

Thomas Heinze

## Evaluation von Forschungsleistungen

### Konzeptionelle Überlegungen und Situationsbeschreibung für Deutschland



Evaluation – der Schlüssel zum rundum erfolgreichen Forschungssystem?

Foto: Volker Derlath

Kenntnisse über die Leistungsfähigkeit des deutschen Wissenschafts- und Forschungssystems sind für Entscheidungsträger in der Forschungs- und Technologiepolitik, der Forschungsplanung und dem Forschungsmanagement von hohem Interesse. Die Frage nach Methoden für die Qualitätsbemessung von Forschungsleistungen gewinnt daher zunehmend an Bedeutung. Um hierzu kompetent Stellung nehmen zu können, bedarf es einerseits klarer konzeptioneller Vorstellungen darüber, was unter „Evaluation von Forschungsleistungen“ zu verstehen ist und andererseits Informationen darüber, wie die gegenwärtige Evaluationspraxis in Deutschland und Europa beschaffen ist. In der Praxis können derzeit unterschiedliche Schwerpunktsetzungen in der Forschungsevaluation beobachtet werden.

Über ein Drittel aller in Deutschland betriebenen Wissenschaft und Forschung wird von der öffentlichen Hand finanziert, knapp zwei Drittel von der Wirtschaft. Die staatlichen Forschungsmittel fließen an Universitäten und außeruniversitäre Forschungseinrichtungen in Form von überwiegend institutioneller, aber auch projekt- und programmbezogener Förderung (BMBF 2002, S. 222 ff.). Traditionell erfolgt die **Finanzierung von Forschungsprojekten** und -programmen unter der Bedingung, die wissenschaftliche Qualität durch Fachkollegen evaluieren zu lassen. Dieses Peer-Review-Verfahren genießt international hohes Ansehen. Zunehmend interessieren sich jedoch auch außerwissenschaftliche Akteure für erzielte Forschungsleistungen. Zum einen fordern Politik und Gesellschaft **Rechenschaft über Leistungen, Qualität und Nutzen** staatlich geförderter Wissenschaft und Forschung. Knappe staatliche Mittel sollen möglichst effektiv und effizient verwendet werden. Zum anderen entsteht im Zuge einer sich internationalisierenden Forschungslandschaft – man denke nur an die geplante Schaffung eines Europäischen Forschungsraumes – bei Verantwortlichen in Forschungsplanung, Forschungsmanagement und Forschungspolitik ein wachsender Bedarf an Informationen darüber, wo exzellente Forschung ihren Schwerpunkt findet. In den letzten Jahren sind deshalb Forschungsinstitutionen, übergreifende Forschungsförderungsprogramme, disziplinäre Forschung an Hochschulen und Forschungsprojekte zum Gegenstand von Forschungsevaluation geworden. Was ist jedoch unter Forschungsleistung zu verstehen? Wie wird sie definiert? Welche Verfahren und Indikatoren sind entwickelt worden, um die Leistungsfähigkeit von Forschungsanstrengungen zu bestimmen und welche Probleme ergeben sich bei ihrer Verwendung? Schließlich: Wie werden in Deutschland gegenwärtig Forschungsleistungen evaluiert? Antworten auf diese Fragen sollen hier diskutiert werden.

#### Definition „Forschungsleistung“

Unter **Forschung** ist der Prozess wissenschaftlicher Arbeit zu verstehen, der unter „Zurverfügungstellung“ bestimmter Ressourcen die **Produktion neuen Wissens und neuer Technologien** zum Ziel hat. Forschung beginnt dieser Vorstellung gemäß mit bestimmten Ressourcen (zum Beispiel Personal, Infrastruktur), diese werden verarbeitet und am Ende stehen Leistungen. Ein solches **Input-Throughput-Output-Modell** des Forschungsprozesses findet sich in unterschiedlichen Spielarten in der Literatur (Liefner 2001, S. 39 f.).

**Was Forschungsleistung materiell bedeutet, lässt sich jedoch nicht ohne Weiteres definieren. Im Gegensatz zur Wirtschaftsleistung, die über das Periodenergebnis von Unternehmen ermittelt wird, verfügt die Wissenschaft über keinen vergleichbar eindeutigen Ansatz, Leistung zu messen**

Was **Forschungsleistung materiell** bedeutet, lässt sich jedoch nicht ohne Weiteres definieren. Im Gegensatz zur Wirtschaftsleistung, die über das Periodenergebnis von Unternehmen ermittelt wird, verfügt die Wissenschaft über keinen vergleichbar eindeutigen Ansatz, Leistung zu messen (Bolsenkötter 1988, S. 41). Forschungsleistung wird aus diesem Grund anhand **unterschiedlicher Kriterien** definiert, etwa als Leistungsmenge, hinsichtlich ihrer Kreativität und Originalität, ihrer Sichtbarkeit und Anschlussfähigkeit, der durch sie erzeugten Wirkung oder der Effizienz ihrer Erstellung. Diese an sich nützlichen, jedoch in Form einer Liste unübersichtlichen Kriterien lassen sich klassifizieren. Naheliegend erscheinen die folgenden drei Dimensionspaare:

- ◆ Quantität und Qualität,
- ◆ Effizienz und Effektivität sowie
- ◆ Produktivität und Wirkung.

Das erste und zweite Dimensionspaar bezeichnen **sachliche Aspekte**. Zum einen geht es um Leistungsmenge und Leistungseigenschaften. Forschungsleistung lässt sich in diesem Sinne „abzählen“, aber auch „wiegen“. Zum zweiten wird Leistungserstellung mit Blick auf **eingesetzte Ressourceneinheiten (Effizienz)** und zum anderen hinsichtlich **definierter Zielgrößen (Effektivität)** evaluiert. Das dritte Dimensionspaar bezeichnet ein Zeitverhältnis und unterscheidet unmittelbar Produziertes von mittelbarer Wirkung. Forschungsleistungen werden differenziert nach bloßer Existenz und erzeugter Resonanz. In der Summe geben die drei Dimensionen somit Auskunft über das „Wie“ von Forschungsleistung: Wie viel? Wie gut? Wie produktiv? Wie effizient? usw.

Neben diesen Wie-Fragen erscheint es sinnvoll, den Leistungsbegriff auch über die Frage zu erschließen, für wen Forschungsleistungen von Interesse sind. Denn weiß man darum, wird ersichtlich, **für welche Zwecke und mit welchen Zielen** Forschungsleistungen überhaupt evaluiert werden. In der Literatur findet sich hierzu die Vorstellung, dass das Wissenschafts- und Forschungssystem ein mit der Produktion neuen Wissens befasstes gesellschaftliches Funktionssystem ist, dessen Leistungen von anderen gesellschaftlichen Teilsystemen, etwa von Politik und Wirtschaft, weiterverwendet werden können (Luhmann 1998). Betrachtet man Wissenschaft aus der Perspektive dieser gesellschaftlichen Umwelten, dann ist Forschungsleistung in folgender Weise relevant:

- ◆ **Wirtschaft:** Unternehmen werden sich vorzugsweise für diejenigen Forschungsleistungen interessieren, die im wirtschaftlichen Sinn verwertbar sind. Eine unter dieser Prämisse durchgeführte Evaluation wird darauf abstellen, wie hoch der Anteil wirtschaftlich anschlussfähigen Wissens an der gesamten Wissensproduktion ist und wie gut es sich für wirtschaftliche Verwertungsprozesse verwenden lässt.
- ◆ **Politik:** Akteure der Forschungs- und Technologiepolitik werden fragen, wie effizient und effektiv aus Steuergeldern finanzierte Forschung durchgeführt wird. Damit werden Forschungserträge zum einen in Relation zu den eingesetzten Ressourcen (Effizienz) und zum anderen in Relation zu politisch definierten Zielvorgaben (Effektivität) gesetzt. Politische Akteure streben auf diese Weise an, Ressourcenallokation in Forschung zu legitimieren.
- ◆ **Binnenperspektive der Wissenschaft:** Sofern Akteure des Wissenschafts- und Forschungssystems an hoher Qualität ihrer Forschungsarbeiten interessiert sind, werden sie danach fragen, wie kreatives originelles Wissen, mit dem innerwissenschaftlich definierte Probleme besser als bisher behandelt oder auch gelöst werden können, hervorgebracht wird. Der Zweck von Evaluation wird folglich darin bestehen, nach den Bedingungen produktiver Forschung zu fragen, also Forschungshypothesen aufzustellen und empirisch zu überprüfen, warum bestimmte Personen, Programme oder Einrichtungen leistungsstärker sind als andere.



**Diplom-Soziologe**  
Thomas Heinze ist wissenschaftlicher Mitarbeiter am Fraunhofer Institut für Systemtechnik und Innovationsforschung, Karlsruhe und Doktorand an der Deutschen Hochschule für Verwaltungswissenschaften, Speyer.

### Stichwörter

Evaluation

Forschung

Wissenschafts- und  
Technologieindikatorik

**summary**

**Professional research evaluation produces information about performances and impacts of the science and research system and is thus of paramount importance for decision-makers in research and technology policy, research management and science planning. The article discusses „research evaluation“ conceptually by delineating aspects of evaluation approaches, sets of indicators and common problems that accompany evaluation. Against this background the current practice of research evaluation in Germany is sketched out.**

Die skizzierten drei Perspektiven sind nicht nur im Sinne eines „interessierten Beobachters“ zu verstehen, der die Erstellung von Forschungsleistung beschaut und es dabei belässt. Vielmehr bezeichnen diese Perspektiven spezifische gesellschaftliche Interessen, die sich Durchsetzung zu verschaffen suchen werden (Kuhlmann, 1998). Im Bereich der Politik und Wirtschaft handelt es sich um (versuchte) „Außensteuerung“, zum einen durch Programme der Forschungs- und Technologiepolitik, zum anderen durch die Geltendmachung wirtschaftlicher Verwertungsorientierungen im Prozess der Wissensproduktion; im Bereich der Wissenschaft um „Selbststeuerung“ durch Verfahren der Qualitätssicherung – etwa durch Peer Review und Wissenschaftsforschung.

**Verfahrensaspekte der Evaluation**

Die empirische Wissenschaftsforschung hat zahlreiche Evaluationsverfahren entwickelt (Stockmann, 2000), die man entlang folgender drei Aspekte zusammenfassen kann:

**Peer Review und „Weite Evaluation“**

Bei der Evaluation von Forschungsleistungen unterscheidet man prinzipiell die gutachterliche Bewertung durch Fachkollegen von der Evaluation unter „weiten“ Gesichtspunkten. Beim ersten Ansatz wird Forschungsleistung unter fachwissenschaftlichen Aspekten bewertet, vorwiegend die Angemessenheit der Theorien und Methoden für die bearbeitete Forschungsfrage, die Originalität und Kreativität des Forschungsdesigns und der Neuigkeitsgrad der erzielten Resultate. Dieser Ansatz ist im Wissenschaftssystem fest institutionalisiert und hat sich als Selbststeuerungsinstrument bewährt (Daniel, 1993). Der zweite, „weite“ Ansatz ist neueren Datums und thematisiert Forschungsleistung unter Gesichtspunkten, die über den fachwissenschaftlichen Kontext hinausreichen und auf Informationsbedürfnisse der gesellschaftlichen Umwelt abstellen (OECD, 1997). Der grundsätzliche Unterschied zwischen diesen beiden Ansätzen liegt darin, dass Peer Review eine wissenschaftliche Eignungsprüfung ist, Evaluation dagegen eine mehrdimensionale Leistungsprüfung. In diesem Sinne ist Peer Review als Teilmenge von Evaluation zu verstehen.

**Ex ante- und ex post-Evaluation**

Peer Review ist aus empirischer Perspektive zumeist – nicht immer! – mit einer ex ante-Prüfung wissenschaftlicher Leistung verbunden. So werden im Normalverfahren der Deutschen Forschungsgemeinschaft (DFG) eingereichte Forschungsprojekte ex ante auf ihre wissenschaftliche Eignung hin überprüft. Eine ex post-Bewertung der produzierten Leistungen erfolgt nicht. Dagegen ist Evaluation als Leistungsprüfung typischerweise eine ex post-Bewertung, die etwa in Form von Vorher/Nachher-Vergleichen, Kontroll- und Vergleichsgruppenansätzen oder Kosten-Nutzen-Analysen entlang der schon erwähnten drei Dimensionspaare erfolgt.

**Interne und Externe Evaluation**

Ein weiterer Verfahrensaspekt betrifft die Durchführungsseite. Fertigen Forschungseinrichtungen **Selbstbeschreibungen** über ihre Forschungsarbeit an, dann ist dies ein interner Vorgang, der zwar nach außen kommuniziert werden kann, aber in der Regie der zu evaluierenden Einrichtung bleibt. **Evaluationssubjekt** und **Evaluationsobjekt** sind identisch. Wird dagegen eine Leistungsbeschreibung durch Akteure angefertigt, die zur Organisationsumwelt gehören, hat man es mit **Fremdbeschreibungen** zu tun. Werden diese Evaluationen von anderen Wissenschaftseinrichtungen in Gang gesetzt, sind solche Fremdbeschreibungen **wissenschaftsintern**, z.B. im Sinne

von empirischer Wissenschaftsforschung. Werden die Beschreibungen dagegen überwiegend von Akteuren des politisch-administrativen Systems in Gang gesetzt, hat man es mit **wissenschaftsexternen** Fremdbeschreibungen zu tun.

### Wissenschafts- und Forschungsindikatoren

Forschungsleistungen lassen sich durch eine Vielzahl von Indikatoren beschreiben (Fisch/Daniel, 1986; Daniel/Fisch, 1988; Weingart u.a., 1991; Stock, 1994; Hornborstel, 1997). Im Folgenden soll eine Auswahl solcher Wissenschafts- und Forschungsindikatoren entlang der drei erwähnten Dimensionspaare vorgestellt werden.

Zunächst umfasst Forschungsleistung **Produktivitätsaspekte**. Hierzu zählen Forschungshandeln, durch welches Wissensprodukte erzeugt werden, sowie die Ausbildung wissenschaftlichen Nachwuchses (Abbildung 1).

	Quantitative Information	Qualitative Information
Forschungshandeln	Projekte	Laufzeit Finanzierung (DFG, Ministerien, Stiftungen) Beitrag zur Profilbildung
	Forschungsaufenthalte	Regelmäßigkeit Finanzierung (Stipendien, Kooperationsverträge)
Wissensprodukte	Publikationen	Monographie/Zeitschriftenbeiträge referiert/nicht-referiert, national/international
	Patente	Technologiegebiet
Ausbildung wissenschaftlichen Nachwuchses	Promotionen	Benotung der Promotion Dauer der Promotion

Abb. 1: Indikatoren zur Messung wissenschaftlicher Produktivität (Auswahl)

Im **quantitativen Sinne** kann Forschungshandeln über die Zahl laufender Projekte und Forschungsaufenthalte bestimmt werden. Diese Indikatoren können in **qualitativer Hinsicht** näher beschrieben werden durch die Länge der Laufzeit oder die Finanzierungsquelle. So deuten längerfristig angelegte DFG-Projekte eher auf Grundlagenforschung und ministeriell finanzierte Forschungsprojekte mit geringer Laufzeit eher auf angewandte oder strategische Forschung hin. Ein solcher **Profilunterschied** wird sich möglicherweise auch in der Wissensproduktion niederschlagen, so dass im erstgenannten Fall mehr und vor allem in referierten Fachzeitschriften publiziert wird, während im letztgenannten Fall Forschungsberichte erstellt werden, die eher für den Auftraggeber als für die wissenschaftliche Gemeinschaft von Interesse sind. Allerdings besteht nicht notwendigerweise ein Zusammenhang zwischen **Projekthäufigkeit** und **Publikationshöhe**. Denkbar ist auch, dass Universitätsforscher, die keine über Drittmittel finanzierten Projekte durchführen, zahlreiche theoretische Aufsätze in international angesehenen Fachzeitschriften publizieren und eine geringe Zahl von Doktoranden beschäftigen, die in kurzer Zeit promovieren. Forschungsproduktivität ist somit kein eindimensionales Konstrukt, sondern über vielfältige Indikatoren beziehungsweise Kombinationen derselben operationalisierbar. Die Frage, welche Indikatoren Produktivität repräsentieren, ist nicht generell, sondern fallbezogen zu entscheiden. Forschungsleistung kann weiterhin über **Wirkungs-Indikatoren** ermittelt werden, die Aufschluss geben, in welchem Umfang Forschungsleistungen in der Wissenschaft selbst sichtbar und anschlussfähig sind, welche wissenschaftliche Reputation sie langfristig erzeugen und wie sie in die gesellschaftliche Umwelt ausstrahlen.

	Quantitative Information	Qualitative Information
Sichtbarkeit (in der Wissenschaft)	Zitation Rezensionen, Repliken	Nachhaltigkeit/Halbwertszeit der Zitationen in referierten/intern. Fachzeitschriften
Reputation (in der Wissenschaft)	Berufungen	Reputation der rufenden Universität
	Forschungspreise, Stipendien	Art und Status des Preises bzw. der Stipendien
Ausstrahlung (in die Gesellschaft)	Lizenzen, Patentzitate	Art der Erlösverwendung Technologiegebiet
	Gutachten, Beratungstätigkeiten	Art der auftraggebenden Einrichtung (Stiftungen, Ministerien)
	Vorträge	Art des Veranstalters (z.B. Industriekammer, Investorenkonferenzen)

Abb. 2: Indikatoren zur Messung wissenschaftlicher Wirkung (Auswahl)

Werden publizierte Schriften zitiert, in **Rezensionen** besprochen (Bücher) und in **Repliken** zum Gegenstand einer Diskussion gemacht (Aufsätze), deutet dies darauf hin, dass geleistete Forschung wissenschaftlich sichtbar und anschlussfähig ist. Die Frage, welche Qualität eine solche Sichtbarkeit hat, kann über ihre **Dauerhaftigkeit** (Halbwertszeit) und den Ort der **Sichtbarkeits-erzeugung** (referierte Fachzeitschriften, Sammelbände etc.) bestimmt werden. Werden Schriften häufig in viel gelesenen Fachzeitschriften aufgegriffen und diskutiert, erzeugt dies eine höhere Sichtbarkeit als beispielsweise in Sammelbänden mit geringer Auflage. Eine weitere Wirkung besteht in aufgebauter **Reputation**, die über erhaltene Forschungspreise, Stipendien oder ergangene Berufungen operationalisiert werden kann. Je nachdem, welche Universität ruft, welcher Preis verliehen und welches Stipendium gewährt wird, können Forscher mehr oder weniger schnell Reputation auf sich ziehen. Weiterhin kann Forschung auch hinsichtlich ihrer **Ausstrahlung** in Politik und Wirtschaft über Indikatoren erfasst werden. Fließen Patentanmeldern Lizenzeinnahmen zu, indiziert dies eine hohe Verwertbarkeit des produzierten Wissens. Wird von Unternehmens- oder Behördenseite Expertise in Form von Gutachten oder Beratungen nachgefragt, so zeugt dies auf außerwissenschaftliche Sichtbarkeit und Verwertbarkeit wissenschaftlichen Wissens hin. Abbildung 2 fasst diese drei Wirkungsaspekte zusammen.

Forschungsleistungen werden in der Evaluationspraxis auf unterschiedlichen **Aggregationsebenen** gemessen. Wie Abbildung 3 im Überblick zeigt, bezieht sich die horizontale Ebene auf Fachdisziplinen und Forschungsfelder, auf Forschungsprogramme und schließlich auf die institutionelle Landschaft, innerhalb welcher Forschung stattfindet. Sie übergreift konkrete Personen und Organisationen. Die vertikale Ebene reicht von Organisationen über Departments und Arbeitsgruppen bis hin zum Einzelwissenschaftler. Sie gibt die Tiefenschärfe an, mit welcher Leistungsindikatoren in Forschungseinrichtungen erhoben werden können. Beide Ebenen sind zumeist verschränkt. So ist das Feld Energieforschung durch den Wissenschaftsrat (1999) auf disziplinärer und organisationaler Ebene untersucht worden.

Abb. 3: Aggregations-Ebenen der Wissenschafts-indikatorik

Horizontale Ebene	Vertikale Ebene
Fachdisziplinen (z.B. Chemie, Biologie), Forschungsfelder (z.B. Energieforschung) Programme (z.B. BMBF, EU-Kommission, DFG) Institutionelle Forschungslandschaft (z.B. Systemevalationen)	Organisation (z.B. Universität) Department (z.B. Fachbereiche, Fakultäten, Abteilungen) Forschergruppen Einzelwissenschaftler

Die diskutierten Indikatoren sind für alle drei Perspektiven Wirtschaft, Wissenschaft und Politik von Interesse. Dennoch gibt es **bereichsspezifische Unterschiede**. Für die Wirtschaft werden Indikatoren zu Patenten, Lizenzen und Promotionen von besonderem Interesse sein, weil diese darstellen, wo und in welchem Umfang verwertungsrelevantes Wissen (kodifiziert als Patent oder gebunden als Humankapital) zu finden ist, das gegebenenfalls transferiert werden kann. Die empirische Wissenschaftsforschung wird insbesondere Indikatoren zum Wissensprodukt Publikation, zum Forschungshandeln sowie zur Reputationsgenese erheben, um Bedingungen produktiver Forschung zu untersuchen. Wissenschaftsmanagement und Akteure der Forschungs- und Technologiepolitik werden dagegen ein über dieses Indikatorenportfolio hinausreichendes Informationsinteresse aufweisen, das bereits unter den Stichworten Effizienz und Effektivität benannt wurde. Darunter ist zu verstehen, dass die Leistungsindikatoren nicht absolut, sondern relativ zu **Referenzgrößen** gesetzt werden. Für Effizienz bedeutet dies den Rückbezug auf spezifische Input-Größen, die in den Forschungsprozess investiert werden. Mithilfe von Input-Indikatoren ist beispielsweise ersichtlich, welcher Verteilung Produktivität und Wirkung in unterschiedlichen Altersstufen (Altersstruktur) und Wissenschaftlersegmenten (Mittelbau versus Professoren) folgen, oder ob es Zusammenhänge zwischen der Qualität der Infrastruktur und der erbrachten Forschungsleistung gibt. Durch den systematischen Rückbezug von Leistungsindikatoren auf diese Input-Indikatoren entstehen **Informationen mit Strukturwert**, die Vorbedingung für erfolgreiches Wissenschaftsmanagement, aber auch für Wissenschaftspolitik sein dürften. Abbildung 4 fasst eine Auswahl von Input-Indikatoren unter den drei Aspekten Personal, Finanzierung und Infrastruktur zusammen.

**keywords**

evaluation

research

science and technology indicators

	<b>Quantitative Information</b>	<b>Qualitative Information</b>
Personal	Anzahl der Wissenschaftler Anzahl der Nicht-Wissenschaftler	Anteil der promovierten Wissenschaftler Relation Wissenschaftler/Nicht-Wissenschaftler Relation Mittelbau/C3- und C4-Stellen Altersstruktur Disziplinenportfolio
Finanzierung	Volumen Veränderung des Volumens	Herkunft (Grundfinanzierung, Drittmittel)
Infrastruktur	Quadratmeter Fläche	Alter, baulicher Zustand
	Anzahl der Personalcomputer und Rechensysteme	technische Leistungsmerkmale

**Abb. 4: Input-Indikatoren (Auswahl)**

Eine weitere Form, Produktivitäts- und Wirkungsindikatoren systematisch auf eine Referenzkategorie rückzubeziehen, besteht in der Analyse von **Zielfunktionen** zur Produktion wissenschaftlichen Wissens. Zweck einer derartigen Evaluation besteht darin, die Effektivität, also den Grad der **Zielerreichung von Forschungshandeln** zu bestimmen. Allerdings ist es eine empirische Frage, inwiefern solche Zielfunktionen tatsächlich existieren. So gilt beispielsweise für das deutsche Forschungssystem, dass Einrichtungen der Max-Planck-Gesellschaft (MPG) und der Fraunhofer-Gesellschaft (FhG) spezifische wissenschaftliche Missionen verfolgen, die sich als Zielfunktionen interpretieren lassen. Die MPG verfolgt die Mission exzellenter, international anschlussfähiger Grundlagenforschung, die FhG die Mission angewandter und strategischer Forschung für und mit Unternehmen, Ministerien und Behörden von Bund und Ländern. Demgegenüber verfügen etwa die Leibniz-Gemeinschaft (WGL) und die deutschen Universitäten über keine spezifisch auf Forschung bezogene Zielfunktion. In diesen Fällen gehört es zur Aufgabe von Forschungsevaluation, sachgerechte Zielfunktionen zu bestimmen.

## Literatur

Bolsenkötter, H./Fisch, R. Daniel, H.-D. (Hrsg.), *Ansätze zur Erfassung und Beurteilung von Forschungsleistungen*, in: *Messung und Förderung von Forschungsleistung*, Konstanz 1986, S. 41-49.

Bommer, R./Ursprung, H. W., Spieglein, Spieglein an der Wand..., *Eine bibliometrische Erfassung des Forschungsoutputs volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz*. Universität Konstanz, Fakultät für Wirtschaftswissenschaften und Statistik, Manuskript, 1997.

Bundesministerium Bildung und Forschung (BMBF) (Hrsg.), *Faktenbericht Forschung*, Bonn 2002.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), *Evaluation der gemeinsam geförderten Forschung in Deutschland*. Thesen der BLK für die Regierungschefs des Bundes und der Länder, 2001, Online im Internet: [http://www.blk-bonn.de/evaluation\\_thesen\\_blk.htm](http://www.blk-bonn.de/evaluation_thesen_blk.htm).

Daniel, H.-D./Fisch, R. (Hrsg.), *Evaluation von Forschung*, Konstanz 1988.

Daniel, H.-D., *Guardians of Science – Fairness and Reliability of Peer Review*, Weinheim 1993.

Daniel, H.-D., *Neuere Entwicklungen und heutiger Stand der Forschungs- und Hochschulevaluation in ausgewählten Ländern*, in: *Wissenschaftsevaluation, CEST*, o. J. Bern.

Deutsche Forschungsgemeinschaft (DFG), *Bewilligungen an Hochschulen und außeruniversitäre Forschungseinrichtungen 1996 bis 1998*, Bonn 2000.

Fisch, R. (Hrsg.), *Ein Rahmenkonzept zur Evaluation universitärer Leistungen*, in: *Daniel/Fisch*, 1988, S. 13-31.

Galtung, J., *Structure, culture and intellectual style. An essay comparing saxon, teutonic, gallic and nipponic approaches*, in: *Social Science Information* 20 (1981) 6, S. 817-856.

Hornbostel, S., *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*, Opladen 1997.

Hochschulrektorenkonferenz (HRK) (Hrsg.), *Hochschulgesetzliche Regelungen zur Qualitätssicherung*, in: *Beiträge zur Hochschulpolitik* Nr. 7, Bonn 2001.

Internationale Kommission, *Forschungsförderung in Deutschland. Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft. Studie im Auftrag der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK)*, herausgegeben von der VolkswagenStiftung, Hannover 1999.

## Probleme bei der Evaluation von Forschungsleistungen

Die Evaluation von Forschungsleistungen ist kein einfaches Unterfangen, sondern stößt auf unterschiedliche Probleme. Ein grundsätzliches Problem besteht in der **Akzeptanz** der Evaluation als Verfahren schlechthin. Insbesondere über den klassischen Peer Review hinausgehende Evaluationen werden von vielen Akteuren des Wissenschafts- und Forschungssystems als Eingriff in das grundgesetzlich verbrieft Selbstbestimmungsrecht wissenschaftlicher Tätigkeit wahrgenommen. Diese **Skepsis** gegenüber Evaluation kann in offene **Ablehnung** umschlagen, wenn evaluative Informationen zur Entscheidungsfindung bei der Allokation von Forschungsressourcen herangezogen werden sollen. Es bedarf wenig Phantasie, um sich vorzustellen, dass Widerstände oder Blockaden gegen Evaluation auch deren Ergebnisse verzerren oder sogar unmöglich machen können. Sofern Evaluation ihren Zweck erfüllen soll, bedarf es der **partnerschaftlichen und respektvollen Zusammenarbeit** von Evaluatoren und zu Evaluierenden.

Ein zweites Problem, das eng mit dem ersten verwandt ist, liegt bei den **verwendeten Indikatoren**. Die vorgestellten Indikatoren (siehe Abbildungen 1, 2 und 4) lassen sich nicht ohne Weiteres in verschiedenen Fachdisziplinen und Forschungsfeldern anwenden. So bestehen zwischen Fachdisziplinen „**kulturelle**“ **Unterschiede** im Publikations- und Zitierverhalten, deren Nichtbeachtung zu **Verzerrungen der Interpretationen** führen kann. Sind für Anglisten und Historiker Monographien zentrale Publikationsformate, so für Biologen und Physiker Aufsätze in Fachzeitschriften (Daniel/Fisch, 1988, Kap. II). Ebenso gibt es nationale **Wissenschaftsstile**, die mehr oder weniger stark normieren, was gute Wissenschaft ist (Galtung, 1981) und welche nur bedingt über Indikatoren abbildbar sind. Ein bloßer zahlenmäßiger Vergleich der Publikationstätigkeit amerikanischer, deutscher oder französischer Universitäten würde ohne Einbeziehung der nationalen Eigenheiten zu kurz greifen. Evaluation von Forschungsleistungen sollte somit „intelligent“ und nicht „mechanisch“ mit dem skizzierten Set an Indikatoren umgehen, um zu sachgerechten Urteilen zu gelangen.

Ein drittes Problem liegt darin, geeignete **Evaluatoren** zu finden. Insbesondere wenn es um externe Evaluation geht, ist es nach Daniel (2001, S.13-14) unerlässlich, **intermediäre Organisationen** zu finden, die von den Akteuren des Wissenschafts- und Forschungssystems akzeptiert sind und sich fernhalten von politischer Instrumentalisierbarkeit. Weitere **Anforderungsprofile** sind ausgewiesenes Expertentum der Evaluatoren im Bereich der Wissenschaftsindikatorik und der Evaluationsverfahren (Fisch, 1988, S. 29). Beispielgebend für die dauerhafte Institutionalisierung einer solchen intermediären Organisation ist in Deutschland der Wissenschaftsrat.

## Evaluationspraxis in Deutschland

Generell gilt, dass Deutschland im Verhältnis zu Ländern wie den Niederlanden, Großbritannien und den Vereinigten Staaten in zurückhaltender Weise die „**weite**“ **Evaluation** wissenschaftlicher Leistungen **institutionalisiert** hat (Daniel, 2001). In erster Linie zeichnet für diese Aufgabe der 1957 gegründete **Wissenschaftsrat** verantwortlich, der als eine von Bund und Ländern getragene Einrichtung Empfehlungen und Stellungnahmen zur Struktur, Entwicklung, Finanzierung und Leistungsfähigkeit von wissenschaftlichen Einrichtungen sowie zu übergreifenden Fragen des Wissenschaftssystems vornimmt. Daneben gibt es eine Reihe von Forschungsinstituten, die auf den Gebieten der Wissenschafts- und Technologiepolitik, Wissenschaftsindikatorik und Evaluation arbeiten und die seit den neunziger Jahren auch professionell in der **Deutschen Gesellschaft für Evaluation (DeGEval)** organisiert sind. Obwohl es seit den neunziger Jahren auch in Deutschland einen „**Evaluationsschub**“ jenseits des Peer Review-Verfahrens gegeben hat, sind die vorhandenen Evaluationsbemühungen bisher nur wenig aufeinander abgestimmt. Im Folgen-

den sollen im Überblick und in Rückbezug auf die konzeptionellen Überlegungen Aspekte der „**Evaluationstopographie**“ Deutschlands skizziert werden.

Seit 1994 hat der Wissenschaftsrat über die ansonsten allgemeinen strukturbezogenen Empfehlungen hinaus drei Forschungsfelder in fach- und organisationaler Perspektive evaluiert: **Umwelt-, Material- und Energieforschung**. Hierzu wurde auf einige der bereits erwähnten Verfahrensweisen und Wissenschaftsindikatoren zurückgegriffen, beispielsweise auf **Selbst- und Fremdbeschreibungen** (etwa schriftliche Befragungen der Einrichtungen, Begehungen) sowie **Produktivitätsindikatoren** (zum Beispiel die Zahl der Publikationen und Patente). Allerdings wurden keine Effektivitäts- oder Effizienzanalysen durchgeführt. Das in den Abbildungen 1, 2 und 4 dargestellte Indikatortableau hat hier keine breite oder systematische Anwendung erfahren.

Weiterhin wurden auf Vereinbarung von Bund und Ländern zwischen 1997 und 2001 alle gemeinsam geförderten **außerhochschulischen Forschungsinstitutionen** Deutschlands mit dem Ziel evaluiert, die Struktur der außerhochschulischen Forschung in Deutschland international konkurrenzfähig zu gestalten und fortzuentwickeln (Bund-Länder-Kommission, 2001). Aus methodischer und inhaltlicher Sicht handelt es sich bei den „**Systemevalutionen**“ um relativ heterogene Schriften, die in unterschiedlicher Weise auf die skizzierte Wissenschaftsindikatrix zurückgegriffen haben. Während für die Helmholtz-Gesellschaft deutscher Forschungszentren (HGF) in vergleichsweise umfänglicher Weise Produktivitätsindikatoren dokumentiert werden (Wissenschaftsrat, 2001), fehlen vergleichbare Maßzahlen für die FhG und die MPG weitgehend. Im Abschlussbericht der Internationalen Kommission (1999) wird lediglich auf hochaggregierter Ebene die Leistungsfähigkeit der Gesamtorganisation bewertet.

Im Bereich der **Evaluation universitärer Einrichtungen** hat es seit den achtziger Jahren von Seiten der empirischen Wissenschaftsforschung **zahlreiche Einzeluntersuchungen** gegeben, in welchen zumeist auf fachlicher (zum Beispiel Wirtschaftswissenschaften, Anglistik), organisationaler (Hochschulen) oder Abteilungsebene (Fakultäten, Fachbereiche) Forschungsleistungen evaluiert wurden. Diese Untersuchungen haben in vielfältiger Weise auf die in den Abbildungen 1 bis 4 skizzierte Wissenschaftsindikatrix zurückgegriffen (Fisch/Daniel, 1986; Daniel/Fisch, 1988; Bommer/Ursprung, 1997). Allerdings ist aus diesen Anstrengungen **kein systematisches Berichtssystem** entstanden, das universitäre Forschungsleistungen in regelmäßiger Form erfasst. Anstrengungen in dieser Richtung gibt es erst seit 1998 mit der **Novellierung des Hochschulrahmengesetzes (HRG § 6)**, durch welche die regelmäßige Qualitätsbewertung von Forschung und Lehre rechtlich festgeschrieben wurde. Die konkrete Ausgestaltung und Implementierung dieser Gesetzesinitiative liegt in Regie der Bundesländer, die dieser Regelung nach einer Übersicht der Hochschulrektorenkonferenz (2001) in unterschiedlicher Weise nachgekommen sind. Derzeitiger „**Vorreiter**“ bei der praktischen Umsetzung dieses Vorhabens ist das Bundesland **Niedersachsen**, wo mit der **Wissenschaftlichen Kommission** eine zentrale, für Forschungsevaluation verantwortliche Institution geschaffen wurde, die nach definierten Zielvorstellungen und mithilfe eines **Methodenmix** die forschungsbezogene Leistungsfähigkeit niedersächsischer Fakultäten und Fachbereiche evaluiert (Abbildung 5). Diese Evaluationen sollen nicht nur Informationszwecken dienen, sondern auch zukünftige Zuwendungsentscheidungen vorbereiten helfen. Die Kommission hat seit 1999 bereits für zahlreiche Fächer Evaluationsberichte veröffentlicht (<http://www.wk.niedersachsen.de/Materialien/material.htm>).

Im Bereich der **Evaluation von Forschungsprogrammen** zeigt sich insbesondere das Bundesministerium für Bildung und Forschung aktiv. Zwischen 1985 und 1993 wurden fünfzig ministerielle Programme evaluiert, die neben Produktivitäts- auch Wirkungsaspekte anhand einer Vielzahl von Indikatoren diskutieren (Kuhlmann/Holland, 1995). Im Bereich der Grundlagenforschung, etwa

Kuhlmann, S./Holland, D., *Evaluation von Technologiepolitik in Deutschland*, Heidelberg 1995.

Kuhlmann, S., *Evaluationsverfahren in der Forschungs- und Technologiepolitik*, in: *Politikmoderation*, Baden-Baden 1998.

Liefner, I., *Leistungsorientierte Ressourcensteuerung in Hochschulsystemen. Ein internationaler Vergleich*, Berlin 2001.

Luhmann, N., *Die Wissenschaft der Gesellschaft*, Frankfurt/Main 1998.

Organization for Economic Co-operation and Development (OECD) (Hrsg.), *Policy Evaluation in Innovation and Technology*, Paris 1997.

Stock, W. G., *Die Bewertung wissenschaftlicher Forschung und Lehre, 10 Diskussionsbeiträge*, in: *Wissenschaftsevaluation*, München 1994.

Stockmann, R. (Hrsg.), *Evaluationsforschung*, Leverkusen 2000.

Weingart, P. (Hrsg.), *Indikatoren der Wissenschaft und Technik*, Frankfurt/Main 1991.

Wissenschaftliche Kommission Niedersachsen (Hrsg.), *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen, Chemie, Bericht und Empfehlungen*, Hannover 1999.

Wissenschaftliche Kommission Niedersachsen (Hrsg.), *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen, Wirtschaftswissenschaften, Bericht und Empfehlungen*, Hannover 2001.

Wissenschaftsrat (Hrsg.), *Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren*, Köln 2001.

<b>Ziele der Evaluation</b>	<ul style="list-style-type: none"> <li>◆ Ermittlung von Forschungsprofilen und deren gezielte Entwicklung</li> <li>◆ Ermittlung von Maßstäben zur eigenverantwortlichen Planung zur Durchführung von Maßnahmen der Qualitätssicherung (auch Mittelvergabe betreffend)</li> <li>◆ Schaffung einer validen Informationsgrundlage für unterschiedliche Interessengruppen (Politik, interessierte Öffentlichkeit)</li> </ul>
<b>Evaluationskriterien</b>	<ul style="list-style-type: none"> <li>◆ Qualität und Relevanz (Innovativität, wissenschaftliche Ausstrahlung, Interdisziplinarität, wiss. Kooperationsintensität national und international, Effektivität Nachwuchsförderung, Kontaktstrukturen zur Wirtschaft)</li> <li>◆ Effektivität und Effizienz (Verhältnis eingesetzter Mittel und erreichter, beabsichtigter Ziele)</li> <li>◆ Strukturpolitische Aspekte</li> </ul>
<b>Evaluationsverfahren</b>	<ul style="list-style-type: none"> <li>◆ Selbstbeschreibung der zu evaluierenden Einrichtung anhand eines abgestimmten Datenrasters</li> </ul>
<b>Aggregationsebenen der Evaluation</b>	<ul style="list-style-type: none"> <li>◆ Fremdbeschreibung durch Gutachtergremium (eingesetzt von Wiss. Kommission)</li> <li>◆ Stellungnahme der Hochschule zum schriftlichen Gutachterbericht</li> <li>◆ Abschließende Ergebnispräsentation und Empfehlungen durch die Wiss. Kommission</li> <li>◆ Evaluation von Fachdisziplinen (disaggregiert nach Fachbereich, Arbeitsgruppe, Lehrstuhl)</li> </ul>

Abb. 5: Evaluation von Forschungsleistungen an niedersächsischen Hochschulen

den Schwerpunktprogrammen der DFG, gibt es solche Untersuchungen bislang nicht. Allerdings dokumentiert die DFG als zentrale Selbstverwaltungseinrichtung der deutschen Wissenschaft seit Ende der neunziger Jahre mit dem so genannten **DFG-Ranking** ihre Forschungsausgaben. Die absolut bewilligten Gesamtvolumina werden in Relation zu Wissenschaftlerstellen, Fachgebieten und Regionen kartographisch dargestellt (DFG, 2000). Interpretiert man eine **DFG-Bewilligung** nicht als bloßen Mittelzufluss und damit als Forschungsinput, sondern als **Qualitätszertifikat** für geplante Forschungsarbeit, dann wird mit dieser Kartierung eine Leistungskennziffer für allgemeine Forschungsleistung zur Verfügung gestellt und ein wichtiger Beitrag zur Diskussion um die Frage nach den „Centers of Research Excellence“ in Deutschland geleistet.

### Fazit

Um zur Frage der Leistungsfähigkeit des deutschen Wissenschafts- und Forschungssystems kompetent Stellung nehmen zu können, bedarf es zum einen klarer konzeptioneller Vorstellungen, was unter Evaluation von Forschungsleistungen zu verstehen ist und zum zweiten Informationen darüber, wie die gegenwärtige Evaluationspraxis in Deutschland beschaffen ist. Das Gesamtbild zeigt sich hier uneinheitlich. Einzelne Forschungsfelder und die außeruniversitären „Säulen“ der deutschen Forschungslandschaft sind evaluiert worden. Dagegen sind Programmevaluationen bisher die Ausnahme und die systematische Evaluation universitärer Forschungsleistungen ist erst in einem Bundesland verwirklicht. Insgesamt wird Forschungsevaluation Deutschland somit auf unterschiedlichen Ebenen von diversen Akteuren mit bisher nicht vereinheitlichten Maßstäben durchgeführt.

Die Evaluation von Forschungsleistungen wird aber angesichts eines zukünftigen „Europäischen Forschungsraumes“ und dem jetzt angelaufenen 6. Forschungsrahmenprogramm der Europäischen Union immer wichtiger. Die deutsche Zurückhaltung bei Programmevaluationen verweist darauf, dass die hiesige Evaluationspraxis von der Evaluation im Ausland lernen kann, insbesondere von Ergebnissen und Erfahrungen, die bei der Evaluation europäischer Forschungsprogramme gemacht wurden.

### Kontakt:

Dipl.-Soz. Thomas Heinze  
 Fraunhofer-Institut Systemtechnik und  
 Innovationsforschung  
 Breslauer Str. 48  
 D-76139 Karlsruhe  
 Tel.: +49 (0) 721/68 09-1 93  
 Fax: +49 (0) 721/68 09-2 60  
 E-Mail: thomas.heinze@isi.fhg.de