



Wissenschaftsevaluation

Neuere Entwicklungen und heutiger Stand
der Forschungs- und Hochschulevaluation
in ausgewählten Ländern

Hans-Dieter Daniel

Center for Science and Technology Studies

The Center for Science & Technology Studies (CEST) compiles and assesses the foundations for political decision-making in the areas of research, higher education, and innovation policy in Switzerland. In this way, it makes its contribution to the development of the Country's scientific, economic, and cultural potential. With this objective in mind, it conducts activities comprising analysis, evaluation, and prospective studies.

Zentrum für Wissenschafts- und Technologiestudien

Das CEST beschafft und überprüft die Grundlagen zur politischen Entscheidungsfindung im Bereich der Forschung, Hochschulbildung und Innovation in der Schweiz. Es leistet damit seinen Beitrag zur Entfaltung ihres wissenschaftlichen, wirtschaftlichen und kulturellen Potentials. Zu diesem Zweck führt es Analyse-, Evaluations- und prospektive Tätigkeiten durch.

Centre d'études de la science et de la technologie

Le CEST rassemble et examine les éléments de base nécessaires à la réflexion et à la décision politique en matière de recherche, d'enseignement supérieur et d'innovation en Suisse. Il contribue ainsi au développement des potentialités scientifiques, économiques et culturelles du pays. C'est dans ce but qu'il procède à des analyses, des évaluations et des études prospectives.

Centro di studi sulla scienza e la tecnologia

Il CEST raccoglie ed esamina gli elementi necessari alla riflessione e alla decisione politica in materia di ricerca, d'insegnamento superiore e d'innovazione in Svizzera. Esso contribuisce così allo sviluppo delle potenzialità scientifiche, economiche e culturali del paese. È a questo scopo che il centro produce delle analisi, delle valutazioni e degli studi prospettivi.

Wissenschaftsevaluation

Neuere Entwicklungen und heutiger Stand der
Forschungs-und Hochschulevaluation
in ausgewählten Ländern

Hans-Dieter Daniel

Februar 2001

CEST 2001/2

Edition	CEST Inselgasse 1; CH-3003 Bern Fon +41-31-322 96 66 Fax +41-31-322 80 70 www.cest.ch
Information	Bernhard Nievergelt und Christian Simon Fon +41-31 322 96 97 bernhard.nievergelt@swr.admin.ch
Graphik	Michael Kurmann
ISBN	3-908194-23-7

The conclusions made in this report engage the author alone.
Die inhaltliche Verantwortung für den Bericht liegt beim Autor.
Le rapport n'engage que son auteur.
L'autore è il solo responsabile del rapporto.

Vorwort des Herausgebers

Das Zentrum für Wissenschafts- und Technologiestudien (CEST) möchte mit dem vorliegenden Bericht von Hans-Dieter Daniel zur kontinuierlichen Reflexion über die Wissenschaftsevaluation anregen. Grundlage dazu bilden die einleitenden Gedanken zur zunehmenden Bedeutung der Wissenschaftsevaluation in allen Feldern (Forschungsorgane, -einrichtungen, -programme, -bereiche), die Darstellung der Evaluationserfahrungen in ausgewählten Ländern und die Diskussion des Forschungsstandes zur Wissenschaftsevaluation. In der Schlussfolgerung findet sich eine allgemeine Einschätzung der Wissenschaftsevaluation. Das CEST veröffentlicht gleichzeitig eine dokumentierte Publikationsliste zu Wissenschaftsevaluation unter www.cest.ch, welche kontinuierlich auf den neuesten Stand gebracht werden soll.

Mit dieser Grundlagenarbeit zur Wissenschaftsevaluation unterstreicht das CEST seine Bemühungen, die eigene Evaluationstätigkeit wissenschaftlich und ethisch zu fundieren und seine Methoden in der Diskussion mit der internationalen Evaluationsforschung und -praxis auszuwählen und zu entwickeln.

Diesen Zielen diene bereits die Publikation "Die Evaluation im Wissenschaftsbereich" (FOP 39/1996). Darin wird die Evaluationstätigkeit des Schweizerischen Wissenschaftsrates reflektiert, die Erfahrungen anderer Institutionen mit Wissenschaftsevaluationen dokumentiert und Grundsätze zu Ethik und Verfahren vorgeschlagen.

Ueber den Autor

Hans-Dieter Daniel ist seit 1995 Professor für empirische Sozialforschung an der Universität Gesamthochschule Kassel. Seine Forschungsaufgaben sind im wissenschaftlichen Zentrum für Berufs- und Hochschulforschung der GH Kassel verankert. Er beschäftigt sich seit Ende der siebziger Jahre mit Fragen der Wissenschaftsentwicklung, der Qualitätssicherung im Wissenschaftsbereich und der Hochschulentwicklung. Besonders markante Beispiele seiner umfangreichen Publikationstätigkeit im Bereich der Wissenschaftsevaluation sind das wegweisende Buch "Evaluation von Forschung: Methoden - Ergebnisse - Stellungnahmen (1988, Hrsg. mit R. Fisch) und seine Habilitation "Guardians of Science: Fairness and Reliability of Peer Review" (1993).

Kontakt: Daniel@hochschulforschung.uni-kassel.de

Summary

Hans-Dieter Daniel provides a survey of how evaluations are conducted in various countries (France, Germany, the Netherlands, Scandinavia, the U.K., the European Union, the USA, and Canada). He discusses the philosophy, methods, and procedures applied with respect to the evaluation of funding agencies, research institutions, universities, research programs, and fields of science and technology. The author makes the observation that recent developments have led to multidimensional approaches, especially to a combination of peer review with quantitative analyses such as bibliometrics. Inasmuch as the linear paradigm (unidirectional knowledge transfer from basic research to practical application) has now become obsolete, indicators for the assessment of trans-disciplinary cooperation and for that of the interface between science and industry are tending to gain in popularity. The author recommends using evaluations as science policy steering instruments. Only an evaluation with clearly defined goals, guaranteed implementation of the results obtained (effect-orientation), and procedures adapted to the nature of the object under evaluation, can be crowned with success.

Zusammenfassung

Hans-Dieter Daniel fasst im Ueberblick zusammen, wie in verschiedenen Ländern (Frankreich, Deutschland, Niederlande, Nordische Länder, Grossbritannien, Europäische Union, USA, Kanada) Wissenschaftsevaluationen durchgeführt werden. Er behandelt die Konzepte, die Methoden und die Praxis für Evaluationen von Institutionen der Forschungsförderung, Forschungsinstitutionen (einschliesslich Hochschulen), Programmen und Wissenschafts- und Technologiebereichen. Der Autor stellt fest, dass der Trend in Richtung mehrdimensionaler Zugänge weist, namentlich die Verbindung von Peer Review mit quantifizierenden Methoden wie Bibliometrie setzt sich durch, während der Abschied vom "Linearen Modell" (von der Grundlagenforschung zur Anwendung) nach Indikatoren ruft, die transdisziplinäre Kooperationen und Schnittstellen zwischen Wissenschaft und Industrie erfassen können. Er empfiehlt, in der Evaluation ein wissenschaftspolitisches Steuerungsinstrument zu sehen. Die Voraussetzungen für dessen erfolgreiche Anwendung sind klare Zielsetzungen, Garantie der Umsetzung der Ergebnisse (Wirkungsorientierung), den Gegenständen angepasste Verfahren.

Résumé

Hans-Dieter Daniel donne un aperçu de l'évaluation telle qu'elle est pratiquée dans différents pays comme la France, l'Allemagne, les Pays-Bas, les pays nordiques, la Grande-Bretagne, l'Union Européenne, les Etats-Unis et le Canada. Il examine les concepts, la méthodologie et la pratique de l'évaluation appliquée aux institutions d'aide à la recherche, aux institutions de recherche (universités inclus), aux programmes et domaines scientifiques et technologiques. L'auteur constate que l'évaluation se développe dans la direction d'une approche pluridimensionnelle, à savoir la combinaison de la peer review avec des méthodes d'analyse quantitative (bibliométrie). Si le modèle linéaire (transfert des connaissances unidirectionnel de la recherche fondamentale aux applications pratiques) est devenu obsolète, les futurs systèmes d'indicateurs privilégieront la coopération entre les disciplines et les interfaces entre la science et l'industrie. L'auteur recommande de se servir de l'évaluation comme instrument de pilotage dans le cadre d'une politique de la science. Des objectifs clairement définis, la garantie de l'application des résultats (accent mis sur les effets de l'évaluation) ainsi que des procédures choisies en fonction des objets sont les conditions préalables à une évaluation réussie.

INHALT

1.	EINLEITUNG	11
1.1	Zielsetzung und Inhalt der Studie	11
1.2	Zur Durchführung der Studie	12
1.3	Zum Verhältnis von Auftraggeber (Staat), Evaluator und Evaluationsobjekt	13
1.4	Wissenschaftsevaluation - ein ubiquitäres und vielfältiges Phänomen	14
2.	EVALUATION VON FORSCHUNGSORGANEN	17
3.	EVALUATION VON FORSCHUNGSEINRICHTUNGEN	19
3.1	Frankreich	19
3.2	Deutschland	21
3.3	Niederlande - das Beispiel der Evaluation von Nederlandse Organisatie voor Wetenschappelijk Onderzoek-Instituten (NWO)	25
3.4	Großbritannien	26
	3.4.1 Das Beispiel der Research Assessment Exercises (RAEs)	26
	3.4.2 Das Beispiel des Economic and Social Research Council (ESRC)	27
3.5	USA: Die Evaluation der Doktoranden-Programme durch den National Research Council (NRC)	28
3.6	Aktueller Stand und Perspektiven der Evaluation von Forschungseinrichtungen	29
4.	EVALUATION VON FORSCHUNGSPROGRAMMEN	31
4.1	Evaluationskonzepte	31
4.2	Programm-Evaluation in ausgewählten Ländern	32
	4.2.1 USA	33
	4.2.2 Kanada	36
	4.2.3 Großbritannien	36
	4.2.4 Deutschland	38
	4.2.5 Finnland	38
	4.2.6 Niederlande - das Beispiel des <i>Assessment of Research Quality</i> der <i>Vereniging van Universiteiten (VSNU)</i>	40
	4.2.7 Europäischen Union	42
4.3	Aktueller Stand und Perspektiven der Programm-Evaluation	44
5.	EVALUATION DER FORSCHUNG IN EINZELNEN DISZIPLINEN UND TECHNOLOGIEBEREICHEN	45
5.1	Finnland	45
5.2	Bundesrepublik Deutschland - das Beispiel der fach- und institutionenübergreifenden Analyse und Bewertung großer Forschungsfelder durch den <i>Wissenschaftsrat</i>	46
5.3	USA - das Beispiel des Experiments in International Benchmarking of US Research Fields durch den <i>National Research Council</i>	46
5.4	Aktueller Stand und Perspektiven der Evaluation von Forschungsfeldern	47
6.	SCHLUSSFOLGERUNGEN UND EMPFEHLUNGEN FÜR DIE WISSENSCHAFTSEVALUATION	49

1. Einleitung

1.1 Zielsetzung und Inhalt der Studie

Modernisierungsdruck und Sparzwang haben dazu geführt, dass seit etwa 1985 immer mehr Länder Evaluationsverfahren für die Wissenschaft entwickeln und einsetzen. Seitdem der quantitative Ausbau des Forschungs- und Bildungssystems zum Stillstand gekommen ist (vgl. Goodstein, 1996) und sich in der Situation des „Steady State“ (Ziman, 1987, 1994) neue Forschungseinrichtungen und Bildungsangebote nur durch Beendigung alter oder veralteter Arbeitsfelder und durch Schließung von Abteilungen oder Instituten realisieren lassen, ist die Suche nach „objektiven Bewertungsmaßstäben“ und reliablen, fairen und validen Evaluationsverfahren weltweit intensiviert worden. In vielen Ländern liegen breitgefächerte Erfahrungen mit jeweils unterschiedlichen Strukturen, Verfahren und Institutionalisierungsformen der Wissenschaftsevaluierung vor.

In historischer Perspektive stand zunächst die Evaluierung individueller Forschungsvorhaben auf der Grundlage des Peer-Review-Verfahrens im Vordergrund. Neben die ex-ante-Evaluation von Forschungsvorhaben ist seit Ende der 80er Jahre zunehmend die ex-post-Evaluation von Forschungsleistungen mit Hilfe bibliometrischer Verfahren getreten. Während das qualitative Peer-Review-Verfahren und die quantitativen Techniken der Bibliometrie zunächst in Konkurrenz zueinander standen, wird heute immer deutlicher, dass sie sich wechselseitig ergänzen sollten, um die Schwächen des jeweils anderen Verfahrens zu kompensieren.

Peer Review, d. h. die Begutachtung wissenschaftlicher Leistungen oder künftiger Leistungsfähigkeit durch ausgewiesene Fachkollegen, gilt einerseits als ein bewährtes Instrument zur Selbststeuerung der Wissenschaft (vgl. Daniel, 1993). Durch kein anderes Evaluationsverfahren scheint in gleicher Weise sichergestellt werden zu können, dass die Auswahl von Bewerbern für leitende Wissenschaftlerstellen, von förderungswürdigen Projektanträgen und von Manuskripten für die Publikation in Fachzeitschriften nach wissenschaftlichen Qualitätsstandards vorgenommen wird. Andererseits ist jedoch nicht zu übersehen, dass eine lediglich auf dem subjektiven Augenschein von Fachkollegen beruhende Evaluation von großen Forschungseinrichtungen und -programmen im allgemeinen erhebliche Schwächen aufweist, wenn sie nicht durch quantitative Verfahren unterstützt wird.

Mit jeweils unterschiedlichen Strukturen, Verfahren und Institutionalisierungsformen werden heute in vielen Ländern evaluiert:

- Forschungsorgane (z. B. Institutionen der Forschungsförderung);
- Forschungseinrichtungen;
- Forschungsprogramme;
- Forschung in einzelnen Disziplinen und Technologiebereichen.

Parallel zur Forschungsevaluierung hat sich die Hochschulevaluierung, insbesondere die Evaluation der Lehre, entwickelt. In zahlreichen Ländern führen die Massenmedien wie bspw. das deutsche Wochenmagazin ‚Der Spiegel‘ seit Ende der 80er Jahre sogenannte Ranking-Studien durch, die dazu beigetragen haben, dass Lehre und Studium heute von sehr unterschiedlichen Akteuren evaluiert werden.

Im Vordergrund der vorliegenden Studie stehen folgende Fragen:

- Wie haben sich Evaluationsmethodik und Evaluationspraxis in der zweiten Hälfte der 90er Jahre international entwickelt?
- Was für Evaluationsansätze wurden international erprobt, welche Erfahrungen wurden mit ihnen gemacht?
- Welche Evaluationsverfahren haben sich besonders bewährt? Welche Auswirkungen hatten die durchgeführten Evaluationen auf das nationale Wissenschaftssystem?
- Wie lässt sich im Hinblick auf Evaluationsmethodik und Evaluationspraxis der heutige Stand charakterisieren?

1.2 Zur Durchführung der Studie

In die Studie einbezogen wurden folgende Länder:

- Kanada - Großbritannien
- USA - Niederlande
- Deutschland - Schweden
- Frankreich - Japan

Die im Zusammenhang mit dem vorliegenden Projekt erarbeitete kommentierte Bibliographie zur Wissenschaftsevaluation (einsehbar unter www.cest.ch) berücksichtigt auch Beiträge aus Dänemark, Finnland und Norwegen, da diese Länder gemeinsam mit Schweden Forschungseinrichtungen unterhalten und Forschungsprogramme aufgelegt haben (vgl. Nordic Council of Ministers, 1998; Helander, 1995), die regelmäßig im Auftrag des *Nordic Council of Ministers* evaluiert werden (vgl. Stähle, 1987) und eine ähnliche, im Vergleich zu anderen Ländern besonders entwickelte Evaluationskultur aufweisen. Am Rande berücksichtigt wurden auch Veröffentlichungen aus Österreich, Italien und Spanien, da diese Länder nicht zuletzt aufgrund des hohen Stellenwertes, den die Europäische Kommission der Wissenschaftsevaluation beimisst, in jüngster Zeit besondere Anstrengungen in diesem Bereich unternommen haben. Die Vorschläge der Europäischen Kommission zur Forschungspolitik sehen beispielsweise vor, dass einzelstaatliche Forschungs- und Technologieprogramme für Teilnehmer aus anderen Ländern geöffnet werden, die Heranziehung internationaler Expertengremien zur Evaluierung einzelstaatlicher Aktivitäten und „Benchmarking“ zur Stimulierung der Forschungstätigkeit in den Regionen. Neben Evaluationsberichten der Europäischen Kommission (vgl. z. B. European Commission, 1997; Fayl et al., 1998) wurden auch Studien aufgenommen, die von der Europäischen Kommission in Auftrag gegeben wurden, um die Methodik der Wissenschaftsevaluation weiterzuentwickeln (vgl. beispielsweise Airaghi et al., 1999). Ferner wurden wichtige aktuelle Veröffentlichungen der OECD (z. B. OECD, 1997), der NATO (vgl. Paces, Pivec & Teich, 1999), der UNESCO (vgl. Cowen, 1996) und der Association of European Universities (vgl. Barblan, 1998) berücksichtigt.

Als Quellen wurden sowohl Publikationen (Zeitschriftenaufsätze, Monographien, Sammelbände, Proceedings) als auch Graue Literatur (z. B. Evaluationsberichte, Empfehlungen und Stellungnahmen, Leitlinien und Kriterienkataloge für Forschungs- und Hochschulevaluationen) recherchiert und ausgewertet. Darüber hinaus wurden Vertreter supranationaler Organisationen (EU-Kommission, OECD, NATO, CRE) sowie nationale Evaluationsexperten und Evaluationsforscher kontaktiert, um den aktuellen Stand der Evaluationsforschung und der Evaluationspraxis in Erfahrung zu bringen. Diesem Ziel diente auch die Teilnahme an zwei internationalen Evaluationskonferenzen (in Berlin und Helsinki) auf Einladung der *Evaluation Unit* der Europäischen Kommission sowie die Teilnahme an einem hochkarätig besetzten Evaluationsworkshop an der Universität Stanford (auf Einladung der Fulbright-Kommission).

Nicht alle Kontakte führten zu den gewünschten Resultaten, da einerseits Evaluationsergebnisse häufig wie Betriebsgeheimnisse behandelt werden und andererseits kommerzielle Evaluatoren, die die Forschungs- und Hochschulevaluation zunehmend als neuen Geschäftsbereich entdecken, darauf achten, dass potenzielle Konkurrenten keinen zu tiefen Einblick in den jeweiligen Evaluationsansatz gewinnen. Darüber hinaus wird die Rezeption nationaler Evaluationserfahrungen dadurch erschwert, dass die entsprechenden Dokumente häufig nur in der jeweiligen Landessprache vorliegen. Dies gilt nach wie vor insbesondere für Japan (in Japan hat der Premierminister 1998 neue Richtlinien für die Forschungsbewertung erlassen; vgl. Science and Technology Agency, 1998; über die Umsetzung dieser Richtlinien informiert ein Bericht der Science and Technology Agency, 1999). Aus den genannten Gründen bilden die in dieser Studie dokumentierten Veröffentlichungen und Evaluationsberichte lediglich die Spitze eines Eisberges.¹

1.3 Zum Verhältnis von Auftraggeber (Staat), Evaluator und Evaluationsobjekt

Es gibt nur sehr wenige politikwissenschaftliche Analysen, die sich aus einer systematisch vergleichenden Perspektive mit dem Verhältnis von Auftraggeber (Staat), Evaluator und Evaluationsobjekt auseinandergesetzt haben. Auch in den Evaluationsberichten selbst werden höchstens am Rande die Rechtsgrundlagen, institutionellen Arrangements und Akteurskonstellationen erläutert. Auf dieser Grundlage ist kaum zu ermitteln, wer wen evaluiert, das heißt, ob eher eine Top-Down-Perspektive politischer Akteure oder eine Bottom-Up-Perspektive wissenschaftlicher Akteure vorliegt (in den Evaluationsberichten findet sich in aller Regel zwar eine Liste der Mitglieder der Evaluationskommissionen; ohne genaue Kenntnis der Personen bleibt jedoch im Dunkeln, ob sie eher dem politischen oder dem wissenschaftlichen System zuzurechnen sind). Die Evaluation von Forschung muss als ein komplexer Interaktionsprozess charakterisiert werden, der nach den vorliegenden Studien in einer großen Vielfalt an institutionellen Arrangements durchgeführt wird, den man kaum entwirren kann. Für die Evaluation von Forschung scheint in besonderer Weise zu gelten, dass „(t)hings economic and social move by their own momentum and the ensuing situations compel individuals and groups to behave in certain ways whatever they may wish to do – not indeed by destroying their freedom of choice but by shaping the choosing mentalities and by narrowing the list of possibilities from which to choose“ (Schumpeter, 1947, zitiert nach Krupp, 1990).

Das Forschungssystem ist zweifellos ein besonders unübersichtliches Gelände, in dem schwere Artillerie (Evaluation) seitens des Staates kaum zum Erfolg führt (vgl. Braun, 1997). Ohne einheimische Späher (Experten, *Peers*) lässt sich aufgrund des Informationsdefizits des politischen Systems Forschung kaum sinnvoll evaluieren. Die Politik ist auf die Mitwirkung der Experten in der Wissenschaft angewiesen. Aus diesem Grunde erfolgt die Evaluation der Forschung in aller Regel durch intermediäre Organisationen zwischen Wissenschaft und Politik. Nicht zuletzt aus Gründen der Wissenschaftsfreiheit evaluiert der Staat als der Geldgeber und die verantwortliche Instanz nur sehr selten selbst die Forschung. In einigen Ländern evaluieren Einrichtungen des Parlaments oder der Regierung von Zeit zu Zeit Forschungseinrichtungen und Forschungsprogramme (Beispiel: *Office of the Comptroller General* in Kanada und in den USA; *U. S. Office of Technology Assessment*). In den meisten Ländern werden stattdessen verselbständigte Evaluationseinheiten mit maßgeblicher Beteiligung von Wissenschaftlern eingeschaltet, die im Auftrag des Staates die ex ante- und ex post-Evaluation von Forschungseinrichtungen und –programmen übernehmen (Deutschland: *Wissenschaftsrat* und

¹ Das CEST veröffentlicht aus Platzgründen diese umfangreiche kommentierte Bibliographie ausschliesslich im Internet: www.cest.ch

Wissenschaftliche Kommission Niedersachsen; Frankreich: *Comité National d'Evaluation de la Recherche* [CNER] und das *Comité National d'Evaluation* [CNE]; England: *Higher Education Funding Council for England* [HEFCE] und *Research Evaluation Steering Group of the Economic and Social Research Council*). Diese intermediären Evaluationssysteme (in Gestalt von Kommissionen oder Agenturen) sind in vielen Ländern ausschließlich dem (Staats-, Bundes-)Präsidenten oder Premierminister verantwortlich, um zu verhindern, dass kurzfristige Ressort- oder Wirtschaftsinteressen „contradict the long-term time-horizon of the scientific mode of operation“ (sog. Haldane-Doktrin). Die intermediären Organisationen beauftragen ihrerseits häufig auf Evaluation spezialisierte Institute (England: *Science Policy Research Unit* [SPRU], *PREST* [*Policy Research in Engineering, Science & Technology*], *Evaluation Associates Ltd.*, *Technopolis Limited*; Frankreich: *Centre de Sociologie de l'Innovation* [CSI]; Kanada: *ARA Consulting Group Inc.*; Niederlande: *Centre for Science & Technology Studies* [CWTS]; USA: *Abt Associates Inc.*, *CHI Research, Inc.*, *COSMOS Corporation*, *Institute for Scientific Information* [ISI], *RAND Corporation*) mit der Durchführung der Evaluation oder mit der Erstellung von Hintergrundmaterialien (z. B. Durchführung von bibliometrischen Studien und Befragungen von Programmteilnehmenden und Stakeholdern). In einigen Ländern führt die Wissenschaft die Evaluation der Forschung vollkommen autonom durch; der Staat beschränkt sich auf die Meta-Evaluation der von der Wissenschaft in eigener Regie durchgeführten Evaluationen (Beispiel: *Association of Universities* [VSNU] in den Niederlanden).

Wissenschaft erweist sich in vielerlei Hinsicht als selbstreferentielles System, das von der Politik kaum regulativ gesteuert werden kann. Im Bereich der Forschung kann der Staat offenbar immer weniger als souveräner Staat agieren (unter anderem auch wegen der zunehmenden staatlichen und privaten Mischfinanzierung der Forschung; staatlicherseits erfolgt die Forschungsförderung darüber hinaus immer häufiger gemeinsam durch Länder/Kantone, Bund und Europäische Kommission). Er tritt deshalb zunehmend als kooperativer Staat auf, der über Verhandlungssysteme operiert, und den Prinzipien „Sachverstand“ und „Akzeptanz“ verpflichtet ist. Eine international vergleichende und systematische Analyse der intermediären Evaluationssysteme stellt ein Desiderat für die zukünftige Wissenschaftsforschung dar. Insbesondere sollte analysiert werden, ob intermediäre Evaluationssysteme die Rationalität politischer Entscheidungen im Bereich der Forschung erhöhen, oder ob intermediäre Organisationen von den Akteuren des Forschungssystems eher genutzt werden, um sich staatliche Kompetenzen und Ressourcen anzueignen (Vereinnahmungsthese), oder gar nach einiger Zeit ein Eigenleben führen, auf das weder politische noch wissenschaftliche Akteure Einfluss haben (Verselbständigungsthese).

1.4 Wissenschaftsevaluation - ein ubiquitäres und vielfältiges Phänomen

Wissenschaftsevaluierung ist heute ein ubiquitäres Phänomen. Das von Daniel & Fisch (1986) für den Zeitraum von 1910 bis 1985 konstatierte exponentielle Wachstum der Literatur zum Themenkomplex Wissenschaftsevaluation hat sich auch in den 90er Jahren weiter fortgesetzt. Während zu diesem Themenkomplex zwischen 1910 und 1949 durchschnittlich weniger als eine Arbeit pro Jahr publiziert wurde, waren es zwischen 1950 und 1959 im Durchschnitt drei Arbeiten pro Jahr, zwischen 1960 und 1969 zehn Arbeiten pro Jahr, zwischen 1970 und 1979 29 Arbeiten pro Jahr und im Zeitraum von 1980 bis 1985 durchschnittlich 35 Arbeiten pro Jahr. Für die letzten 15 Jahre liegen keine genauen statistischen Angaben vor. Eine von Kostoff (1998a) zusammengestellte Bibliographie von neueren Arbeiten zum Themenbereich Forschungs- und Technologieevaluation umfasst 5600 Titel, eine weitere zum *Peer Review* allein 1500 Veröffentlichungen (Kostoff, 1998b). Ähnlich umfangreich ist die Literatur zum Themenbereich Lehrevaluation. Cashin (1988) konnte bereits 1988 1300 empirische Studien zu diesem Teilgebiet der Wissenschaftsevaluierung nachweisen.

Nicht nur die Literatur zum Themenkomplex Wissenschaftsevaluierung, sondern auch die Zahl der einschlägigen Begriffe weist ein exponentielles Wachstum auf. Begriffe wie Auditing, Benchmarking, Bibliometrie, Controlling, Cost Benefit Analysis, Cost Effectiveness Analysis, formelgebundene Mittelverteilung, Leistungsindikatoren, Linkageindikatoren (z. B. für die Vernetzung der Forschungssektoren eines Landes, für den Internationalisierungsgrad der Wissenschaft oder für den Grad der Wissenschaftsabhängigkeit neuer Technologien), Kontraktmanagement, Merit Review, Monitoring, Quality Assessment, Peer Review, Prospektion, Ranking, Systemevaluation und Verbundevaluation machen deutlich, dass es sich beim Themenkomplex Wissenschaftsevaluierung um ein sehr vielfältiges Phänomen handelt. Die Begriffsvielfalt allein im Bereich der Bibliometrie lässt sich auch daran ablesen, dass ein 1994 publiziertes „Dictionary of Bibliometrics“ (Diodato, 1994) 225 Begriffe umfasst.

Angesichts der zunehmenden Zahl von Evaluationen sprechen einige Beobachter bereits von „Evalutis“ und betrachten die sich abzeichnende Evaluationsbürokratie im „evaluative state“ (vgl. Neave, 1998) bzw. die Legitimationsrituale der „audit society“ (vgl. Power, 1997) mit großer Skepsis. Sie befürchten, dass die Verfahren der Qualitätssicherung zu sich selbst genügenden bürokratischen Routinen degenerieren.

2. Evaluation von Forschungsorganen

In den letzten Jahren wurden erstmals auch Forschungsförderungsorganisationen evaluiert. Die in Finnland, den Niederlanden, Schweden und Deutschland durchgeführten Evaluationen von Forschungsorganen durch internationale Expertenkommissionen hatten sehr unterschiedliche Zielsetzungen. In Finnland beispielsweise regte das zuständige Ministerium eine internationale Evaluation der *Academy of Finland* an, um vor allem strukturelle Reformen innerhalb der Forschungsförderungsorganisation durchzusetzen, d. h. die Anzahl der *Research Councils* sollte verkleinert werden, um den Verwaltungsaufwand zu reduzieren und um interdisziplinäre und transdisziplinäre Forschung besser fördern zu können. Durch die internationale Evaluation der *Academy of Finland* (vgl. Blume et al., 1993) konnte dieses Ziel nur teilweise erreicht werden – die sieben *Research Councils* wurden zu vier Forschungsräten (statt zu drei, wie ursprünglich beabsichtigt war) zusammengelegt.

Mit der internationalen Evaluation des *Swedish Natural Science Research Council* (NFR) im Jahre 1998 wurden ähnliche Ziele verfolgt (vgl. Brennan et al., 1998). Die Evaluationskommission hatte den Auftrag, die disziplinäre Gliederung des NFR (Kommissionen für Physik, Chemie, Biologie, Geowissenschaften, Mathematik sowie eine *Analysis, Policy and Strategy Group*) auf ihre Zweckmäßigkeit hin zu überprüfen. Die Kommission empfahl die Beibehaltung der bisherigen Organisationsstruktur und regte an, dass die *Analysis, Policy and Strategy Group* die Bearbeitung multidisziplinärer Forschungsanträge durch die Kommissionen beobachten sollte.

In den Niederlanden wurde 1997 die *Royal Netherlands Academy of Arts and Sciences* (vgl. Brinkman et al., 1997) und 1996 die *Netherlands Organisation for Scientific Research* (vgl. Netherlands Organisation for Scientific Research, 1996) durch internationale Expertenkommissionen evaluiert. Die *Netherlands Organisation for Scientific Research* (NWO) war 1988 gegründet worden. Mit der Evaluation verfolgte das zuständige Ministerium das Ziel, die Leistungen der Forschungsförderungseinrichtung in den ersten sieben Jahren ihres Bestehens zu überprüfen. Die internationale Expertenkommission sollte Antworten auf folgende zehn Fragen finden:

- „Does the work of the NWO adequately reflect the objectives underlying its establishment with reference to both scientific issues (quality, innovation, coordination) and societal issues (promoting research on societal grounds, knowledge transfer)?
- Do the subsidy instruments that the NWO has developed, enable it to carry out its statutory mission? Can the Committee comment at this stage on the way in which the NWO interprets its wider role in the implementation of science policy, a role which is reflected inter alia in the fact that the Minister has delegated certain tasks to the NWO where national priorities, choices and interests are at stake (stimulation programmes and funds for scientific equipment and facilities)?
- Does the NWO pay sufficient attention to evaluation, of both its procedures and the results of its policies?
- What is the Committee's opinion of the functioning of and the relationship between the Government Board and the Boards of each of the scientific fields?
- What is the Committee's opinion of the position of the institutes in the organisation and how they function within it?
- What is the Committee's opinion of the NWO's policy on international cooperation?
- What is the Committee's opinion of the way funds are allocated to the different scientific fields and the considerations on which the allocation of funds is based?

- How are priorities set in respect of promoting research? Are these priorities in line with the actual needs of society, i. e. the needs of companies and society at large?
- Are the tasks to be performed and the available resources in reasonable proportion?
- Does the policy document entitled „Knowledge enriches“ provide a sound basis for future policy in the light of changes taking place in society, and is the NWO adequately equipped to make the changes recommended by the Committee in its future policies?“

Die Stellungnahmen und Empfehlungen der internationalen Expertenkommission können hier nicht im einzelnen wiedergegeben werden. Die Kommission empfahl u. a., daß NWO und die *Royal Netherlands Academy of Arts and Sciences* ein gemeinsames Verfahren für die Evaluation ihrer Forschungsinstitute entwickeln sollten.

In der Bundesrepublik Deutschland hatten die Regierungschefs von Bund und Ländern im Dezember 1996 beschlossen, die gemeinsam geförderten Forschungseinrichtungen zu evaluieren. Bereits im Vorfeld dieses Beschlusses hatten die *Deutsche Forschungsgemeinschaft* (DFG) und die *Max-Planck-Gesellschaft* (MPG) vorgeschlagen, diese beiden Einrichtungen einer gemeinsamen Evaluation im Hinblick auf ihre Stellung im System der gemeinsamen Forschungsförderung zu unterziehen.

Die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) hat daraufhin die Einsetzung einer Kommission international ausgewiesener Experten beschlossen, die klären sollte, „ob im internationalen Vergleich

- die MPG für ihre Institute über geeignete Prinzipien, Verfahren und Möglichkeiten für Neugründungen, Umstrukturierungen und Schließungen sowie für die Qualitätssicherung verfügt und sie erfolgreich anwendet,
- die DFG über geeignete Prinzipien, Verfahren und Instrumente verfügt, um die ihr vorgelegten Anträge angemessen zu bewerten und die richtigen Förderentscheidungen zu treffen,
- die Zusammenarbeit mit den anderen Partnern des Forschungssystems, vor allem den Hochschulen (insbesondere Nachwuchsförderung und Strukturbildung) und dem Wirtschaftsbereich (insbesondere Innovation) funktioniert“ (vgl. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1999).

Die Kommission hat u. a. folgende Empfehlungen vorgelegt:

- „Die Bedingungen für die institutionenübergreifende Kooperation zwischen den einzelnen Bereichen des Forschungssystems sind nachhaltig zu verbessern und Kooperationshemmnisse abzubauen.
- Angesichts der zentralen Bedeutung und Funktion der Universitäten für das Gesamtsystem hält es die Kommission für unerlässlich, dass die Universitäten ihre Aufgaben und Arbeiten in weit größerem Umfange selbstverantwortlich gestalten und angemessene Verfahren der Leistungsbewertung einführen“.

In einem Gespräch mit der BLK über die Konsequenzen der durchgeführten Systemevaluation von DFG und MPG betonte der Präsident der *Max-Planck-Gesellschaft*, Hubert Markl, das Interesse der MPG an einer „Fortsetzung der externen Leistungsbewertung des deutschen Wissenschaftssystems unter Einbeziehung des gesamten Hochschulbereichs“.

3. Evaluation von Forschungseinrichtungen

Beschrieben wird anhand von ausgewählten Beispielen die Evaluation von Forschungseinrichtungen in Frankreich, Deutschland, den Niederlanden, Großbritannien und den USA. Während in Frankreich und den Niederlanden alle staatlich finanzierten universitären und außeruniversitären Forschungseinrichtungen evaluiert werden, werden in Großbritannien die Forschungsleistungen der Universitäten alle fünf Jahre im Rahmen des *Research Assessment Exercise* besonders intensiv, d. h. regelmäßig, flächendeckend und mit Konsequenzen für die Allokation der Forschungsmittel evaluiert. Im Gegensatz dazu werden Universitäten in Deutschland in ihren Forschungsleistungen bisher kaum evaluiert; die Einrichtungen der öffentlich finanzierten außeruniversitären Forschung werden dagegen seit einigen Jahren flächendeckend evaluiert. In den USA hat die Evaluation von Forschungseinrichtungen einen geringeren Stellenwert als die Programmevaluation. Von großer Bedeutung für die amerikanischen Universitäten ist die seit 1982 alle zehn Jahre stattfindende vergleichende Evaluation der Doktoranden-Programme durch den *National Research Council* (z. Z. wird die dritte Studie vorbereitet, die im Jahr 2002 durchgeführt werden soll).

Die Evaluation von Forschungseinrichtungen besitzt - im internationalen Vergleich - in Frankreich den größten Stellenwert. Barré (1999) unterscheidet eine angelsächsische und eine französische Evaluationskultur. Kennzeichnend für die angelsächsische Kultur sei die Finanzierung zeitlich befristeter Forschungsvorhaben durch Forschungsförderungsorganisationen im Rahmen wettbewerblicher Verfahren. Im Zentrum der Evaluation stehen vor allem Forschungsprojekte und -programme. In Frankreich gibt es keine Forschungsförderungseinrichtungen, die beispielsweise mit der *National Science Foundation* in den USA oder den *Research Councils* in Großbritannien vergleichbar wären. Der Grund hierfür ist, dass in Frankreich seit 1982 befristete Arbeitsverträge für Wissenschaftler gesetzlich untersagt sind; Wissenschaftler sind in Frankreich *civil servants* auf Lebenszeit. Es können deshalb keine Drittmittelprojekte durchgeführt werden, da dies die Beschäftigung von Wissenschaftlern mit befristeten Arbeitsverträgen implizieren würde. Konstitutiv für die französische Forschungskultur sind auf Zeit (4 bis 12 Jahre) eingerichtete Forschergruppen (*unité de recherche* oder *laboratoire*), denen universitäre und außeruniversitäre Wissenschaftler auf Dauerstellen angehören. Die Zuweisung von wissenschaftlichem Personal zu Forschergruppen erfolgt vor allem durch das *Centre National de la Recherche Scientifique* (CNRS) und das *Institut National de la Santé et de la Recherche Médicale* (INSERM). In der französischen Forschungskultur zielt die Evaluation aus diesem Grunde nicht auf Projekte und Programme, sondern auf Einzelwissenschaftler, Forschergruppen und Forschungseinrichtungen. „The problematic of evaluation in the French culture is therefore to recruit good people and manage their life-long activity in the structure, placing them in well-organised units with challenging research programmes; it is not to select good projects within the framework of a programme“ (Barré, 1999, S. 46).

3.1 Frankreich

In Frankreich wurde 1984 ein neues Universitätsgesetz verabschiedet, das die systematische und periodische Evaluation der **Universitäten** vorsieht. Zu diesem Zweck wurde das *Comité national d'évaluation* (CNE) gegründet. Das CNE ist vom Ministerium unabhängig und berichtet direkt dem *Président de la République*. Der CNE-Kommission gehören 17 Mitglieder an, die u. a. von beiden Häusern des Parlaments und unabhängigen staatlichen Auditing-Einrichtungen nominiert werden. Die Amtszeit der Kommissionsmitglieder beträgt vier Jahre. Die Kommission wird in ihrer Arbeit von sieben *chargés de mission* und einer Geschäftsstelle unterstützt, der 24 Personen angehören. Das CNE evaluiert jährlich etwa 20 Universitäten. Ende 1994 waren alle Universitäten einmal evaluiert worden.

Der Evaluationsansatz wird als „institutional approach which gives priority to peer reviews over quantitative approaches and indicators“ charakterisiert (vgl. Comité national d'évaluation, 1989; zitiert nach Larédo, 1997, S. 423). Jeweils zwei bis drei Kommissionsmitglieder sind für die Evaluation einer Universität verantwortlich. Die Evaluation wird von einem der sieben *chargés de mission* organisiert, der auch das Informationsdossier über eine Universität erstellt und den Entwurf des Syntheseberichts schreibt.

Das Verfahren umfasst vier Schritte:

- Erstellung eines Informationsdossiers (*tableau de bord*) über die zu evaluierende Universität durch die CNE-Geschäftsstelle. Die Informationen werden mittels Vorortbesuchen und einer *check list* erhoben, die neun Themenbereiche abdeckt (vgl. Chevallier, 1995): (1) Structure of university, faculty, department or service; (2) Personnel; (3) Students; (4) Teaching; (5) Research; (6) Buildings, equipment and resources; (7) Income and expenditure; (8) Libraries and other services; and (9) Institutional environment. Staropoli (1992), der frühere Leiter des CNE, charakterisiert das Informationsdossier als *bill of health*.
- Identifizierung der strategischen Probleme (*crucial issues, evaluation questions*), mit denen sich die Universität auseinandersetzen muss, in gemeinsamen Gesprächen zwischen CNE-Sub-Kommission, externen Experten und Vertretern der Universitätsleitung. Für die Identifizierung der strategischen Probleme der Universität werden auch die Ergebnisse vorausgegangener Evaluationen und seit 1991 der Inhalt des *contrat d'établissement*, d. h. des Vierjahreskontrakts zwischen Universität und Bildungsministerium, berücksichtigt (an der Universität Strasbourg wurde beispielsweise die Biotechnologie als strategisches Problem identifiziert; an der Universität in Rennes war es das bildgebende Verfahren in der Medizin).
- Erstellung von vertraulichen individuellen Expertenberichten zu den gemeinsam identifizierten zentralen Themen der Evaluation (im Durchschnitt werden pro Evaluation neun Experten einbezogen, wobei es sich überwiegend um französische Universitätsprofessoren handelt; jeder Experte bearbeitet jeweils nur ein Thema) sowie Erstellung eines gemeinsam von allen externen Experten verantworteten Syntheseberichts (*rapport d'instruction*). Den Synthesebericht, der zu diesem Zeitpunkt noch keine Stellungnahmen und Empfehlungen enthält, diskutiert die CNE-Sub-Kommission mit Vertretern der Universitätsleitung (Konfrontationsphase).
- Verabschiedung der einstimmig beschlossenen Empfehlungen und Stellungnahmen der CNE-Kommission und Publikation des Berichts mit einer schriftlichen Stellungnahme des jeweiligen Universitätspräsidenten.

Nach Larédo (1997) setzen die Universitäten die Empfehlungen der CNE-Kommission mehr oder weniger um. Als problematisch wird vor allem das wenig systematische *follow up* durch das Bildungsministerium eingeschätzt: „the link with Ministry policy remains erratic, even in cases where the law requires a specific evaluation by the CNE before any new action can be taken“ (S. 425).

1989 richtete das Forschungsministerium in Frankreich das *Comité national d'évaluation de la recherche* (CNER) für die periodische Evaluation aller **staatlich finanzierten außeruniversitären Forschungseinrichtungen** (sog. *research operators*, von denen es in Frankreich etwa 200 gibt) ein (vgl. Larédo, 1997; Larédo & Mustar, 1995). Das Evaluationsverfahren orientiert sich am sog. Garantor-Modell, das 1987 von Chabbal für die Europäische Kommission formuliert wurde. Es handelt sich hierbei um ein *Peer Review*-basiertes Evaluationsverfahren, das weniger die ex post-Evaluation vergangener Leistungen zum Gegenstand hat, als vielmehr Fragen der Organisationsentwicklung und der strategischen Forschungsplanung

sowie die Identifizierung von Stärken und Schwächen des gesamten nationalen Forschungssystems (z.B. wissenschaftlicher Nachwuchs, Bibliotheken, Computerinfrastruktur).

Das Verfahren umfasst vier Schritte (vgl. Comité national d'évaluation de la recherche, 1993a, b, 1994, 1996):

- Erstellung eines Informationsdossiers (*tableau de bord*) über die Forschungseinrichtung durch die CNER-Geschäftsstelle (CNER's *chargés de mission*) unter Beteiligung externer Experten.
- Identifizierung der strategischen Probleme (*crucial issues, evaluation questions*) der Forschungseinrichtung in gemeinsamen Gesprächen zwischen CNER-Kommission (insgesamt zehn Mitglieder, die für sechs Jahre von den wichtigsten Wissenschaftsorganisationen nominiert werden; die Kommission ist vom Forschungsministerium unabhängig und berichtet direkt dem französischen Staatspräsidenten), externen Experten und Vertretern der zu evaluierenden Forschungseinrichtung auf der Grundlage des zuvor von der Forschungseinrichtung akzeptierten Informationsdossiers.
- Erstellung von individuellen Expertenberichten zu den gemeinsam identifizierten zentralen Themen der Evaluation (wobei jeder Experte jeweils nur ein Thema bearbeitet) sowie Erstellung eines gemeinsam von allen externen Experten verantworteten Syntheseberichts (*rapport d'instruction*); Diskussion des Syntheseberichts mit Vertretern der evaluierten Forschungseinrichtung und Aufforderung zu schriftlicher Stellungnahme (Konfrontationsphase).
- Publikation der einstimmig beschlossenen Empfehlungen und Stellungnahmen der CNER-Kommission zu der evaluierten Forschungseinrichtung.

Das Verfahren ist sehr zeitaufwendig, führt jedoch zu Empfehlungen, die von den evaluierten Forschungseinrichtungen bisher offenbar ausnahmslos akzeptiert und umgesetzt wurden. Kritisch angemerkt wird von Larédo (1997) jedoch, dass im Vergleich zu den lokalen Effekten der Einfluss auf das nationale Forschungssystem vergleichsweise gering ist. Er sieht die Ursache hierfür im Garantor-Modell von Chabbal (1987), das zwar zu robusten, relevanten und glaubwürdigen Evaluationsergebnissen führt, allerdings von der falschen *follow up*-Annahme ausgeht, dass „the mere existence of written evaluations is sufficient to stimulate ‚political‘ debate and change“. Larédo (1997) empfiehlt deshalb „the creation of ‚hybrid fora‘ or ‚arenas‘ for debate“.

Das Centre National de la Recherche Scientifique (CNRS) hat 1989 ein Auditing-Verfahren eingeführt. Die *Délégation des audits* hat die Aufgabe, Fragen von zentraler Bedeutung für das CNRS aufzugreifen und der Leitung zu berichten. Die Themen werden von der CNRS-Leitung bestimmt (z. B. interdisziplinäre Forschung, Computerinfrastruktur, Beurteilung der Wissenschaftler durch das *Comité national*, vgl. Centre National de la Recherche Scientifique, 1995). Gemeinsam mit einer internen Expertengruppe wählt die *Délégation des audits* externe Auditoren aus. Deren Analysen werden innerhalb des CNRS diskutiert und von der internen Expertengruppe zu einem Bericht zusammengefasst. Der ganze Prozess wird von dem *Comité consultatif des audits* beobachtet, das als Garantor die Qualität der Berichte gewährleisten soll.

3.2 Deutschland

In Deutschland hat sich nach dem 2. Weltkrieg ein plurales und dezentrales Forschungssystem entwickelt, dem neben den Universitäten und Fachhochschulen die außerhochschulischen Forschungseinrichtungen der *Max-Planck-Gesellschaft* (MPG), der *Fraunhofer-*

Gesellschaft (FhG), der *Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren* (HGF) sowie der *Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz* (WGL) angehören.

Die Sicherung der wissenschaftlichen Qualität der Forschung geschieht in Deutschland einerseits durch Evaluation und andererseits durch Teilnahme an wettbewerblichen Verfahren. Abgesicherte institutionelle Förderung gilt zwar weiterhin als unabdingbare Voraussetzung für mittel- und langfristig angelegte Forschung, die nicht auf unmittelbare Verwertungserfolge angelegt ist und sich aus ihnen finanziert. Bei Fehlen von Qualitätssicherungsverfahren kann die Grundausstattung jedoch auch die Bereitschaft zur Teilnahme am wissenschaftlichen Wettbewerb lähmen. Es wird daher zunehmend als sinnvoll erachtet, die institutionelle finanzielle Förderung durch Qualitätssicherungsverfahren wie regelmäßige Berichterstattung, Kosten- und Leistungsrechnung, interne und externe Evaluierung zu ergänzen.

Die Regierungschefs von Bund und Ländern hatten im Jahre 1996 die Evaluation aller gemeinsam geförderten Forschungseinrichtungen bei Erhalt der Ressourcen für die Forschung und mit Stellentransfermöglichkeiten zwischen den Bereichen gemeinsam finanzierter Forschungseinrichtungen beschlossen.

Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (Blaue Liste-Einrichtungen): Seit 1979 besteht ein externes wissenschaftliches Evaluierungsverfahren für die Institute der *Blauen Liste*, das Vorbild für die Evaluierung der Institute der ehemaligen *Akademien der Wissenschaften* in der DDR gewesen ist. Ziel der externen Evaluierung war und ist, den Qualitätsstandard nachzuweisen und Institute, die den Ansprüchen an wissenschaftliche Qualität nicht oder nicht mehr genügen, aus der gemeinsamen Förderung von Bund und Ländern herauszunehmen oder im Rahmen einer auf Dauer angelegten Umstrukturierung ihre Finanzierung zurückzufahren. Alle 79 Institute der Leibniz-Gemeinschaft werden derzeit durch den Wissenschaftsrat einer Evaluierung durch externe Gutachter unterzogen. Diese Evaluierung, die 1995 begonnen wurde, wird Mitte 2000 abgeschlossen sein (Leitfaden und Fragebogen für die Bewertung der Einrichtungen sind dokumentiert in Wissenschaftsrat 1999a, b). Das von Bund und Ländern gewünschte Evaluierungsverfahren hat für mehrere Institute kritische Resultate ergeben. Sie bestätigen die Notwendigkeit der Evaluierung und legen nahe, weitere Maßnahmen der Qualitätssicherung einzuführen. Nach Beendigung der jetzigen Bewertungsrunde durch den Wissenschaftsrat wird die laufende Qualitätskontrolle in ähnlicher Form in den Händen des im November 1998 konstituierten Senats der WGL liegen.

Die Qualitätssicherung der Forschung in der *Max-Planck-Gesellschaft* und der *Fraunhofer-Gesellschaft* und ihren Instituten folgt jeweils eigenen Prinzipien.

Max-Planck-Gesellschaft: Für die MPG ist nach wie vor das Harnack-Prinzip wichtig, wonach für einen wichtigen Forschungsgegenstand um eine herausragende Forscherpersönlichkeit ein Institut gebildet wird. Da die Identität von Institut und Forschungskonzeption einer einzelnen Forscherpersönlichkeit bei der heutigen Komplexität und Arbeitsteilung nicht mehr in Reinheit durchzuhalten ist, treten andere Qualitätssicherungsverfahren hinzu (vgl. Krull, 1995). Bei der MPG sind dies die Einrichtung wissenschaftlicher Beiräte an ihren Instituten (die Regelungen für das Fachbeiratswesen sind beschrieben in *Max-Planck-Gesellschaft*, 1998; für den notwendigen Inhalt der Statusberichte der MPG-Institute vgl. *Max-Planck-Gesellschaft*, 1999), regelmäßig stattfindende Begehungen von Instituten sowie personalwirtschaftliche Maßnahmen, die bei Ausscheiden der Leitungspersonlichkeiten auch personell einen Neuanfang und eine Umorientierung der Institute und ihrer Abteilungen ermöglichen. Die MPG hat darüber hinaus eine Systemevaluation der Gesellschaft in die Wege geleitet, deren Ergebnisse 1999 vorgelegt wurden (vgl. Krull, 1999).

Fraunhofer-Gesellschaft: In der FhG wird die permanente Evaluation der Institute durch den Markt durch eine regelmäßige interne Evaluation, zum Teil unter Heranziehung externer Experten, ergänzt. Die FhG hat darüber hinaus ihre Stellung, Effizienz und Effektivität im System der gemeinsam geförderten Forschungseinrichtungen durch eine unabhängige Kommission bewerten lassen (vgl. Fraunhofer-Gesellschaft, 1998).

Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren: Die HGF-Zentren unterziehen sich in regelmäßigen Abständen einer Erfolgskontrolle durch unabhängige externe Beratungsgremien und Gutachtergruppen. Der Senat der HGF überprüft z. Z. die Rahmenbedingungen für Strukturkommissionen und Begutachtungsverfahren und erarbeitet Empfehlungen für ihre Ausgestaltung. Grundsätze für die Erfolgskontrolle bei Großforschungseinrichtungen hat der Bundesminister für Forschung und Technologie 1988 formuliert.

Deutsche Forschungsgemeinschaft: An den Hochschulen und in Teilen der außerhochschulischen Forschungseinrichtungen wird die Qualitätssicherung mittels Wettbewerb durch die Kriterien und Verfahren der *Deutschen Forschungsgemeinschaft* (DFG) nachhaltig unterstützt. Die Strukturen ihrer Förderung hat die DFG in einem gemeinsamen Verfahren mit der MPG durch eine internationale Kommission evaluieren lassen (vgl. Krull, 1999).

Die Bundesforschungsministerin und die Forschungsminister der Länder diskutierten am 20. 3. 2000 mit den Präsidenten der *Deutschen Forschungsgemeinschaft* und der *Max-Planck-Gesellschaft* (MPG) Konsequenzen der im Jahre 1999 durchgeführten Systemevaluation beider Organisationen. Der Präsident der *Max-Planck-Gesellschaft* betonte in diesem Gespräch das Interesse der MPG an einer „Fortsetzung der externen Leistungsbewertung des deutschen Wissenschaftssystems unter Einbeziehung der staatlichen Forschungseinrichtungen und des gesamten Hochschulbereichs“. In der *Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung* (BLK) besteht Einvernehmen, dass dabei Verfahren der Kosten- und Leistungsrechnung und der Output-Steuerung mit dem Ziel der Effizienzsteigerung zu entwickeln sind.

Die gemeinsam von Bund und Ländern geförderten Forschungseinrichtungen sollen künftig Kosten- und Leistungsrechnungen erstellen, die eine wissenschaftliche und ökonomische Bewertung ermöglichen. Seitens der *Bund-Länder-Kommission* wird antizipiert, dass die Bewertung auf der Leistungsseite Quantifizierungsprobleme bereiten wird. Sie hält es daher für notwendig, Annäherungskriterien wie die Publikation in referierten international bedeutenden Fachzeitschriften, eingeladene Vorträge auf wissenschaftlichen Kongressen, internationale und nationale Forschungsk Kooperationen, Drittmitteleinwerbung oder die Zahl der angemeldeten Patente zu verwenden. Diese Leistungsindikatoren werden nach Meinung der BLK zur Transparenz beitragen und damit eine Grundlage für die Bewertung bieten.

Die universitäre Forschung wird bisher in Deutschland kaum systematisch evaluiert. Lediglich eines der insgesamt 16 Bundesländer hat eine wissenschaftliche Kommission eingerichtet, die Forschungsevaluationen an Hochschulen durchführen soll. Es handelt sich hierbei um die *Wissenschaftliche Kommission Niedersachsen*, die im Oktober 1997 auf der Grundlage eines Kabinettsbeschlusses und in Abstimmung mit der Landeshochschulkonferenz auf Dauer eingerichtet wurde. Die *Wissenschaftliche Kommission Niedersachsen* hat 16 Mitglieder, von denen zehn stimmberechtigt sind. Der Vorsitzende und die Mitglieder der *Wissenschaftlichen Kommission* werden für einen Zeitraum von drei Jahren auf Vorschlag des Ministeriums für Wissenschaft und Kultur durch den Ministerpräsidenten berufen. Ihre Amtszeit kann einmalig um eine weitere Periode verlängert werden. Die stimmberechtigten Mitglieder des Gremiums sind renommierte Wissenschaftlerinnen und Wissenschaftler aus anderen Bundesländern. Sie vertreten ein breites Spektrum von Fachrichtungen. Weiterhin gehören der Kommission sechs

Vertreter von niedersächsischen Institutionen aus den für die Wissenschafts- und Forschungspolitik wichtigen Bereichen (Wissenschaft, Wirtschaft, Politik und andere Fördereinrichtungen) an. Zu den Aufgaben der *Wissenschaftlichen Kommission Niedersachsen* gehört die Entwicklung eines leistungsfähigen Evaluationsverfahrens für die Forschung. Das Verfahren für die Evaluation der Forschung in den beiden Pilotdisziplinen Chemie und Geschichte ist im Oktober 1999 angelaufen.

Die Evaluation der Forschung soll vor allem dazu dienen (vgl. Wissenschaftliche Kommission Niedersachsen, 1999):

- die Hochschulen bei der Entwicklung eines eigenen, klar definierten Forschungsprofils und bei der Standortbestimmung im nationalen und internationalen Vergleich zu unterstützen;
- den Hochschulen Kriterien für die eigenverantwortliche Planung und Durchführung von Maßnahmen zur Qualitätssicherung und –verbesserung an die Hand zu geben;
- Wissenschaftlern und Studierenden aus Niedersachsen, anderen Bundesländern und dem Ausland, die im Zusammenhang mit wissenschaftlichen Kooperationen, Arbeitsplatz- oder Studienortwahl Informationen suchen, eine valide Einschätzung der Forschungsschwerpunkte und Forschungsleistungen niedersächsischer Einrichtungen zu ermöglichen;
- die Profilbildung der Hochschulen gezielter von seiten des Landes durch Ausstattung, Berufungspolitik etc. zu fördern;
- zur Entwicklung von Kriterien für die qualitätsorientierte Mittelvergabe durch das Land im Rahmen der Einführung von Globalhaushalten für die Hochschulen beizutragen.

Das Verfahren ist fachbezogen und landesweit institutionenübergreifend angelegt. Dabei sollen Vergleiche mit Institutionen außerhalb des Landes herangezogen werden. Außeruniversitäre Forschungseinrichtungen, die aus Landesmitteln finanziert werden, können in die Evaluation einbezogen werden. Die Evaluation soll nach dem Prinzip des *informed peer review* erfolgen. Die Einschätzungen und Empfehlungen der Gutachter werden in Abschlußberichten niedergelegt. Die betroffenen Hochschulen erhalten die Möglichkeit, vor der Veröffentlichung der Berichte zu diesen Stellung zu nehmen. Die Bewertungsmaßstäbe, die generell in allen Verfahren berücksichtigt werden sollen, lassen sich in drei Bereiche untergliedern:

- Qualität und Relevanz (folgende Aspekte sollen im einzelnen Berücksichtigung finden: Innovativität der an der Institution geleisteten Forschung; wissenschaftliche Ausstrahlung; Interdisziplinarität oder besonderer Stellenwert als Einzeldisziplin; Kooperation mit anderen Forschungseinrichtungen auf regionaler und nationaler Ebene; Intensität und Qualität der internationalen Zusammenarbeit, z. B. durch Forschungsk Kooperationen, EU-Projekte, gemeinsame Veröffentlichungen, Gastwissenschaftler, gemeinsam betreute und gegenseitig anerkannte Promotionen, Internationalisierung von Nachwuchsforschern im Rahmen von Hochschulpartnerschaften und Mobilitätsprogrammen);
- Effektivität und Effizienz (das Evaluationsverfahren soll die Frage nach dem Verhältnis von Aufwand und Erfolg beantworten. Hierbei soll der Mitteleinsatz insbesondere unter Zugrundelegung folgender Daten bewertet werden: Personalausstattung, Sachausstattung, Drittmittel);
- Strukturpolitische Bedeutung von Forschungseinrichtungen auf Landesebene.

Die Evaluationen soll nach folgendem Schema ablaufen:

- Entscheidung über durchzuführende Evaluationen in der Lenkungsgruppe (hierbei handelt es sich um Hochschulvertreter, die von der LHK benannt werden, um ein Mitglied

des Wissenschaftsministeriums sowie um Mitglieder der *Wissenschaftlichen Kommission Niedersachsen*);

- Selbstbericht der Institute unter Einbeziehung der eigenen Zielvorstellung, des angestrebten Profils – auch im überregionalen Kontext – und der Stellung und Bedeutung des betroffenen Fachs in der Hochschule, einschließlich Datenerhebung nach einem bestimmten Raster;
- Benennung einer Gutachtergruppe durch die Lenkungsgruppe;
- Zusammenstellung der Berichte durch die Geschäftsstelle und Weiterleitung an die Gutachter;
- Vorbereitende Besprechung mit den Gutachtern;
- Besuche der Institute mit möglichst weitgehender Kontinuität in der Gutachtergruppe;
- Zusammenfassung der Evaluationsergebnisse und Empfehlungen der Gutachter in einem Abschlußbericht;
- Stellungnahme der Hochschulleitung und des evaluierten Fachs/Fachbereichs;
- Diskussion der Berichte in der Lenkungsgruppe;
- Beratung der Evaluationsergebnisse unter strukturellen Gesichtspunkten in der *Wissenschaftlichen Kommission* zusammen mit den Ergebnissen der Lehrevaluation durch die ZEvA; Freigabe zur Veröffentlichung;
- Empfehlungen an das Land und ggf. die einzelnen Hochschulen zur weiteren Planung durch die *Wissenschaftliche Kommission*;
- Nachfolgende Maßnahmen: Nach zwei Jahren berichten die Hochschulen, ob und wie die Gutachterempfehlungen umgesetzt wurden.

3.3 Niederlande - das Beispiel der Evaluation von Nederlandse Organisatie voor Wetenschappelijk Onderzoek-Instituten (NWO)

Die NWO hat die Aufgabe, die Forschung an den niederländischen Universitäten und in ihren eigenen Instituten zu fördern. Für die Evaluation der NWO-Institute werden international besetzte Evaluationskommissionen eingerichtet. Den Kommissionen gehören „topnotch foreign researchers“ an sowie mindestens ein niederländischer Wissenschaftler, der mit der spezifischen niederländischen Situation vertraut ist. Die NWO-Institute können Gutachter nominieren. Die Mitglieder der Evaluationskommissionen werden jedoch von den *NWO Research Councils* bestellt. Der Evaluationsauftrag beinhaltet sowohl retrospektive als auch prospektive Elemente: „The retrospective element should be an assessment of the qualitative and quantitative output of the institute in an international perspective, while the prospective element should be an assessment of the institute’s research plans for the coming period. Overall it should consider the necessity for the institutes continued existence“.

Auf einer vierstufigen Notenskala – unsatisfactory, satisfactory, good, excellent – sollen folgende Aspekte beurteilt werden:

- „The (technical) scientific quality of the research conducted over the last six years as well as the research topics chosen for the coming six years, seen in an international perspective and with the emphasis on the institute’s contribution to (developing) relevant national and international programmes;
- The volume (critical mass) of the institute’s research activities, including where relevant the operation of existing facilities, their coherence, and the prospective demand for these facilities by various (sub-) disciplines;

- Where appropriate, the facility's infrastructure – including assistance provided by the institute – for the conduct of research by (potential) users from various (sub-) disciplines;
- Where appropriate, the institute's plan for further development of present/future facilities in view of user demand;
- The balance, in terms of both of volume and of quality, between the regular funding of the institute by NWO and funding obtained in competition (from funding agencies, including NWO, and in the form of contracts from industry and other private or non-profit organisations and government departments);
- The appropriateness of the organisation and its management“ (vgl. Nederlandse Organisatie voor Wetenschappelijk Onderzoek, 1998).

3.4 Großbritannien

3.4.1 Das Beispiel der Research Assessment Exercises (RAEs)

In Europa war Großbritannien das erste Land, das eine systematische und flächendeckende Evaluation der universitären Forschung auf den Weg gebracht hat. Die sog. *Research Assessment Exercises* (RAEs) wurden bisher 1986, 1989, 1992 und 1996 durchgeführt. Die nächste Evaluation wird im Jahr 2001 stattfinden (nähere Informationen stehen im Internet unter der URL <http://www.rae.ac.uk/> zur Verfügung). Methodik und Prozedere der letzten *Research Assessment Exercise* lassen sich wie folgt beschreiben:

Zu Beginn der Evaluation wird die gesamte britische Hochschulforschung in Disziplinen (*Units of Assessment*) untergliedert. 1996 waren es insgesamt 69 Disziplinen. Für jede Disziplin wird ein eigenes *Assessment Panel* – bestehend aus durchschnittlich sechs bis zehn Experten – eingerichtet, das für die Evaluation der Forschung im engeren Sinne zuständig ist. Jedes Universitätsinstitut wird einer Disziplin zugeordnet (in Ausnahmefällen werden Universitätsinstitute aufgeteilt und verschiedenen Disziplinen zugeordnet).

Die Universitätsinstitute übermitteln dem zuständigen *Assessment Panel* umfangreiche Informationen über ihr Forschungsprofil. Diese Informationen decken folgende Bereiche ab: Überblick über das Personal; Anzahl der Studenten in der post-gradualen Phase; Stipendien für Forschungsstudenten; Drittmittel; Forschungsprogramme und Forschungspläne. Die wichtigste Informationsbasis sind jedoch die Publikationen: Jeder forschungsaktive Wissenschaftler muss dem *Assessment Panel* seine vier besten Publikationen (oder andere öffentlich zugängliche Manifestationen von Forschungs-Output) zur Verfügung stellen. Das britische Evaluationsmodell gibt somit eindeutig der Qualität der Forschung gegenüber der Quantität des Forschungsoutputs den Vorzug.

Auf der Basis der übermittelten Informationen bewertet das *Assessment Panel* die Qualität der Forschungsleistung jedes einzelnen Instituts. Die Bewertung erfolgt anhand einer siebenstufigen Notenskala mit den Ausprägungen 1, 2, 3b, 3a, 4, 5 und 5*. 5* stellt die beste Note dar und impliziert höchste Qualität der Forschungsleistungen im internationalen Vergleich (*international excellence*). Die Ergebnisse der Bewertung werden in Form eines Universitätsrankings pro Disziplin veröffentlicht.

Die Evaluationsergebnisse fließen direkt in die Finanzierungsformel des *Higher Education Funding Councils* (HEFC) für die Grundfinanzierung der universitären Forschung ein. Die Formel lautet vereinfacht ausgedrückt: Volumen mal Qualität, d. h. die Gesamtzahl der forschungsaktiven Wissenschaftler (die funktionsabhängig unterschiedlich gewichtet werden)

wird mit der Note aus der Forschungsevaluation multipliziert. Die Grundfinanzierung, die eine Universität für die Forschung von dem HEFC erhält, entspricht der Summe der Einzelbeträge für die einzelnen Universitätsinstitute.

Die *Research Assessment Exercises* sind in den ersten Jahren heftig kritisiert worden. Sie wurden für den vermeintlichen Niedergang der britischen Forschung verantwortlich gemacht. Die vorliegenden bibliometrischen Studien stützen diese Einschätzung jedoch nicht (vgl. z. B. Braun, Glänzel & Schubert, 1989). Nach einer in *Science* publizierten Studie von Sir Robert M. May (1998), Chief Scientific Adviser to the UK Government, steht Großbritannien gemessen an seiner Forschungseffizienz nach wie vor an der Weltspitze (Indikator: Papers published in science, including medicine and engineering, in 1990, 1993, and 1996 per £1 million of science base expenditure in 1987, 1990, and 1993, respectively). Auch die große Mehrheit der Teilnehmer an einem Expertenseminar des *British Council* stimmte in einem abschließenden Votum gegen die Behauptung, daß die Forschungsevaluierung der britischen Wissenschaft geschadet habe (British Council, 1998, S. 28).

3.4.2 Das Beispiel des Economic and Social Research Council (ESRC)

Der *Economic and Social Research Council* (ESRC) evaluiert die vom ihm unterstützten *Research Centres* entweder acht oder elf Jahre nach ihrer Gründung. Die Evaluation findet im achten Jahr statt, wenn das Forschungszentrum einen Antrag auf Verlängerung der Förderung um weitere fünf Jahre stellt. Die Evaluation bezieht sich dabei auf folgende Themen:

- the extent to which a centre met its agreed objectives;
- the contribution of a centre to its research field in terms of capacity building and methodological and theoretical development;
- the quantity and quality of a centre's work in terms of its outputs;
- the impact of the centre's work on its user communities;
- the value added by concentrating resources in a centre; and
- the potential value of further investment in the research area.

Das Evaluationsverfahren beschreibt der *Economic and Social Research Council* (1997) wie folgt:

„Centres are required to submit a report summarising their academic and non-academic achievements, together with examples of publications or other output. The report and supporting material is subjected to peer review and is then assessed, together with the reviewers comments, by a panel of experts drawn from the relevant academic and user constituencies.

The peer reviewers and the membership of the evaluation panel itself are arranged by the Research Evaluation Committee (REC) in consultation with the Research Priorities Board, the ESRC's Research Support Teams and the relevant Centre Director. Issues such as the nature of a centre, and its size and complexity, are taken into account when deciding on the size and make-up of the review panel and the number of peer reviewers to be approached. Up to thirty peer reviewers (academics, 'users' and referees based overseas) are approached and an evaluation panel is normally composed of six or seven members. A member of the Policy and Evaluation Division (PED) acts as secretary to the panel.

In addition to the peer review assessments, the panel conducts a site visit to the centre to gather further evidence. A typical site visit will consist of interviews with Centre staff, presentations of research by the Centre and interviews with university representatives. The site visit allows an opportunity for the centre to highlight directly to the panel parti-

cular aspects of its work and for issues of concern to both parties to be explored in depth.

Following the site visit the panel prepares its evaluation report to the REC.

The Centre has the opportunity to respond to the evaluation report, and this response is attached to the report before its submission to the REC. Once approved by the Committee, the report is forwarded to the Research Priorities Board. The Board then considers the evaluation report, alongside the bid for further support. The centre's bid to the Research Priorities Board for further support is submitted after the presentation of its report to the evaluation, and is peer reviewed independently by the Research Priorities Board“.

Ein Forschungszentrum wird nach elf Jahren evaluiert, wenn es keinen Antrag auf Fortsetzung der Förderung stellt. In diesem Fall wird die ex post-Evaluation des Zentrums auf der Grundlage eines Ergebnisberichts durch einen unabhängigen und besonders ausgewiesenen Wissenschaftler („an independent senior figure with an established research record in the area“) durchgeführt. Dieser legt den Ergebnisbericht *Peers* mit der Bitte um Stellungnahme vor, befragt Mitarbeiter des Zentrums und andere relevante Gruppierungen.

3.5 USA: Die Evaluation der Doktoranden-Programme durch den National Research Council (NRC)

In den USA bieten über 300 Universitäten Ph.D.-Programme an. Die Qualität der Programme wird seit 1966 landesweit vergleichend beurteilt. Die beiden ersten Studien wurden vom *American Council on Education* publiziert (vgl. Cartter, 1966; Roose & Anderson, 1970). Im Jahr 1982 hat der *National Research Council* die Schirmherrschaft über die Studie übernommen und seither zwei weitere Studien vorgelegt (vgl. Jones, Lindzey & Coggeshall, 1982; Goldberger, Maher & Ebert Flattau, 1995). Die nächste vergleichende Evaluation der Ph.D.-Programme ist für das Jahr 2002 geplant.

In der 1995 veröffentlichten Studie „Research-Doctorate Programs in the United States“, die sich auf das Studienjahr 1992/93 bezieht und Doktoranden-Programme in 41 Disziplinen vergleicht, wurden folgende Bewertungsmaßstäbe zugrunde gelegt (vgl. Goldberger et al., 1995):

1993 Ratings (Reputationseinschätzungen im Rahmen des NRC National Survey of Graduate Faculty):

- Scholarly quality of program faculty;
- Program effectiveness in educating research scholars and scientists;
- Change in program quality in the last five years.

Faculty

- Total number of faculty participating in the program;
- Percentage of full professors participating in the program;
- Percentage of program faculty with research support (1986 - 1992).

For Arts and Humanities:

- Total number of awards and honors attributed to program faculty for the period 1986 - 1992;
- Percentage of program faculty that have received at least one honor or award for the period 1986-1992.

For the fields in Engineering and the Sciences:

- Percentage of program faculty publishing in the period 1988 to 1992;
- The ratio of the total number of program publications in the period 1988 - 1992 to the number of program faculty;
- Gini coefficient for program publications, 1988 - 1992;
- The ratio of the total number of program citations in the period 1988 - 1992 to the number of program faculty;
- Gini coefficient for program citations, 1988 - 1992.

Students

- The number of full and part time graduate students enrolled in the Fall of 1992;
- The percentage of full and part time female graduate students enrolled in the Fall of 1992;
- The number of Ph.Ds produced by that program for the period academic year 1987 - 1988 to 1991 - 1992.

Doctoral Recipients

- The percentage of Ph.D.s awarded to women during the period July 1986 - June 1992;
- The percentage of Ph.D.s known to be awarded to underrepresented minorities during the period July 1986 - June 1992;
- The percentage of Ph.D.s known to be awarded to U.S. Citizens and Permanent Residents during the period June 1986-June 1992;
- The percentage of Ph.D.s having research assistantships who reported their primary form of support;
- The percentage of Ph.D.s having teaching assistantships who reported their primary form of support;
- Median time lapse from entering graduate school to receipt of Ph.D. in years.

Die Studie „Research-Doctorate Programs in the United States“ stellt zweifellos die weltweit umfassendste und gründlichste Ranking-Studie dar (für eine kritische Würdigung vergleiche Webster & Skinner, 1996). Zu der Fülle US-amerikanischer Ranking-Studien gibt es im Internet eine WWW-Seite (URL: <http://www.library.uiuc.edu/edx/rankings.htm>) sowie weiterführende Literaturhinweise (URL: <http://www.library.uiuc.edu/edx/rankbib.htm>). In dem von Walleri & Moss (1995) herausgegebenen Sammelband werden weitere wichtige Ranking-Studien vorgestellt und kritisch gewürdigt.

3.6 Aktueller Stand und Perspektiven der Evaluation von Forschungseinrichtungen

Die Evaluierung der Leistung von Forschungseinrichtungen erfolgt zunehmend mehrdimensional, d. h. anhand eines Indikatorenbündels. Die Evaluationskriterien beziehen sich in der Regel auf Umfang, Qualität (vgl. Buchholz, 1995) und Relevanz der erzielten Forschungsergebnisse. Sie berücksichtigen häufig auch die jeweils selbstgesteckten Ziele der Forschungseinrichtung und legen in der Regel einen internationalen Bewertungsmaßstab an.

Bei der Bildung bibliometrischer Indikatoren sollte in Zukunft stärker darauf geachtet werden, dass Spezifika der Datenbanken (vgl. Münzinger & Daniel, 1992; Quoniam et al., 1995) und der auf Indikatorenbildung spezialisierten Institute (z. B. *Institute for Scientific Information*

[ISI]; *CHI Research, Inc.*; *Centre for Science and Technology Studies at Leiden University [CWTS]*, *Information Science and Scientometrics Research Unit [ISSRU]* der *Ungarischen Akademie der Wissenschaften* in Budapest) die Ergebnisse nicht verzerren bzw. verfälschen (vgl. Glänzel, 1996; Moed, 1996).

Ein fairer Leistungsvergleich von Forschungseinrichtungen sollte in Zukunft auch immer bestehende Größen und Ausstattungsunterschiede berücksichtigen, damit nicht fälschlicherweise als Leistung ausgewiesen wird, was in Wirklichkeit nur Ausdruck der Größe oder der personellen und infrastrukturellen Ausstattung der Institutionen ist. Johnes (1996) konnte für Großbritannien beispielsweise zeigen, dass 80 % der Leistungsunterschiede zwischen den Universitäten durch Unterschiede in den Input-Variablen erklärt werden konnten.

Da eine absolute Quantifizierung von Forschungsleistungen nicht möglich ist, sollte die Höhe der Leistung immer nur im Vergleich zu anderen „Leistungserbringern“ bestimmt werden (z. B. nach dem Benchmarking-Ansatz, vgl. Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine, 1999; Adams et al., 1998; Murphy, 1995). Der Auswahl der Vergleichsobjekte kommt dabei eine besondere Bedeutung zu.

Da die Leistungsunterschiede innerhalb der Forschungsinstitutionen häufig wesentlich größer sind als die Leistungsunterschiede zwischen den Forschungseinrichtungen, sollte im Rahmen von Evaluationen immer auch ermittelt werden, welche Wissenschaftler die eigentlichen Leistungsträger sind (vgl. Narin & Breitzman, 1995; Narin & Hamilton, 1996; Vitt, 1998). Dieser Punkt ist ganz besonders wichtig, weil keine Forschungseinrichtung als Ganzes einem Kriterium für Spitzenforschung genügen kann. Forschungsergebnisse werden immer von Personen erarbeitet, die sich hinsichtlich ihrer Produktivität und Reputation in hohem Maße unterscheiden. Innerhalb jeder Einrichtung existieren ganz erhebliche Leistungsunterschiede zwischen den einzelnen Wissenschaftlern (in der Regel in der Größenordnung von eins zu zehn). „Das Leben der Universität hängt“, um Karl Jaspers (1961) zu zitieren, „an den Persönlichkeiten, nicht an der Institution, welche nur Bedingung ist“ (S. 124). Dies gilt nicht nur für Universitäten, sondern für Forschungseinrichtungen insgesamt (vgl. z. B. National Institute of Science and Technology Policy, 1994).

4. Evaluation von Forschungsprogrammen

4.1 Evaluationskonzepte

Studien zur Evaluation von Forschungsprogrammen haben im wesentlichen vier Wurzeln, die ihr jeweiliges Konzept in unterschiedlichem Maße prägen:

- Verfahren zur Bewertung von Forschung und Technologie: sie wurden ursprünglich von der *scientific community* entwickelt; ihr wichtigstes Verfahren ist das *peer review* – die Bewertung von Leistungen oder Forschungsvorhaben durch „Fachgenossen“;
- Verfahren zur Überwachung der Verwendung öffentlicher Finanzmittel: sie werden von den Institutionen der Haushaltskontrolle angewendet (z. B. *Bundesrechnungshof, General Accounting Office*);
- Ansätze der Wirkungsforschung: sie kommen aus der sozialwissenschaftlichen Wirkungs- und Implementationsforschung und aus der amerikanisch geprägten Tradition der *policy analysis*;
- Neuerdings auch Elemente des Projektmanagements und des F&E-Controlling: diese Verfahren stehen in der Tradition der Betriebswirtschaftslehre und der Managementlehre.

Programmevaluationen lassen sich nach Evaluationstyp, Untersuchungsgegenstand, Evaluationsmethode und Rahmenbedingung unterscheiden:

Evaluationstyp:

- ex ante-Evaluation
- mid-term-Evaluation
- begleitende Evaluation (Monitoring)
- ex post-Evaluation

Untersuchungsgegenstände:

- Zugrundeliegende Annahmen des Programms
- Klientel des Programms
- Beabsichtigte Effekte (Output)
- Unbeabsichtigte Effekte
- Langzeiteffekte (Outcome)
- Zielerreichungsgrad (Effektivität)
- Verhältnis von Aufwand und Ertrag (Effizienz)
- Mitnehmereffekte
- Alternative Politik-Instrumente
- Administrative Effizienz des Programms

Evaluationsmethoden und Datenerhebungsverfahren:

- Vorher/Nachher-Vergleich
- Kontrollgruppen-Ansatz
- Fallstudien, Interviews

- Methoden-Mix
- Perzeptive Daten (Umfragen, Interviews etc.)
- Quasi-objektive Daten (Publikations- und Patentstatistiken etc.)

Rahmenbedingungen:

- Ressourcen, Zeitrahmen für Evaluationsstudie
- Untersuchtes Zeitfenster
- Unabhängigkeit der Evaluation
- Nutzung der Ergebnisse.

Aus der Kombination der jeweiligen Optionen resultieren in der Praxis der Programmevaluation sehr unterschiedliche Evaluationskonzepte.

Nach Meinung vieler Experten sind Forschungsprogramme schwerer zu evaluieren als Forschungseinrichtungen, da die Programmziele in aller Regel „weich“, „qualitativ“ und dabei in wachsendem Maße komplex formuliert werden. Eine Ausnahme stellt das ROAME-System des britischen *Departments of Trade and Industry* dar, das *verifiable objectives* verwendet (ROAME = **R**ationale, **O**bjectives, **A**ppraisal, **M**onitoring, **E**valuation, vgl. URL: <http://www.miti.go.jp/info-e/cM99809e.html>).²

Wegen der genannten Schwierigkeit kommen wirtschaftswissenschaftlich motivierte Ansätze der Messung ökonomischer Wirkungen staatlicher Forschungsprogramme nur selten zur Anwendung. Link (1996) beschreibt, wie mit Hilfe ökonometrischer Modelle und spieltheoretischer Überlegungen Kosten/Nutzen-Vergleiche staatlich geförderter Forschungsprogramme durchgeführt werden können. Wilk (1997) erläutert das Verfahren der Kosten-/Nutzen-Analyse am Beispiel der Gesundheitsforschung. Ökonomische Verfahren für die ex ante- und für die ex post-Evaluation von Programmen im Bereich der agrarwissenschaftlichen Forschung beschreiben Alston, Norton & Pardey (1998). Da für ökonometrische Analysen lange Zeitreihen von Input-, Output- und Outcome-Daten benötigt werden (mindestens 25 bis 30 Jahre), die in der Regel für derart lange Zeiträume nicht zur Verfügung stehen, empfehlen die Autoren für die Evaluation von „noncommodity research programs“ ökonomische „surplus“-Methoden. Der Nutzen ökonomischer Evaluationsverfahren wird vor allem darin gesehen, dass alle Beteiligten gezwungen sind, sich darüber klar zu werden, „what they are trying to achieve – and how best to get there“ (S. 512).

4.2 Programm-Evaluation in ausgewählten Ländern

Dargestellt wird die Evaluation von Forschungsprogrammen in den USA, in Kanada, Großbritannien, Deutschland, Finnland und den Niederlanden, da für diese Länder die Praxis der Programmevaluation vergleichsweise gut dokumentiert ist (die Evaluation von Forschungsprogrammen spielt in Frankreich keine Rolle: „the evaluation culture is not oriented towards projects and programme evaluation, but on individual, laboratory and institutional levels“, vgl. Barré, 1999). Darüber hinaus wird das Verfahren der Europäischen Kommission für die Evaluation der RTD-Programme der Europäischen Union beschrieben. Die Europäische Kommission versucht im Sinne einer Evaluationspyramide ex ante-, mid-term- und ex-post-Evaluation zu integrieren und Programmevaluation und strategische Planung zu verklammern.

² Vgl. auch Punkt 4.2.3

4.2.1 USA

Die USA haben eine lange Tradition in der Durchführung von Programm-Evaluationen im Bereich der Forschungs- und Technologiepolitik (vgl. Chubin, 1987; Cozzens, 1987, 1995; Cozzens & Melkers, 1997; Feller, 1988; Kostoff, 1998b; Mullins, 1987; Narin, 1987; Roessner, 1999). Eine prominente Rolle kommt dabei der Legislative zu: den Haushalts- und Fachausschüssen der beiden Häuser des Parlaments und die dem Parlament zuarbeitenden Einrichtungen wie das *General Accounting Office* (GAO), der *Congressional Research Service* und das *Office of Technology Assessment* (das 1995 aufgelöst wurde), die sowohl aus eigener Initiative wie auch im Auftrag des Parlaments Evaluationen durchführen (vgl. z. B. General Accounting Office, 1996). Die z. Z. wichtigste Evaluationsagentur ist das *General Accounting Office*, der US-amerikanische Rechnungshof, zu dessen gesetzlichen Aufgaben seit 1970 die Evaluation staatlicher Programme gehört.

Nach Cozzens (1995) hatte die ex post-Programmevaluation in den USA allerdings in den letzten Jahren nur einen geringen Stellenwert: „Over the last decade, ex post research assessment at the program level in the United States has seemed much less active than the equivalent activities in Europe, both west and east. 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“ (S. 351). Für den Zeitraum bis 1993 konstatieren Roessner & Melkers (1995) eine relativ geringe Reichweite der Fragestellungen in den amerikanischen Programmevaluationen sowie eine begrenzte Nutzung, die sie auf die Fragmentierung der Macht im amerikanischen politischen System zurückführen. Programme zur Förderung der Grundlagenforschung würden kaum kontrovers diskutiert und lösten daher nur eine geringe Nachfrage für externe Evaluationen aus; sie würden vorwiegend intern evaluiert und die Ergebnisse eher genutzt, um kleinere Programmanpassungen vorzunehmen. Nach Eldon & Devine (1985) haben Programmevaluationen in den USA mehr symbolische als substantielle Bedeutung.

Cozzens (1995) erwartet jedoch, dass die Bedeutung von Programmevaluationen in den USA in den nächsten Jahren stark zunehmen wird. Das 1993 verabschiedete *Government Performance and Results Act* verpflichtet alle Bundesbehörden nach einem mehrjährigen Moratorium „to do strategic planning, to report annually in terms of performance indicators, and to use program evaluation in its strategic planning and performance reporting efforts“ (S. 352). Alle Bundesbehörden (einschließlich NSF und NIH) mussten dem Kongress bis 1998 einen mindestens 5-jährigen strategischen Forschungsplan vorlegen (in die Beratungen über die Entwicklung des Forschungsplans ist der Kongress einzubeziehen; der strategische Forschungsplan ist alle drei Jahre zu aktualisieren), ab 1999 ihre aus dem Forschungsplan abgeleiteten Leistungsziele für das kommende Haushaltsjahr schriftlich fixieren und ab dem Jahr 2000 dem *Office of Management and Budget* und dem Kongress einen jährlichen Leistungsbericht vorlegen, in dem dargelegt werden muss, ob die im Vorjahr genannten Ziele erreicht worden sind (Kanada, Neuseeland und Großbritannien haben ähnliche Gesetze verabschiedet).

Das Gesetz dürfte die Praxis der Programmevaluation in den USA stark beeinflussen, da „quantitative evaluation studies are relatively rare in the United States, and are concentrated in the fundamental science agencies, NSF and NIH, and at the General Accounting Office. Most of the assessment of federal research programs is descriptive, and far removed from the sort of quantification of performance the Government Performance and Results Act requires. A large array of quantitative tools for evaluation has been described in the literature, but few of them are used in practice. To respond to GPRA, research programs and agencies thus face the challenge of either choosing among an array of options they have largely avoided in the past, or developing new ones“ (Cozzens, 1995, S. 354).

Um in Erfahrung zu bringen, auf welche Weise das Gesetz implementiert werden kann, hat die *National Science Foundation* mehrere Pilotprojekte durchgeführt. Für das Programm

„Science and Technology Centers“, durch das 25 Kompetenzzentren an amerikanischen Universitäten langfristig gefördert werden, wurden beispielsweise zunächst folgende Ziele und Leistungsindikatoren formuliert (vgl. Cozzens, 1995, S. 357):

Goal 1: Address interdisciplinary research problems beyond the capabilities of single investigators or small research groups.

Performance Indicators:

- amount, quality, and interdisciplinary nature of research products, such as publications and collaborations;
- novel integrative approaches to research problems;
- use and development of unique research instruments and facilities;
- size and duration of collaborative research projects;
- sponsorship of leadership and coordinating activities for the community.

Goal 2: Promote partnerships with states, industry, and national research laboratories, for knowledge and technology transfer from academia to other sectors.

Performance Indicators:

- creation of mechanisms for knowledge/technology transfer: alliances, contractual arrangements, organizational structures;
- tangible products of knowledge/technology transfer such as amount, nature, and quality of joint research and personnel exchange;
- patent disclosures, licensing arrangements;
- new products, spin-off companies;
- use and acquisition of shared research instrumentation;
- reciprocal training programs and hiring practices.

Goal 3: Produce graduates at all levels with unique interdisciplinary science and engineering training.

Performance Indicators:

- number of students trained, including minorities and females;
- inter-departmental/interdisciplinary nature of advising committees;
- career tracks of graduates;
- creation of interdisciplinary courses, curricula, majors, programs.

An den genannten Leistungsindikatoren monierte das *Office of Management and Budget* fehlende Angaben zur jeweiligen Ausgangssituation (*baseline*) und zu den Grenzwerten für gute und weniger gute Leistung (die *Strategic Plans* und *Performance Plans* der Ministerien und Bundesbehörden, einschließlich der *National Science Foundation*, stehen im Internet unter der folgenden URL: <http://www.gao.gov/new.items/gpra/gpra.htm>).

Das *Committee on Science, Engineering, and Public Policy* (COSEPUP) der *National Academy of Sciences*, der *National Academy of Engineering* und des *Institute of Medicine* hat eine Reihe von Expertenworkshops durchgeführt, um zu klären, wie das *Government Performance and Results Act* im Bereich der Grundlagenforschung umgesetzt werden kann (vgl. COSEPUP, 1999b). Da der praktische Nutzen der Grundlagenforschung nur selten im voraus be-

stimmt und quantifiziert werden kann, sei es notwendig, „to evaluate the performance of basic research programs by using measures not of practical outcomes but of performance, such as the generation of new knowledge, the quality of research, the attainment of leadership in the field, and the development of human resources“ (COSEPUP, 1999b, S. 30). COSEPUP empfiehlt deshalb für die Evaluation von Forschungsprogrammen im Bereich der Grundlagenforschung ein *Expert Review*-Verfahren einzusetzen, das aus den drei Komponenten *Peer Review*, *Relevance Review* und *International Benchmarking (leadership)* besteht: „The most effective way to evaluate research programs is by expert review. The most commonly used form of expert review of quality is peer review. This operates on the premise that the people best qualified to judge the quality of research are experts in the field of research. This premise prevails across the research spectrum, from basic research to applied research. A second form of expert review is relevance review, in which potential users and experts in other fields or disciplines related to an agency’s mission or to the potential application of the research evaluate the relevance of research to the agency’s mission. A third form of expert review is benchmarking, in which an international panel of experts compares the level of leadership of a research program relative to research being performed worldwide“ (COSEPUP, 1999, S.9-10).

Darüber hinaus empfiehlt COSEPUP (1999b, S. 32), dass Expertenpanels von Fall zu Fall

- die Ergebnisse von bibliometrischen Studien und internationalen Benchmarking-Vergleichen berücksichtigen sowie
- den langfristigen Nutzen von Forschungsprogrammen anhand von ökonomischen Erfolgsindikatoren und den Hinweisen aus Fallstudien und retrospektiven Analysen abschätzen sollten.

Die Pro- und Kontra-Argumente zu den einzelnen Methoden der Forschungsevaluation fasst COSEPUP (1999b, S. 19) wie folgt zusammen:

Table: Current methods used for evaluating research

Methods	Pro	Con
Bibliometric analysis	Quantitative; useful on aggregate basis to evaluate quality for some programs and fields	At best, measures only quantity; not useful across all programs and fields; comparisons across fields or countries difficult; can be artificially influenced
Economic rate of return	Quantitative; shows economic benefits of research	Measures only financial benefits, not social benefits (such as health-quality improvements); time separating research from economic benefit is often long; not useful across all programs and fields
Peer review	Well-understood method and practice; provides evaluation of quality of research and sometimes other factors; already an existing part of most federal-agency programs in evaluating the quality of research projects	Focuses primarily on research quality; other elements are secondary; evaluation usually of research projects, not programs; great variance across agencies; concerns regarding use of „old boy networks“; results depend on involvement of high-quality people in process
Case studies	Provides understanding of effects of institutional, organizational, and technical factors influencing research process, so process can be improved; illustrates all types of benefits of research process	Happenstance cases not comparable across programs; focus on cases might involve many programs or fields making it difficult to assess federal-program benefit
Retrospective analysis	Useful for identifying linkages between federal programs and innovations over long intervals of research investment	Not useful as a short-term evaluation tool because of long interval between research and practical outcomes
Benchmarking	Provides a tool for comparison across programs and countries	Focused on fields, not federal research programs

4.2.2 Kanada

Kanada wird von vielen Experten als weltweit führend in der Evaluation von staatlichen Programmen bezeichnet. Seit 1977 müssen alle von Ministerien und staatlichen Behörden durchgeführten Programme periodisch evaluiert werden (in Kanada fördern elf Bundesministerien Forschung und Technologie; allein dem Industrieministerium unterstehen 13 Behörden, die Forschung und Technologie fördern. Dazu gehören z. B. die *Canadian Space Agency*, der *National Research Council of Canada*, der *National Sciences and Engineering Research Council of Canada* und der *Social Sciences and Humanities Research Council of Canada*). Unter Programmevaluation wird die periodische, unabhängige und objektive Überprüfung und Bewertung eines Programms verstanden, um die Angemessenheit seiner Ziele, seines Designs und seiner intendierten und nicht intendierten Ergebnisse zu ermitteln. Evaluationen sollen die Daseinsberechtigung von Programmen überprüfen. Untersucht werden jeweils die Programmbeurteilung, die öffentliche Auswirkungen und die Kosteneffizienz. Die politischen Ziele und Zwecke der Programmevaluation werden innerhalb der Regierung vom *Treasury Board* formuliert. Nach McDonald & Teather (1997) hat sich an der politischen Zielsetzung von Programmevaluationen in den letzten 20 Jahren nur wenig geändert. Die Autoren konstatieren allerdings in jüngster Zeit „a shift in focus from resources, activities and outputs to achievements“.

Im Jahre 1978 richtete das Parlament das *Office of the Auditor General* ein. Es hat u. a. die Aufgabe, die Evaluationsaktivitäten der Ministerien zu überwachen und entsprechende Beratungsfunktionen auszuüben. Es berichtet direkt dem Parlament. Die Abteilung für Programmevaluation innerhalb des *Office of the Auditor General* hat detaillierte Anleitungen für die Durchführung von Programmevaluationen ausgearbeitet (vgl. Auditor General, 1993a, b), die Ministerien/Behörden sind jedoch weitgehend autonom in ihren Entscheidungen über die Durchführung und Reichweite der von ihnen durchgeführten Evaluationen. Die überwiegende Anzahl von Evaluationen wird von externen Beratern nach öffentlichen Ausschreibungen durchgeführt (vgl. z. B. ARA Consulting Group, 1997; Muir & Williams, 1994).

Die Nutzung von Evaluationsergebnissen im Bereich der Forschungs- und Technologiepolitik erfolgt nach Roessner & Melkers (1995) auch in Kanada nur begrenzt. Die Richtlinien für Programmevaluationen des *Office of the Auditor General* sehen lediglich vor, dass ein Plan für angemessene Umsetzungs-Aktionen der Evaluationsempfehlungen entwickelt werden soll.

4.2.3 Großbritannien

Eine wichtige Rolle im Bereich der Programmevaluation spielt in Großbritannien das *Department of Trade and Industry* (DTI). In den frühen 80er Jahren überprüfte das DTI alle staatlichen Programme daraufhin, ob sie einen Zusatznutzen (*additionality*) erbringen, der ohne staatliche Förderung (etwa im Bereich der Informationstechnologie oder der Materialforschung) sonst nicht erfolgt wäre. Das DTI hat ein Programmplanungsverfahren entwickelt, das die Evaluation formalisiert in den Politikplanungsprozess einbezieht. Das Programmplanungsverfahren ROAME sieht für alle Programmvorschlüsse ein standardisiertes Format vor, das sicherstellen soll, dass die Begründung für ein Programm explizit gemacht wird, um Programm-Management und Programm-Evaluation zu erleichtern. Das Akronym ROAME steht für fünf Themenkomplexe, zu denen bei der Programmplanung explizit Stellung zu nehmen ist: *Rationale, Objectives, Appraisal, Monitoring, Evaluation* (gelegentlich wird *Policy Feedback* als ein sechster Themenkomplex genannt, d. h. die Berücksichtigung der Evaluationsergebnisse bei der Planung zukünftiger Programme). Neben das Rationale und den Zielen des Programms muss Stellung genommen werden zum Verfahren und zu den Kriterien für die Auswahl des Programms (*appraisal*) und zur Beobachtung der Programmdurchführung (*monitoring*); im Abschnitt *Evaluation* sind vorläufige Vorschläge für die Programmbewertung

vorzulegen, insbesondere sind der Zeitpunkt und ein Ablaufplan für das endgültige Design der Evaluation festzulegen. Das ROAME-Programmplanungsverfahren wird in Großbritannien von den meisten Ministerien eingesetzt, so zum Beispiel auch vom Umweltministerium (vgl. Cunion, 1995).

Im Jahre 1983 beauftragten das DTI, das Verteidigungsministerium und der *Science and Engineering Research Council* das *Programme of Policy Research in Engineering, Science and Technology* (PREST) an der Universität Manchester und die *Science Policy Research Unit* (SPRU) an der Universität Sussex mit der Evaluation des britischen Forschungsprogramms im Bereich der Informationstechnologie. Die Evaluation des sog. Alvey-Programms, die 1991 abgeschlossen wurde, ist die umfangreichste Evaluation eines Forschungs- und Technologieförderprogramms, die bisher durchgeführt wurde (vgl. Guy et al., 1991). Neben den wissenschaftlichen und technischen Leistungen, wurde auch der Beitrag des Programms zur Vernetzung des Universitäts- und Industriesektors untersucht sowie der Beitrag des Programms zur Revitalisierung der britischen IT-Industrie (Georghiou & Roessner, 1999, stellen Methoden und Techniken der Evaluation von Technologieprogrammen im Überblick dar).

1990 hat das DTI eine rollende Fünfjahresplanung eingeführt, die Programmevaluation und strategische Planung miteinander verklammern soll. Der Hauptzweck von Programmevaluationen scheint seither darin zu bestehen, gegen Ende eines Programms gegebenenfalls neuen oder zusätzlichen Forschungsbedarf zu identifizieren.

Alle Forschungsprogramme des Handels- und Industrieministeriums (DTI) werden von der *Technology and Standards Assessment Unit* (AU) innerhalb des Direktorats *Technology, Economics, Statistics and Evaluation* (TESE) intern evaluiert (vgl. Barber, 1999). Der AU gehören fünf hauptberufliche Evaluatoren an; die AU verfügt jährlich über etwa £ 200.000, um externe Berater zu engagieren. Das DTI versucht die Qualität der Programmevaluationen durch eine *Evaluation Methodology Group* (EMG) zu sichern, der Evaluationsexperten aus dem gesamten Ministerium angehören. Der *Evaluation Methodology Group* (EMG) müssen alle Evaluationsvorhaben zur Begutachtung vorgelegt werden. Sie nimmt auch Stellung zu allen Evaluationsberichten und berichtet dem *Evaluation and Policy Improvement Committee* (EPIC), dem der Finanzdirektor des Ministeriums vorsitzt.

Für die Forschungsförderung in Großbritannien sind die *Research Councils* von großer Bedeutung. Gegenwärtig sind sieben Research Councils installiert:

- Biotechnology and Biological Sciences Research Council;
- Economic and Social Research Council;
- Engineering and Physical Sciences Research Council;
- Medical Research Council;
- Natural Environment Research Council;
- Particle Physics and Astronomy Research Council;
- Council for the Central Laboratory of the Research Councils.

Der *Economic and Social Research Council* (ESRC) hat im Jahre 1998 Richtlinien für die ex post-Evaluation von Forschungsprogrammen publiziert. Der ESRC lässt pro Jahr drei bis vier abgeschlossene Förderprogramme evaluieren. Die Evaluation wird etwa zwölf Monate nach Abschluss des letzten im Rahmen des Programms geförderten Projekts begonnen. Das *Research Evaluation Committee* (REC) des ESRC beauftragt nach einer Ausschreibung der geplanten Evaluation einen Berater (*consultant*) mit der Durchführung der Programmevaluation. Die Evaluation umfasst in der Regel folgende Bereiche:

- „Evaluation of the programme’s achievements in relation to its overall aims and objectives;
- Assessment of the programme’s academic dissemination activity;
- Review of interaction with research users, and impact on these groups;
- Examination of the ‚value-added‘ by organising the research within the programme funding-mode;
- Assessment of the contribution of individual projects to the overall programme objectives;
- Assessment of the Programme Director’s contribution to the programme;
- Consideration of the Programme’s potential contribution to future research developments“.

Die Evaluation stützt sich auf folgende Informationsquellen: Abschlussberichte der einzelnen Projekte; Publikationen und zur Veröffentlichung eingereichte Manuskripte der Projektleiter; Endbericht des Programmdirektors mit Angaben zum Beitrag jedes einzelnen Projekts zu den Programmzielen; Auswertung der übrigen Outputs auf der Basis der ESRC *Research Grants Archive and Database* (REGARD); Interviews mit dem Programmdirektor, dem Vorsitzenden des Programmbeirates, den Projektleitern und Repräsentanten der wichtigsten Nutzergruppen; schriftliche Befragung von Projektleitern zu Disseminationsaktivitäten und von potentiellen nicht-akademischen Anwendern zum wahrgenommenen Nutzen des Programms.

4.2.4 Deutschland

Evaluationsstudien zur Wirkungsanalyse forschungs- und technologiepolitischer Programme finden in der Bundesrepublik Deutschland seit etwa 1980 zunehmend Anwendung. Anders als in anderen Ländern werde solche Studien in Deutschland überwiegend nicht vom politisch-administrativen System selbst, sondern von verwaltungsexternen, unabhängigen Forschungsinstituten durchgeführt. Anfang der 90er Jahre beauftragte der Bundesminister für Forschung und Technologie (BMFT) das *Fraunhofer-Institut für Systemtechnik und Innovationsforschung* (FhG-ISI), eine Metaevaluation der über 50 Evaluationsstudien durchzuführen, die der BMFT seit 1985 in Auftrag gegeben hatte (vgl. Kuhlmann & Holland, 1995).

Vergleichsweise häufig wurden die Wirkungen der Förderprogramme in den Bereichen marktnahe industrielle Forschung und Entwicklung, Technologie-Transfer und Technik-Diffusion evaluiert; seit dem Ende der 80er Jahre wurden darüber hinaus zunehmend auch Programme, die Forschung und Entwicklung bestimmter Technologielinien stärken wollen, analysiert (z. B. Gesundheits-, Laser-, Materialforschung, Dünnschichttechnologien, Mikrosystemtechnik, Umweltforschung und –technologie, Biotechnologie). Kaum untersucht wurden bisher in Deutschland im Rahmen von Evaluationsstudien die Förderung von Programmen im Bereich der Grundlagenforschung.

Die Zahl von Programmevaluationen hat in den letzten Jahren zwar zugenommen, im ganzen gesehen ist sie in der Bundesrepublik Deutschland allerdings noch lange nicht selbstverständlich. Dies lässt sich nicht zuletzt daran ablesen, dass weniger als 0,3 Prozent der Fördergelder in die Evaluation bzw. Qualitätskontrolle fließt (vgl. Harhoff, D. & Röller, L.-H., 1999; Kuhlmann & Holland, 1995, S. 72).

4.2.5 Finnland

Der finnische Wissenschaftsbereich wird auf der Basis des „dual support system“ finanziert: die Universitäten und außeruniversitären Forschungseinrichtungen erhalten eine öffentliche Basisfinanzierung sowie zweckgebundene öffentliche Mittel, die von intermediären Organisa-

tionen wie der *Academy of Finland* und dem *Technology Development Center of Finland* (TEKES) vergeben werden. Die zweckgebundene Finanzierung ist das eigentliche Steuerungsinstrument der öffentlichen Hand. Aus diesen Mitteln werden Forschungsprojekte, Forschungsprogramme, Kompetenzzentren und Ausbildungsprogramme (z. B. Graduate schools) finanziert. In Finnland werden seit Mitte der 90er Jahre alle öffentlich geförderten Forschungsprogramme durch international besetzte Expertenkommissionen evaluiert (vgl. Luukkonen, 1997). Die *Academy of Finland* beschreibt das Verfahren für die ex post-Evaluation durch international besetzte Expertenkommissionen wie folgt: „The Programme Group (members include representatives of the Academy and other funding organizations, people from universities, research institutes, and industry) decides what kind of evaluation procedures should be used. In many cases, an external (foreign) evaluation panel is nominated. Both the projects and the programme itself is evaluated. Since scientific level is hard to evaluate, bibliometrical indicators and doctoral degrees are about the most important indicators“ (Tero Majamaa, persönliche Mitteilung). Von Fall zu Fall werden auch potentielle Anwender und Kooperationspartner befragt: „Some evaluations include surveys to industrial or social users of research as well as co-operative partners. The aim is to combine a quality assessment with one of relevance“ (Helander, 1995, S. 394). Auch TEKES lässt alle geförderten Projekte und Programme ex post evaluieren (vgl. Numminen & Hämäläinen, 1995). Die *Academy of Science* führt darüber hinaus im Auftrag der Regierung seit 1982 und TEKES seit 1988 eine Fülle von Programmevaluationen durch (vgl. Helander, 1995). So hat die *Academy of Science* beispielsweise das finnische „Research Programme of Climate Change“ evaluiert. Allein TEKES hat in den letzten Jahren etwa 50 Evaluationen von Programmen vorgenommen (z. B. „Energy Technology Programme, 1993-1998“, „Pharmaceutical Technology Programme“, „Synthesis Technology Programme, 1992-1996“). Neben der *Academy of Finland* und TEKES führt das *Technical Research Centre of Finland* (VTT), die größte staatliche Forschungseinrichtung Finnlands, seit 1983 ebenfalls zahlreiche Programmevaluationen durch (z.B. „Research Programme of Molecular Modelling“, „Research Programme of Chemical Reaction Mechanisms“). Das VTT erstellt z. Z. eine Metaanalyse der bisher durchgeführten Evaluationen in Finnland. Das Projekt „Research Evaluation in Finland: Practices and Experiences, Past and Present“ (Projektbearbeiter Juha Oksanen) soll bis Mai 2000 abgeschlossen sein. Die finnischen Evaluationsaktivitäten werden vom *Science and Technology Policy Council* (STPC) begleitet, der aus seinen Beobachtungen Empfehlungen für die Weiterentwicklung der Evaluationskultur in Finnland ableitet. Die aus den Programmevaluationen sich ergebenden Erkenntnisse werden in hohem Maße zur Optimierung des finnischen Forschungs- und Technologiesystems genutzt, das nach einer Studie des IMD zu den erfolgreichsten der Welt zählt (vgl. Garelli, 1999).

Der *Science and Technology Policy Council* hat 1998 eine Liste grundlegender Fragen entwickelt, auf die ex post-Evaluationen Antworten finden sollen:

Effects to research:

- What are the effects to the level and orientation of research (the balance between scientific orientation and market orientation)?
- What are the implications to public/private partnerships in research?

Effects to human capital:

- What are the effects to the human capital development?
- Has the balance between the supply and demand of skilled labour improved?

Economic effects:

- What is the contribution to the structural renewal of the Finnish economy?

- What is the contribution to the knowledge intensity / productivity / competitiveness of industry?
- What new businesses or firms have been created?
- What is the contribution to the growth of knowledge intensive services?
- Is there any indication of decreasing or increasing marginal utility of R&D investment?

Employment effects:

- What are the direct / indirect effects to job creation (including the substitution effects) in short and long terms?
- What is the contribution of process and product innovation to job creation?

Systemic effects:

- Is there a balance between efforts to promote knowledge generation / distribution / utilisation?
- What new co-operation structures have been created (networking effects)?
- What is the contribution to institutional renewal in the R&D / higher education system?
- What are the effects to the leverage of public R&D intervention?
- What are the implications to the innovation system as a whole?

International effects:

- What are the effects to Finland's ability to benefit from international R&D co-operation?
- Is there any indication that knowledge is ,hollowing-out' more than before?
- What are the implications to Finland's attractiveness as a home base for high tech industries?

4.2.6 Niederlande - das Beispiel des *Assessment of Research Quality* der *Vereniging van Universiteiten* (VSNU)

In den Niederlanden legte das Wissenschaftsministerium im Jahre 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 gegebenenfalls –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. Die zur Zeit laufende zweite Runde der fachbezogenen Forschungsevaluationen der niederländischen Universitäten erstreckt sich von 1998 bis 2003.

Für die Durchführung der Forschungsevaluation wurde ein offizielles zweisprachiges (Niederländisch/Englisch) Dokument, das *Protocol 1998*, erstellt, das die wichtigsten Grundsätze,

Richtlinien und die zeitliche Abfolge der einzelnen Evaluationsschritte festhält (vgl. Verening van Universiteiten, 1998).

Die Ziele der Forschungsevaluation werden im *Protocol 1998* wie folgt beschrieben:

- „Assessment of the quality of the research programmes on the basis of self-evaluations prepared by faculties or institutes; appraisal of how the work is related to the group’s mission, as well as to the mission of the faculty or the institute of which the group’s programme is a part. This will enable the committees to place the activities of the group within the context of the group’s environment;
- Evaluation of the missions of faculties or institutes, as well as of the group’s own mission;
- Appraisal of the state-of-the-art in the discipline or academic area concerned“.

In den Niederlanden wurde das gesamte Spektrum der Hochschulforschung – ähnlich wie in Großbritannien – in 28 Disziplinen unterteilt. Für jede Disziplin wurde ein international besetztes *Review Committee* gebildet. Im Unterschied zu den britischen *Research Assessment Exercises* wird die Evaluation nicht zugleich in allen Disziplinen durchgeführt, sondern gestaffelt über einen Zeitraum von sechs Jahren.

Im niederländischen Evaluationsmodell sind nicht die Universitätsinstitute – wie bei der *Research Assessment Exercise* in Großbritannien – die kleinste institutionelle Einheit die evaluiert wird, sondern die Forschungsprogramme (*research programmes*). Die Forschung jedes einzelnen Fachbereichs untergliedert sich in mehrere Forschungsprogramme, wobei die Forschungsprogramme auch fachbereichsübergreifend angelegt sein können. Die Strukturierung der universitären Forschung in Forschungsprogramme wurde in den Niederlanden in den 80er Jahren eingeführt.

Die *Review Committees* erhalten von jedem Forschungsprogramm eine Fülle von Informationen über die Forschungsleistungen der letzten fünf Jahre sowie einen Entwicklungsplan. Dazu gehören: die fünf besten Publikationen, eine Publikationsliste, Angaben zum wissenschaftlichen Personal sowie ein *Research Mission Statement*.

Die Beurteilung der Qualität der einzelnen Forschungsprogramme erfolgt entlang vier verschiedener Dimensionen, wobei die Gutachter einen internationalen Maßstab anlegen sollen:

- Wissenschaftliche Qualität (*the quality of the academic activities*): Damit ist Qualität im klassischen Sinne gemeint. Dazu gehören die Originalität der Ideen und der Methodik, die Signifikanz des Forschungs-Outputs für die Leistung der Disziplin sowie die interne Kohärenz des Forschungsprogrammes und der wissenschaftliche *Impact* der Forschung;
- Wissenschaftliche Produktivität (*the quantity of the academic output*): Im wesentlichen zielt diese Dimension auf die Gegenüberstellung von Input und Output. Dabei gilt als wichtigstes Output-Maß die Gesamtzahl der Publikationen, wobei diese nach verschiedenen Qualitätskategorien differenziert werden. Von großer Bedeutung sind Publikationen in internationalen Fachzeitschriften mit hohen Impact-Faktoren. Ferner sind die Anzahl der abgeschlossenen Dissertationen, Patentanmeldungen und die Entwicklung von Software, Prototypen und Testverfahren von Bedeutung. Die Input-Maße beziehen sich im wesentlichen auf die monetären Ressourcen und auf das Personal, welche in die Forschung ‚investiert‘ werden;
- Wissenschaftliche und gesellschaftliche Relevanz (*the academic and the social relevance*): Diese Dimension betrifft die Frage, ob die erzielten Forschungsergebnisse von anderen Forschern im Fach rezipiert und für die Entwicklung des Fachs als wichtig angesehen werden, sowie die Frage, ob die Ergebnisse zu Anwendungen geführt hat, die für die Gesellschaft oder für die Technologieentwicklung nützlich waren;

- Perspektive des Forschungsprogramms (*the academic perspective*): Beurteilt werden soll das Potenzial des Forschungsprogramms vor dem Hintergrund nationaler und internationaler Entwicklungen.

Jedes Forschungsprogramm wird entlang dieser vier Dimensionen auf einer fünfstufigen Notenskala beurteilt. Die Ausprägungen dieser Notenskala sind 1 (*poor*), 2 (*unsatisfactory*), 3 (*satisfactory/average*), 4 (*good*), 5 (*excellent*).

Alles in allem stellt das niederländische Evaluationssystem eine Kombination aus Selbstevaluation und externer Evaluation dar. Die Forschung der universitären Forschungsprogramme wird dezidiert aus einer internationalen Perspektive und unter Anlegung eines internationalen Qualitätsmaßstabes beurteilt. Im Gegensatz zu Großbritannien wird jedoch keine Verbindung zwischen Evaluationsergebnis und Mittelallokation hergestellt. Die Evaluation der Forschungsprogramme hat nach Meinung niederländischer Experten das Niveau der Forschung an den Universitäten insgesamt erhöht und die Selbststeuerungs-Fähigkeiten der Universitäten und die universitätsinternen Entscheidungsprozesse wurden allgemein professionalisiert.

4.2.7 Europäischen Union

Die Europäische Union fördert Forschung und Technologie durch das *RTD Rahmenprogramm* und durch das *Euratom Rahmenprogramm*. Die Europäische Kommission führt darüber hinaus am *Joint Research Centre* (das in sieben Institute unterteilt ist) selbst Forschungsprojekte durch. Im Rahmen des *Vierten Rahmenprogramms* (1994-1998) förderte die EU mit 13 Mrd. Euro 18 verschiedene Programme. Durch die Programme wurden etwa 10.000 Forschungsprojekte unterstützt, an denen etwa 30.000 Wissenschaftler beteiligt waren (vgl. European Commission, 1997).

Der erste Aktionsplan für die Evaluation der Forschungsprogramme der EU wurde 1983 implementiert. Von 1989 bis 1992 förderte die Europäische Kommission im Rahmen der SPEAR-Komponente des MONITOR-Programms die Entwicklung von Evaluationsmethoden. Bis 1994 wurden mehr als 70 Programme evaluiert; an den Evaluationen waren über 500 Experten beteiligt (vgl. Fayl et al., 1998).

1995 wurde ein neues Evaluationssystem eingeführt, das ex ante-Evaluation, fortlaufende Beobachtung der Implementierung und des Fortschritts der Programme (Monitoring), mid-term-Evaluation und ex post-Evaluation in neuartiger Weise zu integrieren versucht (vgl. Dumont et al., 1999). Zu den laufenden Einzelprogrammen (mit einer Laufzeit von fünf Jahren) erstellen unabhängige Experten jährliche Monitoring-Berichte. Die alle fünf Jahre stattfindende Evaluation der einzelnen Programme (Assessment) durch externe Experten ist so terminiert, dass sie die letzten drei Jahre des vorausgegangenen Programms (ex post-Evaluation), die ersten zwei Jahre des laufenden Programms (mid-term appraisal) und die Vorbereitungsphase für das Nachfolgeprogramm (strategische ex ante-Evaluation) umfasst. Alle Einzelprogramme werden auf diese Weise beobachtet und jeweils zum selben Zeitpunkt evaluiert. Auf diese Weise stehen die Evaluationsergebnisse im Idealfall rechtzeitig als Input für die Entwicklung des nächsten Rahmenprogramms zur Verfügung (vgl. Durieux & Fayl, 1997).

Der Begleitung der Einzelprogramme kommt dabei folgende Aufgabe zu: „the monitoring examines progress in relation to the original objectives and whether these objectives, priorities and financial resources are still appropriate to changing circumstances“. Die fortlaufende Begleitung der Programme soll darüber hinaus sicherstellen, dass bereits während der Programmdurchführung jene Daten erhoben werden, die später für die Fünfjahrevaluation (five-year assessment) benötigt werden.

Die Fünfjahresevaluation der einzelnen Programme erstreckt sich vor allem auf folgende Aspekte: „the relevance of the initial objectives in the face of major new developments (taking into account current S&T state of the art and economic and societal needs in relation to S&T), the cost-effectiveness of programme implementation, and its effectiveness in achieving its original objectives. The exercise also results in the identification of major achievements and lessons learned from programme implementation, and provides recommendations for future activities“ (Durieux & Fayl, 1997, S. 374 f.).

Die jährlichen Monitoring-Berichte bilden die Grundlage für die alle fünf Jahre stattfindende Evaluation der Einzelprogramme. Die Evaluationsberichte zu den einzelnen Programmen stellen die Basisinformationen für die Evaluation des gesamten Rahmenprogramms zur Verfügung. Für den Fall, dass alle Berichte rechtzeitig vorliegen, entsteht auf diese Weise eine Evaluationspyramide.

Den unabhängigen und heterogen zusammengesetzten Evaluationskommissionen gehören Mitglieder aus Universitäten, staatlich finanzierten außeruniversitären Forschungseinrichtungen und aus der Industrieforschung an. Die Mitglieder der Evaluationskommissionen interviewen in aller Regel Repräsentanten aller Stakeholder-Gruppen und berücksichtigen Ergebnisse aus Umfragen unter Programmteilnehmern und abgelehnten Antragstellern sowie Leistungs- und finanzstatistische Kennzahlen.

Eine Analyse von 61 Evaluationsberichten, die im Zeitraum von 1980 bis 1995 erstellt wurden, zeigt, dass sich die Evaluation von EU-Programmen in der Vergangenheit sehr stark am traditionellen Modell des *Peer Review* orientiert hat (Cunningham, 1997). In den Evaluationsberichten wird relativ häufig Bezug genommen auf Interview- und Fragebogendaten; nur vergleichsweise selten werden finanzstatistische Kennzahlen und bibliometrische Daten (z. B. Publikationen, Patente, internationale Ko-Autorenschaften) herangezogen.

Gegenwärtig exploriert das *European Technology Assessment Network* (ETAN) der Europäischen Kommission, welche Möglichkeiten es gibt, die wissenschaftlichen, die sozio-ökonomischen und die *Policy*-Effekte der europäischen RTD-Programme besser nachzuweisen. Die ETAN-Arbeitsgruppe orientiert sich dabei an folgendem Schema (Airaghi, 1999, S.8):

Table: Impact dimensions of public RTD spending

Main domains of impact of public RTD spending	Direct impacts		Indirect Impacts	
	Short-term	Long-term	Short-term	Long-term
Science („Wissenschaft“)	scientific findings	knowledge	improved teaching	industrial spill-overs
Economy and society	improved technology	improved technical know-how	increased productivity	improved competitiveness
Policy	improved understanding	problem-solving	increased problem awareness	increased general satisfaction

4.3 Aktueller Stand und Perspektiven der Programm-Evaluation

Die internationale Erfahrung zeigt, dass es kein generell empfehlenswertes „ideales“ Evaluationskonzept gibt. Die Ziele, die Forschungsfelder, die Förderinstrumente und die Wirkungen von Forschungsprogrammen sind zu vielfältig, als dass sie mit einem Einheitskonzept bewertet werden könnten. Wissenschaftliche Forschung und Technologieentwicklung finden heute in turbulenten Feldern statt. In diesem Umfeld sollte die Evaluation von Programmen nicht als buchhalterische Übung absolviert werden. Programmziele sollten auch während einer laufenden Maßnahme modifizierbar bleiben (moving targets), auch wenn dies die Evaluation des Zielerreichungsgrades von Programmen erschwert.

Programmevaluation scheint sich international immer mehr zu einem Planungsmittel zu entwickeln, das vorausschauende (ex ante) Analysen, begleitende (monitoring) und rückblickende (ex post) Evaluationen umfasst. Dabei wird eine systematische Verknüpfung von ex ante-, begleitender und ex post-Evaluation angestrebt. Im Idealfall entsteht auf diese Weise ein rollendes System der Informationsversorgung. Ex ante-Evaluationen werden in Zukunft stärker als bisher auch Ergebnisse von Delphi-Befragungen und anderen Foresight-Initiativen sowie von Technologiefolgeabschätzungen berücksichtigen (vgl. Kuhlmann, 1999; Kuhlmann et al., 1999; Martin, 1996); die begleitende Evaluation wird verstärkt Coaching-Funktionen beinhalten und für die ex ante-Evaluation wird ein programmspezifischer Methoden-Mix zur Anwendung kommen, mit dessen Hilfe neben den wissenschaftlichen auch die technologischen, ökonomischen, sozialen und ökologischen Auswirkungen von Forschungsprogrammen abgeschätzt werden können. Neben *Peer Reviews* (z. B. in Form von Abschlusskolloquien und Statusseminaren) werden vermutlich auch immer öfter bibliometrische Daten und im Bereich wissenschaftsbasierter technologischer Innovationen auch Statistiken über Patente, Firmengründungen (*spin-off companies*) und Nutzern herangezogen werden.

5. Evaluation der Forschung in einzelnen Disziplinen und Technologiebereichen

In den nordischen Ländern (Dänemark, Finnland, Norwegen und Schweden) hat die Evaluation der Forschung in einzelnen Disziplinen eine lange Tradition. Nach Luukkonen (1999) fand in den 80er Jahren „a diffusion of evaluation models from one Nordic country to another“ statt. In Finnland beispielsweise wurde 1982 mit der Evaluation von Disziplinen begonnen. Das Verfahren für die Evaluation von Disziplinen orientierte sich am Modell des *Swedish Natural Science Research Council*. In Deutschland hat der Wissenschaftsrat in den letzten Jahren mit der fach- und institutionenübergreifenden Analyse und Bewertung großer Forschungsfelder Neuland beschritten. In den USA hat der *National Research Council* 1997 damit begonnen, den Rang der US-amerikanischen Forschung in ausgewählten Forschungsfeldern durch internationales Benchmarking zu bestimmen.

5.1 Finnland

In Finnland hat die Academy of Science seit 1982 etwa 25 Disziplinen evaluiert. Hierbei wurde ein internationaler Bewertungsmaßstab angelegt. Die Evaluation bezog sich – im Unterschied zum schwedischen Modell - jeweils auf die gesamte Forschung in dem jeweiligen Fach in Finnland: „The idea of evaluating research and the method used were borrowed from the practices of the Swedish Natural Science Research Council, although the method was applied in a slightly different way, i. e. to whole areas of research irrespective of whether or not they were funded by the Academy“ (Luukkonen, 1997).

Das Evaluationsverfahren soll im folgenden am Beispiel der elektrotechnischen Forschung in Finnland dargestellt werden (vgl. Academy of Finland, 1997):

Im September 1995 beauftragte die Leitung der Akademie den zuständigen Forschungsrat (in diesem Fall den *Research Council of Natural Sciences and Engineering*) mit der Evaluation der elektrotechnischen Forschung in Finnland. Es wurde eine Kommission eingerichtet, die wiederum zwei ausländische Experten bestimmte, die für die eigentliche Evaluation des Faches zuständig waren.

Das Forschungsfeld Elektrotechnik wurde auf einige Teildisziplinen eingegrenzt, auf die sich die Evaluation konzentrierte. Innerhalb dieser Teildisziplinen wurden 28 universitäre und außeruniversitäre Forschergruppen ausgewählt, deren Forschung evaluiert wurde.

Die Forschungstätigkeit jeder Forschungsgruppe wurde hinsichtlich der folgenden sechs Kriterien bewertet:

- die ‚Mission‘, Ziele und Forschungsplanung der Gruppe;
- die der Forschergruppe zur Verfügung stehenden Ressourcen, sowie die Effizienz deren Nutzung;
- die wissenschaftliche Kompetenz der Forschergruppe sowie deren Innovationsgrad;
- die technologische Kompetenz der Forschergruppe sowie das Ausmaß der Kooperation mit anderen Forschergruppen und wissenschaftlichen Einrichtungen, mit der Industrie und mit den Anwendern (*appliers*) der Forschungsergebnisse;
- die nationale und internationale Bedeutung der Forschergruppe und ihrer Forschungsergebnisse für die *scientific community* sowie allgemein für die weitere Qualifikation von Wissenschaftlern;
- die Relevanz der Forschergruppe und ihrer Forschungsleistungen für die Industrie.

Auf dieser Grundlage erstellte die Evaluationskommission einen Bericht, der neben spezifischen Stellungnahmen und Empfehlungen zu einzelnen Arbeitsgruppen, eine allgemeine Analyse der Situation der elektrotechnischen Forschung in Finnland beinhaltet.

5.2 Bundesrepublik Deutschland - das Beispiel der fach- und institutionenübergreifenden Analyse und Bewertung großer Forschungsfelder durch den *Wissenschaftsrat*

Neues Wissen entsteht immer weniger im Kontext einer einzelnen Disziplin. Vielmehr bilden sich auf der Suche nach Problemlösungen neue transdisziplinäre Forschungszusammenhänge heraus, wie beispielsweise die Umwelt-, Material- oder Energieforschung. Für die Beurteilung der Qualität und des Erfolgs der Forschung in diesen Feldern kommen zusätzliche Kriterien sozialer, ökonomischer und politischer Art ins Spiel. „Gute Wissenschaft“ zu bestimmen ist in diesen Bereichen komplexer und schwieriger.

Seit 1994 hat der *Wissenschaftsrat* drei derartige Forschungsfelder evaluiert. Erstmals wurden über die Grenzen der klassischen Fachdisziplinen und der Sektoren der öffentlich finanzierten Forschungseinrichtungen hinweg Qualität und Perspektiven der Forschung bewertet (vgl. *Wissenschaftsrat*, 1994, 1999c, d). In den Studien zur Umwelt-, Material- und Energieforschung wurden verschiedene Instrumente des Vorgehens erprobt und jeweils Handlungsoptionen für die Ausrichtung und Koordination der Forschungsförderung erarbeitet. In dieser fach- und institutionenübergreifenden Analyse und Bewertung großer Forschungsfelder sieht der *Wissenschaftsrat* eine wichtige Aufgabe zum Erhalten und zur Steigerung der Leistungsfähigkeit der deutschen Forschung. Deswegen wird er die Erfahrungen mit den abgeschlossenen Querschnittsbewertungen hinsichtlich des methodischen Vorgehens und ihrer Bedeutung für das deutsche Wissenschaftssystem auf der Grundlage eines Vergleichs mit ähnlichen Initiativen im Ausland analysieren und darauf aufbauend Planungen für weitere Stellungnahmen ausarbeiten.

5.3 USA - das Beispiel des Experiments in International Benchmarking of US Research Fields durch den *National Research Council*

Im Zusammenhang mit dem *Government Performance and Results Act* hat der *National Research Council* vorgeschlagen, die Qualität der amerikanischen Forschung vor allem daran zu messen, ob sie weltweit führend ist (*International Leadership*). Um die internationale Position der USA in einzelnen Forschungs- und Technologiefeldern bestimmen zu können, hat das *Committee on Science, Engineering, and Public Policy* (COSEPUP) der *National Academy of Sciences*, der *National Academy of Engineering* und des *Institute of Medicine* drei Pilotstudien für die Fächer Mathematik (vgl. COSEPUP, 1997), Materialforschung (vgl. COSEPUP, 1998) und Immunologie (vgl. COSEPUP, 1999) durchgeführt. In dem Bericht „Experiments in International Benchmarking of US Research Fields“ hat COSEPUP (2000) das Verfahren und seine Erfahrungen mit dem Benchmarking-Ansatz zusammengefasst.

Die von COSEPUP eingesetzten Expertengruppen verwendeten jeweils verschiedene Methoden, um ein Forschungsfeld zu beurteilen:

- „The ‚virtual congress‘ (each panel asked leading experts in the field to identify the ‚best of the best‘ researchers for particular subfields, anywhere in the world);
- Citation analysis;
- Journal-publication analysis;

- Quantitative data analysis (for example, number of graduate students, degrees, and employment status);
- Prize analysis;
- International-congress speakers.“

COSEPUP (2000) beurteilt den Benchmarking-Ansatz ausgesprochen positiv und gibt folgende Empfehlungen für den Einsatz und die Weiterentwicklung der Methode:

- „Because of the panel’s use of expert judgment, the choice of panelists is a key to the credibility of the results. A tendency toward national biases can be mitigated by ensuring diverse geographical membership of panels; the same is true of the groups that select the panel members. In particular, it is critical to include non-US participants in the selection of panelists and as panel members because they provide perspective and objectivity;
- Because major fields of research change slowly, benchmarking can probably detect important changes in quality, relevance, and leadership in fields when conducted at intervals of 3-5 years. It is unlikely that changes can be detected by annual benchmarking;
- The choice of research fields to be evaluated is both challenging and critical. A ‚field‘ might best be considered the array of related domains between which investigators can move without leaving their primary area of expertise;
- Benchmarking produces information that administrators, policy-makers, and funding agencies find useful as they make decisions as to what activities a federal research program should undertake and respond to demands for accountability, such as the Government Performance and Results Act;
- If federal agencies use benchmarking, the wide variation in agency missions dictates that each agency tailor the technique to its own needs;
- Use of indicators that provide information on degree of uncertainty and reliability of benchmarking results might enhance the presentation of panel assessments of leadership status“;

Internationales Benchmarking wird von COSEPUP als effizientes und reliables Evaluationsverfahren empfohlen, das nicht nur zu bestimmen erlaubt, welchen Rang ein Land in einem Forschungsfeld einnimmt, sondern auch Hinweise liefert, wie die Situation der Forschung in einem Fach verbessert werden kann.

5.4 Aktueller Stand und Perspektiven der Evaluation von Forschungsfeldern

Um die Herausforderungen einer globalisierten Forschungslandschaft und weltweiter Technologiemärkte aufnehmen zu können, haben viele Länder die auf Vannevar Bush (1945/1990) zurückgehende Vorstellung aufgegeben, es existiere eine lineare Entwicklung von der Grundlagenforschung bis zur Anwendung. Längst hat sich in der Wissenschaftsforschung die Erkenntnis durchgesetzt, dass in allen Stufen des Innovationsprozesses das Zusammenwirken mit der Grundlagenforschung notwendig sein kann und daher möglich sein muss (vgl. Stokes, 1997). Viele Regierungen tragen den zu beobachtenden Veränderungen im Verhältnis von Grundlagenforschung, anwendungsorientierter Forschung und Technologietransfer durch die Bildung innovativer Netzwerke von Unternehmen und Forschungseinrichtungen Rechnung. Die Evaluation derartiger Netzwerke und Kompetenzzentren hat in den letzten Jahren weltweit stark zugenommen (vgl. z. B. Hicks, 1993; Hicks & Katz, 1996; Schmoch, 1997; Turpin et al., 1999; Webster, 1994).

In diesem Zusammenhang müssen in Zukunft vermehrt quantitative Indikatoren für die Schnittstelle von Forschung und Technologie entwickelt werden. Das Augenmerk richtet sich auf Patente aus dem universitären Forschungsbereich (vgl. Henderson et al., 1998; Mansfield, 1995; Patent and Trademark Office, 1998), auf Publikationen von Industrieforschern in wissenschaftlichen Fachzeitschriften (vgl. Godin, Barker & Landry, 1995) sowie auf Verweise auf wissenschaftliche Literatur in Patentdokumenten (vgl. Kroll, Ault & Narin, 1998; Narin, 1994; Narin, Hamilton & Olivastro, 1997).

Um die Herausforderungen einer globalisierten Forschungslandschaft und weltweiter Technologiemarkte aufnehmen zu können, müssen sich Forschungseinrichtungen international stärker vernetzen (Lipsett & Holbrook, 1996). Umfang und Erfolg der Internationalisierungsbemühungen werden zunehmend an Indikatoren wie internationale Ko-Autorenschaft (vgl. Melin, 1996; Persson et al., 1997), Zahl der Forschungsstätten im Ausland sowie Höhe der Forschungs- und Entwicklungsaufwendungen im Ausland abgelesen (nach einer Studie des US Handelsministeriums unterhielten 1997 beispielsweise schweizerische Unternehmen 42 Forschungsstätten in den USA; die Forschungsaufwendungen beliefen sich 1993 auf 2423 und 1997 auf 3382 Millionen US Dollar).

Mit anderen Worten: Die Evaluation transdisziplinärer Forschungs- und Technologiefelder wird in Zukunft stark an Bedeutung gewinnen. Dabei werden Linkageindikatoren beispielsweise für die Vernetzung der Forschungssektoren eines Landes, für den Internationalisierungsgrad der Wissenschaft oder für den Grad der Wissenschaftsabhängigkeit neuer Technologien verstärkt zum Einsatz kommen werden.

6. Schlussfolgerungen und Empfehlungen für die Wissenschaftsevaluation

Evaluation sollte nach Krull (1996) stets als ein *wissenschaftspolitisches Steuerungsinstrument* verstanden und angewendet werden. Sie darf kein *l'art pour l'art* bleiben, sich nicht auf ein bloßes Ranking von Ländern und Wissenschaftseinrichtungen oder auf die Entwicklung technisch immer sauberer Indikatorensysteme und Messmethoden beschränken. Es ist vielmehr ein verbindlicher Handlungsrahmen erforderlich, bei dem alle Beteiligten über das Vorgehen und die angestrebten Folgewirkungen grundsätzlich Konsens haben sollten. Modelle aus Nordeuropa aber auch der *Max-Planck-Gesellschaft* sind hierfür vorbildhaft. Sie basieren allesamt auf dem Prinzip, nach Möglichkeit die jeweils weltweit führenden Wissenschaftler für eine Evaluation zu gewinnen und die Ergebnisse zur Grundlage für künftige Ressourcen-Allokation zu machen.

Eine klare *Zieldefinition* ist für den Erfolg einer Evaluation von entscheidender Bedeutung. Dabei ist zwischen *ex ante*-, *mid-term*- und *ex post*-Evaluation zu unterscheiden sowie insbesondere darauf zu achten, dass eine Evaluation auf ihrer jeweiligen Ebene in das Netzwerk der verschiedenen anderen Qualitätsbewertungen einzupassen ist, um Evaluationsredundanz und Evaluationsmüdigkeit zu vermeiden. Der Evaluation nationaler Forschungssysteme sollte beispielsweise die Evaluation der Forschungsorgane zugrunde liegen, deren Evaluation wiederum baut auf der Evaluation der Förderprogramme und Forschungseinrichtungen auf, deren Evaluation schließlich auf der Bewertung einzelner Projekte und Projektverbünde beruhen sollte (Evaluationspyramide).

Die *organisatorischen Rahmenbedingungen* müssen nicht nur die Durchführung der Evaluation, sondern auch die Implementation ihrer Ergebnisse gewährleisten. Das Ineinandergreifen von internen und externen Bewertungsverfahren bedarf sorgfältiger Abstimmung. Es gilt, die Glaubwürdigkeit und methodische Adäquatheit zu sichern, die Unabhängigkeit der Bewertung zu gewährleisten sowie die Durchsetzungsfähigkeit der Empfehlungen zu sichern.

Verfahren, Methoden und Techniken sind nicht universell anwendbar, sondern müssen für den jeweiligen Fall adaptiert werden. Quantitative und qualitative Verfahren ergänzen einander. Weder *peer review*-basierte Bewertungen noch lediglich auf Indikatoren basierende Rangreihungen sind letztlich auf der Aggregationsebene von Institutionen, Programmen und Ländern tragfähig. Das A und O einer jeden Evaluation sind Vergleiche; ein komparatives Vorgehen („Benchmarking“) ist von entscheidender Bedeutung, um für Stärken oder auch für Schwächen die Augen zu öffnen.

Jede Bestandsaufnahme und Bewertung muss mit klaren *Schlussfolgerungen bzw. Empfehlungen* verbunden werden, um handlungsrelevant werden zu können. Letztlich ist Evaluation Teil eines konsens-orientierten Prozesses, bei dem die Evaluierten und die Evaluierenden im Interesse einer Leistungserhaltung und/oder Leistungssteigerung gemeinsam nach adäquaten Mitteln und Wegen suchen sollten.

Ein *Berichtssystem* mit klaren Adressatenbezügen ist für die Wirksamkeit von Evaluationen entscheidend. Dabei stellt sich immer wieder die Frage, ob einer vertraulichen Behandlung oder der Veröffentlichung von Evaluationsberichten der Vorzug gegeben werden sollte. Ferner ist zu überlegen, ob die jeweilige Expertengruppe direkte oder indirekte Berichtsmöglichkeiten gegenüber den Entscheidungsträgern haben sollte. Die Filterfunktion von übergeordneten Gremien ist oftmals hilfreich. Sie führt aber auch gelegentlich zu einer Verwässerung der Ergebnisse. In jedem Fall ist es wichtig, die verschiedenen Adressaten eines Evaluationsberichts um schriftliche Stellungnahmen zu bitten.

Die *Implementation und Überwachung von Folge-Aktivitäten* ist besonders wichtig, um strukturelle Veränderungen und Re-Allokationen von Ressourcen durchzusetzen. Die Rückkopplung zu den Expertengruppen und zum Garantor - d. h. zu einer Steuerungsinstanz, die das Vertrauen aller Beteiligten genießt - durch Berichte über den Stand der Umsetzung von Evaluationsempfehlungen hat sich in vielen Fällen als hilfreich erwiesen.

Ganz entscheidend für die Zukunft der Wissenschaftsevaluation wird sein, ob es gelingen wird, Evaluationen nicht nur methodisch adäquat durchzuführen, sondern auch *wirkungsvoll umzusetzen*. Dies ist von weitaus größerer Bedeutung als das akribische Bemühen um technisch saubere Indikatorensysteme und als eine immer weitergehende Verfeinerung der Messmethoden.

Literatur

- A** Academy of Finland (1997). Evaluation of Electronics Research in Finland. Helsinki: Edita.
- Adams, J., Bailey, T., Jackson, L., Scott, P., Small, H. & Pendlebury, D. (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).
- Airaghi, A., Busch, N. E., Georghiou, L., Kuhlmann, S., Ledoux, M. J., Raan, A. F. J. van & Viana Baptista, J. (1999). Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Brussels: Commission of the European Communities, ETAN.
- Alston, J. M., Norton, G. W. & Pardey, P. G. (1998). Science under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting. New York: CAB International.
- 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 (1993a). Program Evaluation in the Federal Government. The Case for Program Evaluation. Treasury Board of Canada.
- Auditor General (1993b). The Program Evaluation System Making it Work. Treasury Board of Canada.
- B** Barber, J. M. (1999). Creating an Anglo-Saxon Innovation Culture. In: S. Bühler & S. 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.
- Barblan, A. (1998). The CRE Perspective on Quality Assurance. In J. P. Scheele, P. A. M. Maassen & D. F. Westerhijden (Eds.), To be continued ... Follow-up of Quality Assurance in Higher Education (S. 64-70). Maarssen: Elsevier/De Tijdstroom.
- Barré, R. (1999). French Culture. In: S. Bühler & S. 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.
- Blume, S., Heløe, L. A., Larsen, P. O. & Posner, M. V. (1993). The Academy of Finland – An International Evaluation 1992. Helsinki: Ministry of Education.
- Braun, D. (1997). Die politische Steuerung der Wissenschaft – Ein Beitrag zum »kooperativen Staat«. Frankfurt am Main: Campus.
- Braun, T., Glänzel, W. & Schubert, A. (1989). Assessing Assessments of British Science: Some Facts and Figures to Accept or Decline. *Scientometrics*, 15, 165-170.
- Brennan, M., Burke, K. C., Forsén, S., Stadtman, T. C. & Gidefeldt, L. (1998). Review of the Organisation and Working Methods of the NFR. Stockholm: Naturvetenskapliga Forskningsrådet (NRF)/Swedish Natural Science Research Council.
- Brinkman, H. J., Bladel, J. van, Schaper, W. K. A. & Dommele, J. van (1997). Externe evaluatie Koninklijke Nederlandse Akademie van Wetenschappen. Amsterdam: KNAW.
- British Council (1998). „Research Assessment Has Damaged British Science“ – A Debate. Conference Report. Köln: The British Council.
- Buchholz, K. (1995). Criteria for the Analysis of Scientific Quality. *Scientometrics*, 32(2), 195-218.
- 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.
- Bush, V. (1945/1990). Science – The Endless Frontier: A Report to the President of a Program for Postwar Scientific Research. Washington, DC: National Science Foundation.

- C Cartter, A. (1966). *An Assessment of Quality in Graduate Education*. Washington, DC: American Council on Education.
- Cashin, W. E. (1988). *Student Ratings of Teaching. A Summary of the Research*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Centre National de la Recherche Scientifique (1995). *Audit du Comité national de la recherche scientifique*. Paris: CNRS éditions.
- Chabbal, R. (1987). *Organisation of the Research Evaluation in the Commission of the European Communities. Report to the European Commission, EUR 11545*. Luxembourg: Office for Official Publications of the European Communities.
- Chevallier, Thierry (1995). *Quality Assessment in the French Higher Education System*. Dijon: Institute for the Management of Education, Université de Bourgogne.
- Chubin, D. E. (1987b). *Designing Research Program Evaluations: A Science Studies Approach*. *Science and Public Policy*, 14, 82-90.
- Comité national d'évaluation (1989). *Priorités pour l'Université. Rapport de fin de mandat*. Paris: La Documentation Française.
- Comité national d'évaluation (1993). *Universités: la recherche des équilibres*. Paris: La Documentation Française.
- Comité national d'évaluation de la recherche (1993a). *L'Evaluation de la recherche: réflexions et pratiques*. Paris: La Documentation Française.
- Comité national d'évaluation de la recherche (1993b). *Un autre regard sur la recherche – Sept évaluations 1990-1993*. Paris: La Documentation Française.
- Comité national d'évaluation de la recherche (1994). *Réflexions sur l'appareil national de recherche et de développement 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*.
- 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.
- Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1999a). *International Benchmarking of US Immunology 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 (1999b). *Evaluating Federal Research Programs - Research and the Government Performance and Results Act*. 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 (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.
- Cowen, R. (Ed.). (1996). *World Yearbook of Education 1996: The Evaluation of Higher Education Systems*. London: Kogan Page.
- Cozzens, S. E. (1987). *Expert Review in Evaluating Programs*. *Science and Public Policy*, 14, 71-81.
- Cozzens, S. E. (1995). *U. S. Research Assessment: Recent Developments*. *Scientometrics*, 34 (3), 351-362.
- Cozzens, S. E. & Melkers, J. E. (1997). *Use and Usefulness of Performance Measurement in State Science and Technology Programs*. *Policy Studies Journal*, 25(3), 425-435.
- Cunio, K. M. (1995). *U. K. Government Departments Experience of RT&D Programme Evaluation and Methodology*. *Scientometrics*, 34(3), 375-389.
- Cunningham, P. (1997). *The Evaluation of European Programmes and the Future of Scientometrics*. *Scientometrics*, 38(1), 71-85.

- D** Daniel, H.-D. (1993). *Guardians of Science - Fairness and Reliability of Peer Review*. Weinheim/New York/Basel/Cambridge/Tokyo: VCH.
- Daniel, H.-D. & Fisch, R. (1986). *Messung von Forschungsleistungen - Eine annotierte Bibliographie (1910 - 1985) und Synopsis*. Erlangen: Institut für Gesellschaft und Wissenschaft.
- Der Bundesminister für Forschung und Technologie (1988). *Voraussetzungen für die Tätigkeit von Großforschungseinrichtungen und Kriterien für eine Erfolgskontrolle*. Bonn.
- Diodato, V. (1994). *Dictionary of Bibliometrics*. New York: Haworth Press.
- 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: S. Bühler & S. 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.
- E** 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 Three: Evaluating Research Centres*. London: ESRC, Policy and Evaluation Division.
- Eldon, R. E. & Devine, C. M. (1985). *Government's Research and Evaluation Gap*. *Public Relations Review*, 11, 47-56.
- 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.
- F** Fayl, G., Dumont, Y., Durieux, L., Karatzas, I. & O'Sullivan, L. (1998). *Evaluation of Research and Technological Development Programmes: A Tool for Policy Design*. *Research Evaluation*, 7(2), 93-97.
- Feller, I. (1988). *Evaluating State Advanced Technology Programs*. *Evaluation Review*, 12, 232-252.
- Fraunhofer-Gesellschaft (Hrsg.). (1998). *Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission*. München.
- G** Garelli, S. (Ed.). (1999). *World Competitiveness Yearbook 1999*. Lausanne: International Institute for Management Development (IMD).
- 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.
- Georghiou, L. (1999). *Accountability to the Public and Value-for-Money*. In: S. Bühler & S. 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, L. & Roessner, D. (1999). *Evaluating Technology Programs: Tools and Methods*. Manchester: University of Manchester, PREST, and Atlanta, GA: Georgia Institute of Technology. Unpublished Manuscript.
- Glänzel, W. (1996). *The Need for Standards in Bibliometric Research and Technology*. *Scientometrics*, 35(2), 167-176.
- Godin, B., Barker, R. S. & Landry, M. (1995). *Besides Academic Publications: Which Sectors Compete, or are there Competitors?* *Scientometrics*, 33(1), 3-12.
- Goldberger, M. L., Maher, B. A. & Ebert Flattau, P. (Eds.). (1995). *Research-Doctorate Programs in the United States - Continuity and Change*. Washington, DC: National Academy Press.
- Goodstein, D. (1996). *The Big Crunch: The End of Expansion in Science*. Washington, D. C.: George C. Marshall Institute.

- Guillaume, H. & Zegveld, W. (1995). The Technology Development Centre of Finland (TEKES) – An International Evaluation. Helsinki: Finnish Ministry of Trade and Industry, Publications 5/1995.
- Guy, K., Georghiou, L., Quintas, P., Cameron, H., Hobday, M. & Ray, T. (1991). Evaluation of the Alvey Programme for Advanced Information Technology. London: HMSO.
- Guy-Ohlson, D. (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.

H Harhoff, D. & Röller, L.-H. (1999): Entwicklungen und Perspektiven der bundesrepublikanischen Beihilfepolitik. In: M. Kaase & G. Schmid (Hrsg.), *Eine lernende Demokratie – 50 Jahre Bundesrepublik Deutschland*. WZB-Jahrbuch 1999 (461-498). Berlin: edition sigma.

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

Henderson, R., Jaffe, A. B. & Trajtenberg, M. (1998). Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988. *Review of Economics and Statistics*, 119-127.

Hicks, D. (1993). University-Industry Research Links in Japan. *Policy Sciences*, 26, 361-395.

Hicks, D. & Katz, J. S. (1996). Science Policy for a Highly Collaborative Science System. *Science and Public Policy*, 23(1), 39-44.

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

J Jaffe, A. B., Fogarty, M. & Banks, B. 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).

Jaspers, K. (1961). Die Idee der Universität. In: K. Jaspers & K. Rossmann (Hrsg.), *Die Idee der Universität* (S. 41-165). Berlin.

Johnes, J. (1996). Performance Assessment in Higher Education in Britain. *European Journal of Operational Research*, 89, 18-33.

Jones, L. V., Lindzey, G. & Coggeshall, P. E. (Eds.). (1982). *An Assessment of Research-Doctorate Programs in the United States*. Washington, DC: National Academy Press.

K Kostoff, R. N. (1998a). *Science and Technology Metrics*. Arlington, VA: Office of Naval Research.

Kostoff, R. N. (1998b). *Research Program Peer Review: Principles, Practices, Protocols*. Arlington, VA: Office of Naval Research.

Kroll, P., Ault, G. & Narin, F. (1998). Tracing the Influence of Basic Scientific Research on Biotechnology Patents. *Patent World*, 100, 38-46.

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

Krull, W. (1996). *Autonomie - Qualität - Evaluation*. Vortrag vor dem Österreichischen Universitätskuratorium in Wien.

Krull, W. (Hrsg.). (1999). *Forschungsförderung in Deutschland – Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft*. Hannover: Neue Medien.

Kuhlmann, S. (1999). Distributed Intelligence for Innovation Policy Planning: Integrating Evaluation, Foresight and Technology Assessment. In: S. Bühner & S. 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, S., Boekholt, P., Georghiou, L., Guy, K., Héraud, J.-A., Larédo, P., Lemola, T., Loveridge, D., Luukkonen, T., Polt, W., Rip, A., Sanz-Menendez, L. & Smits, R. (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.

Kuhlmann, S. & Holland, D. (1995). *Evaluation von Technologiepolitik in Deutschland*. Heidelberg: Physica-Verlag.

Krupp, H. (Hrsg.). (1990). *Technikpolitik angesichts der Umweltkatastrophe*. Heidelberg: Physica-Verlag.

- L** Larédo, P. (1997). *Evaluation in France: A Decade of Experience*. In OECD (Ed.), *Policy Evaluation in Innovation and Technology Towards Best Practices* (Chapter 24). Paris: OECD.
- Larédo, P. & Mustar, P. (1995). *France, the Guarantor Model and the Institutionalisation of Evaluation*. *Research Evaluation*, 5(1), 11-21.
- Link, A. N. (1996). *Economic Performance Measures for Evaluating Government-sponsored Research*. *Scientometrics*, 36(3), 325-342.
- Lipsett, M. S. & Holbrook, J. A. (1996). *Reflections on Indicators of International Cooperation in S&T*. Vancouver, BC: Centre for Policy Research on Science and Technology, Simon Fraser University.
- Luukkonen, T. (1999). *Finnish (Nordic) Culture*. In: S. Bühner & S. 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, T. (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.
- Luukkonen, T. (1995). *The Impacts of Research Field Evaluations on Research Practice*. *Research Policy*, 24, 349-365.
- M** Mansfield, E. (1995). *Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing*. *Review of Economics and Statistics*, 57, 55-65.
- Martin, B. R. (1996). *Technology Foresight: Capturing the Benefits from Science-related Technologies*. *Research Evaluation*, 6(2), 158-168.
- Max-Planck-Gesellschaft (1998). *Regelungen für das Fachbeiratswesen*. München: MPG.
- Max-Planck-Gesellschaft (1999). *Statusbericht für den Fachbeirat: notwendiger Inhalt*. München: MPG.
- May, R. M. (1998). *The Scientific Investments of Nations*. *Science*, 281, 49-51.
- McDonald, R. & Teather, G. (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.
- Melin, G. (1996). *The Networking University – A Study of a Swedish University Using Institutional Co-authorships as an Indicator*. *Scientometrics*, 35(1), 15-31.
- Moed, H. F. (1996). *Differences in the Construction of SCI Based Bibliometric Indicators Among Various Producers: A First Overview*. *Scientometrics*, 35(2), 177-191.
- Münzinger, F. & Daniel, H.-D. (1992). *Die Forschung in der ehemaligen DDR im Spiegel bibliometrischer Indikatoren: Möglichkeiten und Grenzen von Online-Datenbanken*. In: W. Neubauer & K.-H. Meier (Hrsg.), *Deutscher Dokumentartag 1991: Information und Dokumentation in den 90er Jahren - Neue Herausforderungen, neue Technologien* (S. 303-319). Frankfurt am Main: Deutsche Gesellschaft für Dokumentation.
- Muir, L. R. & Williams, D. (1994). *Management of R&D Program Evaluations: A Case Study of Canada's Energy R&D Program*. *Research Evaluation*, 3(2), 97-106.
- Mullins, N. C. (1987). *Evaluating Research Programs: Measurement and Data Sources*. *Science and Public Policy*, 14(2), 91-98.
- Murphy, P. S. (1995). *Benchmarking Academic Research Output in Australia*. *Assessment & Evaluation in Higher Education*, 20(1), 45-57.
- N** Narin, F. (1987). *Bibliometric Techniques in the Evaluation of Research Programs*. *Science and Public Policy*, 14(2), 99-106.
- Narin, F. (1994). *Patent Bibliometrics*. *Scientometrics*, 30(1), 147-155.

- Narin, F. & Breitzman, A. (1996). Inventive Productivity. *Research Policy*, 24, 507-519.
- Narin, F. & Hamilton, K. S. (1996). Bibliometric Performance Measures. *Scientometrics*, 36(3), 293-310.
- Narin, F., Hamilton, K. S. & Olivastro, D. (1997). The Increasing Linkage between U. S. Technology and Public Science. *Research Policy*, 26, 317-330.
- 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.
- Neave, G. (1998). The Evaluative State Reconsidered. *European Journal of Education*, 33(3), 265-284.
- Nederlandse Organisatie voor Wetenschappelijk Onderzoek (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.
- Nordic Council of Ministers (1998). Nordic Publications 1997/98. Copenhagen.
- Numminen, S. & Hämäläinen, O. (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.
- O** OECD (1997). The Evaluation of Scientific Research: Selected Experiences. Paris: OECD, Committee for Scientific and Technological Policy. Document OECD / GD(97)194.
- P** Paces, V., Pivec, L. & Teich, A. H. (1999). Science Evaluation and its Management. Ohmsha: IOS Press, NATO Science Series: Science & Technology Policy, vol. 28.
- Patent and Trademark Office (1998). U. S. Colleges and Universities – Utility Patent Grants 1969-1997. Washington, DC: U. S. Department of Commerce, Patent & Trademark Office.
- Persson, O., Melin, G., Danell, R. & Kaloudis, A. (1997). Research Collaboration at Nordic Universities. *Scientometrics*, 39(2), 209-223.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Q** 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.
- R** Rip, A. (1999). Societal Challenges for Evaluation. In: S. Bühner & S. 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.
- Roessner, D. (1999). New Approaches to Evaluating Research Programmes for Management Purposes. In V. Paces, L. Pivec & A. H. Teich (eds.), *Science Evaluation and Its Management (36-50)*. Ohmsha: IOS Press, NATO Science Series: Science & Technology Policy, vol. 28.
- Roessner, D. & Melkers, J. (1995). Evaluation of National Research and Technology Policy Programs in the United States and Canada. In: S. Kuhlmann & D. Holland (Hrsg.), *Evaluation von Technologiepolitik in Deutschland: Konzepte, Anwendung, Perspektiven (243-296)*. Heidelberg: Physica-Verlag.
- Roose, K. D. & Anderson, C. J. (1970). *A Rating of Graduate Programs*. Washington, DC: American Council on Education.
- S** Schmoch, U. (1997). Indicators and the Relations between Science and Technology. *Scientometrics*, 38(1), 103-116.
- 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).

- Science and Technology Agency (1998a). 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>).
- Science and Technology Agency (1998b). 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).
- 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).
- Stähle, B. (1987). Evaluation of Research: Nordic Experiences. Copenhagen: Nordic Science Policy Council, FPR-publication no. 5.
- Staropoli, A. (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.
- Stokes, D. E. (1997). *Pasteur's Quadrant – Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.
- T** Turpin, T., Aylward, D., Garrett-Jones, S., Speak, G., Grigg, L. & Johnston, R. (1999). *University and Industry Research Partnerships in Australia. An Evaluation of ARC/DETYA Industry-linked Research Schemes*. Canberra: Department of Education, Training and Youth Affairs, Higher Education Division.
- V** Vereniging van Universiteiten (1998). *Protocol 1998 (Series: Assessment of Research Quality)*. Utrecht: VSNU.
- Vitt, J. (1998). *Schlüsselerfinder in der industriellen Forschung und Entwicklung*. Wiesbaden: Deutscher Universitätsverlag.
- W** 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).
- Webster, A. (1994). International Evaluation of Academic-Industry Relations: Contexts and Analysis. *Science and Public Policy*, 21(2), 72-78.
- Webster, D. S. & Skinner, T. (1996). Rating PhD Programs: What the NRC Report Says ... and Doesn't Say. *Change*, 28(3), 34-44.
- 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.
- Wissenschaftliche Kommission Niedersachsen (1999). *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen - Grundzüge des Verfahrens*. Hannover.
- Wissenschaftsrat (1994). *Stellungnahme zur Umweltforschung in Deutschland (zwei Bände)*. Köln.
- Wissenschaftsrat (1999a). *Leitfaden für die Bewertung von Einrichtungen der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibnitz (WGL) (Blaue Liste)*. Drs. 3927/99. Köln.
- Wissenschaftsrat (1999b). *Fragebogen für die Bewertung der Forschungseinrichtungen und Museen der Blauen Liste*. Drs. 3926/99. Köln.
- Wissenschaftsrat (1999c). *Stellungnahme zur außeruniversitären Materialwissenschaft*. Köln.
- Wissenschaftsrat (1999d). *Stellungnahme zur Energieforschung*. Köln.
- Z** Ziman, J. (1987). *Science in a „Steady State“: The Research System in Transition*. London: Science Policy Support Group.
- Ziman, J. (1994). *Prometheus Bound - Science in a Dynamic Steady State*. Cambridge: Cambridge University Press.

Publications edited by the Center for Science & Technology Studies (CEST) can be accessed at the following site: www.cest.ch. They can be either consulted and printed out in a PDF format, or requested in hard copy form at the Science Policy Documentation Center (edith.imhof@swr.admin.ch).

It is also possible to order publications of the former Swiss Science Council which are indexed at the same address.

Die Publikationen des Zentrums für Wissenschafts- und Technologiestudien (CEST) finden sich unter www.cest.ch und können entweder als PDF-file eingesehen und ausgedruckt oder als Papierversion bei der Dokumentationsstelle für Wissenschaftspolitik (edith.imhof@swr.admin.ch) bezogen werden. Die Publikationen des ehemaligen Schweizerischen Wissenschaftsrates (SWR) und dessen Geschäftsstelle können ebenfalls unter den genannten Adressen eingesehen und bestellt werden.

On trouvera les publications du Centre d'études de la science et de la technologie (CEST) à l'adresse: www.cest.ch; elles peuvent ou bien être consultées et imprimées en format PDF ou demandées en version papier auprès du Centre de documentation de politique de la science (edith.imhof@swr.admin.ch).

Il est également possible de commander les publications de l'ancien Conseil suisse de la science (CSS), elles sont répertoriées à la même adresse.

Si possono trovare le pubblicazioni del Centro di studi sulla scienza e la tecnologia (CEST) all'indirizzo seguente: www.cest.ch. Esse sono disponibili in formate PDF, o essere ordinate in una versione scritta presso il Centro di documentazione di politica della scienza (edith.imhof@swr.admin.ch).

È inoltre possibile comandare le pubblicazioni dell'ex Consiglio Svizzero della Scienza (CSS), anch'esse repertorate allo stesso indirizzo.