

Gewogen und für zu leicht befunden?

Zur Bewertung von Wissenschafts- und Forschungsleistungen

von Margrit Seckelmann, Felicia Lauer und Andreas Jug

Die Frage nach der bestmöglichen Verwendung öffentlicher Ressourcen wird zunehmend auch für den Bereich von Forschung und Lehre gestellt; gleichzeitig wird die Qualität wissenschaftlicher Leistungen in Evaluationen auf den Prüfstand gestellt. Doch Evaluationen sind längst nicht mehr nur eine Qualitätskontrolle wissenschaftlicher Arbeit, sondern dienen immer häufiger als Grundlage für weitreichende Allokationsentscheidungen. Vor diesem Hintergrund wird die Legitimität der Bewertungen von Forschungs- und Wissenschaftsleistungen diskutiert. In einem zweiten Schritt werden die verschiedenen Verfahren zur Messung der Qualität wissenschaftlicher Leistungen analysiert und Perspektiven für eine Modifikation der bisherigen Evaluationspraxis aufgezeigt.

Einleitung

Nach welchen Kriterien bemisst sich wissenschaftliche Exzellenz? Diese Frage befindet sich schon so lange in der Diskussion, wie die Ordinariuniversität alten Typs von Modellen herausgefordert wird, die sich stärker gegenüber der Öffentlichkeit oder ihren share- und stakeholdern hinsichtlich ihrer Ressourcenverwendung rechtfertigen müssen, wie es das amerikanische tut.¹ Diese Diskussion dauert schon

länger, als es den meisten heute bewusst ist. Schon 1922 prognostizierte Max Weber hellseherisch: »Unser deutsches Universitätsleben amerikanisiert sich, wie unser Leben überhaupt in wesentlichen Punkten [...]. Die technischen Vorzüge sind ganz unzweifelhaft, wie bei allen kapitalistischen und zugleich bürokratisierten Betrieben. Aber der ›Geist‹, der in ihnen herrscht, ist ein anderer als die althistorische Atmosphäre der deutschen Universitäten. Es besteht eine außerordentlich starke Kluft, äußerlich und innerlich, zwischen dem Chef eines solchen kapitalistischen Universitätsunternehmens und dem gewöhnlichen Ordinarius alten Stils.«²

Auch in der deutschen Diskussion ist – angesichts knapper Kassen – die Ressourcenverwendung durch Universitäten und Forschungsinstitute in die Diskussion geraten. Weil eine inputorientierte Ex-ante-Steuerung durch den Staat zu suboptima-

len Ergebnissen führt, gilt die Umstellung auf eine Outputsteuerung und damit einhergehend eine Ergebnis- statt einer Verfahrenskontrolle als »Mittel der Wahl«.³

Die »Konjunktur« von Evaluationen ist alleine schon daran erkennbar, dass sich – nach Jahren eher autodidaktisch ausgeübter Evaluatorentätigkeit – langsam ein Ausbildungsprofil für (Forschungs-)Evaluatoren herausbildet und Aufbaustudiengänge entwickelt werden, zu denen etwa der »Master of Evaluation« der Universität des Saarlandes oder der Nachdiplomstudiengang »Evaluation« der Universität Bern gehören.

Nachdem die DFG Ende der 1960er Jahre damit begann, ihre Sonderforschungsbereiche zu evaluieren, sind Evaluationen Teil der Organisationswirklichkeit aller deutschen Wissenschafts- und Forschungseinrichtungen geworden. Die Zielsetzung von Evaluationen besteht darin, Informationen zu objektivieren und diesbezüglich Transparenz herzustellen. Sind die Bezugsobjekte von Evaluationen zwar verschieden, bilden doch Qualität und Relevanz sowie Effektivität und Effizienz die entscheidenden Bewertungskriterien von Wissenschafts- und Forschungsleistungen.⁴ So unstrittig diese Kriterien in ihrer Allgemeinheit sind, so schwierig ist die Definition der einzelnen Faktoren. Die Unmöglichkeit einer einzigen gültigen Definition von Forschungsqualität vor Augen, erscheint daher die Annäherung über Indikatoren als tragfähige Lösung.

Doch Evaluationen sind längst nicht mehr nur eine Qualitätskontrolle. Denn zunehmend gekoppelt an Allokationsentscheidungen, sind Evaluationen vielmehr »ein Spiel auf Leben und Tod«⁵ von Wissenschaftseinrichtungen, wie sich aktuell am Beispiel des »Center for Advanced Eu-



Dr. Margrit Seckelmann ist Geschäftsführerin des Deutschen Forschungsinstituts für öffentliche Verwaltung Speyer.



Dr. Felicia Lauer ist Mitarbeiterin des Kanzlerbüros der Johannes Gutenberg Universität Mainz.



Diplom-Kaufmann Andreas Jug ist Mitarbeiter in der Institutsleitung des Deutschen Forschungsinstituts für öffentliche Verwaltung Speyer.

1 Zum amerikanischen Modell: Rau 1988: 210; zur Entstehungsgeschichte: Curtius 2003.

2 Weber 1922/1996.

3 Engels 2001: 396.

4 Füller 2004: 11.

5 ebd.

ropean Studies and Research« (CAESAR) verfolgen lässt.⁶

Vor diesem Hintergrund wird nicht nur grundsätzlich die Legitimität von Bewertungen von Forschungs- und Wissenschaftsleistungen diskutiert, sondern auch die Aussagefähigkeit und Qualität der verschiedenen Bewertungsinstrumente und -methoden vorgestellt und kritisch beleuchtet. Schließlich sollen Perspektiven für eine Modifikation der bisherigen Evaluationsverfahren aufgezeigt werden.

Rechtliche Rahmenbedingungen

Die Einführung wettbewerblicher Elemente in der öffentlichen Verwaltung vollzieht sich innerhalb der Grenzen, die durch das

Grundgesetz gezogen werden. Diese betreffen hinsichtlich der Evaluation von Universitäten und Forschungseinrichtungen insbesondere die Frage nach der Zulässigkeit einer extern durchgeführten Bewertung von Wissenschafts- und Forschungsleistungen. Denn diese kann von großer faktischer Bindungswirkung für die spätere Mittelvergabe durch öffentliche Einrichtungen sein, die für von Bund und Ländern gemeinschaftlich finanzierte Einrichtungen etwa der Ausschuss »Forschungsförderung« der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) trifft. Je mehr also die Grenzen zwischen Beratung und Entscheidung verschwimmen können, desto stärker stellt sich die Frage nach der demokratischen Legitimation von Evaluationsgremien oder -agenturen. Bereiten diese Entscheidungen der öffentlichen Verwaltung nur vor oder präjudizieren sie faktisch die spätere staatliche Entscheidungsfindung? Eine solche faktische Bindungswirkung wäre rechtfertigungsbedürftig hinsichtlich der grundgesetzlichen Verbürgung, dass alle Staatsgewalt vom Volke ausgeht (Artikel 20 Absatz 2 Satz 1 Grundgesetz). Das Erfordernis der demokratischen Legitimation von Staatsgewalt fungiert mithin als eine organisatorische Schranke, wenn Beteiligungs- und Entscheidungsrechte auf hochschulexterne Gremien ausgelagert werden.⁷ Sollten die Grenzen einer bloßen Entscheidungsvorbereitung durch Private überschritten werden, so erfordert dies einen Beleihungsakt auf gesetzlicher Grundlage.⁸

Daneben setzt die Wissenschaftsfreiheit des Artikel 5 Absatz 3 Grundgesetz der Beurteilung von Wissenschafts- und Forschungsleistungen Grenzen. Die gewährte Staatsfreiheit der Forschung nach Artikel 5 Absatz 3 Grundgesetz beinhaltet ein individuelles Abwehrrecht wie eine institutionelle Garantie der Wissenschaftsfreiheit. Hinsichtlich des Abwehrrechts wirkt die staatliche Beurteilung von wissenschaftlichen Einzelleistungen Probleme auf.⁹ Als grundsätzlich weniger bedenklich sind Vergleiche der wissenschaftlichen »Leistungen« von Hochschulen, Fachbereichen oder Instituten zu beurteilen. Das Bundesverfassungsgericht hat Evaluationen von

das Ende des wissenschaftlichen »Produktionsprozesses« verlagert wurde.¹⁴

Je mehr Evaluationen von eigens damit beauftragten Privaten, Gesellschaften oder Agenturen durchgeführt werden, desto deutlicher stellen sich schließlich auch Probleme der Gleichbehandlung. Es ergeben sich noch ungelöste Probleme hinsichtlich der Akkreditierungssicherung und der Transparenz der Auswahl dieser Akteure sowie ihrer notwendigen Qualifikationen. Diese Fragen sind in Hinblick auf den Gleichheitsgrundsatz des Artikel 3 Absatz 1 Grundgesetz sowie die Berufsfreiheit (Artikel 12 Absatz 1 Grundgesetz) von Bedeutung. Das Feld des Vergaberechts ist berührt, zumal der Europäische Gerichtshof auch für die Besorgung von Geschäften durch öffentlich-rechtliche Körperschaften für ebensolche zunehmend die Vergaberechtopflichtigkeit bejaht.¹⁵ Hier stellt sich eine Fülle von Problemen, insbesondere auch hinsichtlich der Notwendigkeit einer europaweiten Ausschreibung. Die Komplexität »eines fairen, transparenten und wettbewerbsorientierten Verteilungsverfahrens« muss jedoch in Kauf genommen werden, wenn man »marktwidrigen Privilegierungen« nicht »Tür und Tor öffnen« will.¹⁶

»Die Entwicklung der Evaluation von der Qualitätskontrolle zur existenzentscheidenden Instanz hat den Alltag der Wissenschaftler verändert.«

Instituten und Universitäten für grundsätzlich vereinbar mit Artikel 5 Absatz 3 Grundgesetz erklärt, soweit diese zur Grundlage für eine Ressourcenentscheidungen gemacht werden.¹⁰ Das Bundesverfassungsgericht hat insbesondere die Kopplung von Ressourcenentscheidungen an Leistungsvergleiche (»benchmarking«) zugelassen und diese von einer genuinen inhaltlichen Beurteilung von Forschung durch den Staat unterschieden. Ein »Verbot der Bewertung wissenschaftlicher Qualität oder ein Verbot, an die Bewertung Folgen bei der Ressourcenverteilung zu knüpfen,« lasse sich, so das Bundesverfassungsgericht, Artikel 5 Absatz 3 Satz 1 Grundgesetz nicht entnehmen.¹¹

Dennoch ergeben sich nach dem vom Bundesverfassungsgericht entwickelten Grundsätzen des »Grundrechtsschutzes durch Verfahren«¹² bestimmte verfahrensmäßige Anforderungen insbesondere an die Transparenz des Verfahrens. Die Rechtsweggarantie nach Artikel 19 Absatz 4 Grundgesetz fordert zudem, dass gegen die Entscheidungen, die auf der Grundlage der Beurteilungen von Evaluationsagenturen getroffen werden, der Rechtsweg offen stehen muss. Dies ist insbesondere deswegen der Fall, da unter dem Aspekt der Governance von Forschungseinrichtungen¹³ und der Umstellung von einer Input- auf eine Outputsteuerung die Kontrolle grundsätzlich vom Beginn an

Bewertungen von Wissenschafts- und Forschungsleistungen finden auf den verschiedenen Ebenen des Wissenschafts- und Forschungsprozesses statt und reichen von der individuellen Selbstbewertung der Arbeitsergebnisse durch die Forschenden oder Lehrenden oder die Fremdbewertung von Individuen – etwa im Rahmen der lei-

Instrumente der Bewertung von Wissenschafts- und Forschungsleistungen

6 Horstkotte 2006.
7 Musil 2005: 365.
8 Zu weiteren anderen Gestaltungsmöglichkeiten, Privaten öffentlich-rechtliche Leistungspflichten aufzuerlegen, etwa durch öffentlich-rechtliche Verträge oder die Gründung eines gemischt-wirtschaftlichen Unternehmens oder die Beteiligung an einem solchen vgl. Voßkuhle 2004: 20f.
9 Musil 2005: 365; zur strittigen Frage der Verfassungsmäßigkeit von Lehrevaluationen vgl. Hufen 1995: 1 ff.; zur Anwendbarkeit des Artikels 5 Absatz 3 auf außeruniversitäre Forschungseinrichtungen: Meusel 1999: 148 ff.
10 BVerfG vom 26.10.2004, NVwZ 2005, 315 (317); dazu Musil 2005: 340f.
11 Musil 2005: 342.
12 BVerfGE 49, 89; 53, 30.
13 Dazu Arnold/Groß 2005.
14 Behrens 1996: 219.
15 Dazu: Ziekow/Siegel (2005a); dieselben (2005b).
16 Voßkuhle 2004: 23f.

stungsbezogenen Besoldung – über die Evaluation von Forschergruppen bis hin zu Programm- und Systemevaluationen der Wissenschaftsorganisationen. Bewertungen von Wissenschafts- und Forschungsleistungen können also unterschieden werden nach ihrer Bezugsebene (individuelle/kollektive/organisatorische Leistung), nach dem Evaluationsinhalt (Projekt/Programm/Organisation/Institution), nach der Veranlassung und Durchführung der Evaluation (intern/extern) sowie nach der Zielsetzung. So kann mit einer Evaluation das Ziel einer Qualitätsförderung verbunden sein, es kann aber auch ein Vergleich, eine Positionierung in einem Feld, eine Erfolgsmessung eines Projekts oder Programms oder eine Grundlage einer Finanzierungsentscheidung angestrebt werden. Im Bereich der überindividuellen Evaluation kann grundsätzlich zwischen folgenden Formen differenziert werden:

- Institutionelle Beurteilungen von Wissenschafts- und Forschungsleistungen von Hochschulen oder Instituten (institutionelle Einzelbegutachtungen)
- Querschnitzevaluationen bestimmter Fachgruppen (zum Beispiel durch die Präsidentenkommission der Max-Planck-Gesellschaft)
- Systemevaluationen (zum Beispiel der Einrichtungen der Blauen Liste)
- Strukturuntersuchungen einzelner Fächer.¹⁷

Als häufigste Bewertungsform begegnet die Evaluation von Wissenschafts- und Forschungsleistungen von Hochschulen und Forschungsinstituten, die in Form einer Selbstbewertung mit anschließender Begutachtung durch »peers«, also Fachkollegen, durchgeführt wird. Dieser Evaluationsablauf wurde zuerst durch die HIS für den Bereich der Lehre konkretisiert¹⁸ und hat sich dort als gängiges Modell etabliert.

In Bezug auf die Evaluation der Forschung hat sich in der Praxis eine Kombination verschiedener Evaluationssysteme herausgebildet. Diese verbindet Elemente einer internen mit denen einer externen Evaluation,¹⁹ also indikatorenbasierte Vergleiche (Ratings) mit expertengestützten Bewertungen (Rankings).²⁰ Dieses funktioniert dergestalt, dass die Wissenschaftseinrichtungen zunächst selbst quantifizierbare Leistungskennzahlen (performance indicators) ermitteln. Diese werden von der Hochschulverwaltung erfasst und für die Zwecke der Evaluierung aufbereitet. Die Hochschul- oder Institutsverwaltung ermittelt ihre statistischen Daten auf Grund der Angaben, die ihr von den einzelnen Fachbereichen oder Institutsabteilungen/Sektionen gemacht werden. Der Bericht selbst wird dann zur Grundlage der weiteren Evaluation, die

durch Experten vorgenommen wird (peer review). Diese nehmen neben der Analyse der internen Leistungsbilanz auch ein subjektives Qualitätsurteil vor.²¹ Die Kopplung einer Selbst- und Fremdbewertung sowie die »Begehung«, also die Inaugenscheinnahme, das persönliche Gespräch zwischen den zu Evaluierenden und den Evaluatoren, bilden daher immer noch den Kern jeder Evaluation, die die Analyse der statistischen Daten ergänzt.²²

Messung wissenschaftlicher Qualität über Forschungsindikatoren

Zur Heranziehung quantifizierbarer Messgrößen im Allgemeinen

Im privatwirtschaftlichen Sektor werden quantifizierbare Messgrößen zur Beurteilung von Effizienz und Effektivität von Leistungen und Leistungspotentialen in Unternehmen herangezogen (performance measurement). Das performance measure-

kürlichen, vorher festgelegten Annahmen beruhen, also Konstruktionen sind, die »bestenfalls das wiedergeben, was Wissenschaftler für gut, wichtig oder nützlich halten.«²⁴ Ein Indikator kann mithin immer nur bezogen auf eine bestimmte Zielformulierung interpretiert werden, aus der eine zu erreichende »benchmark« hergeleitet wurde.²⁵ Er ist zudem kontextfrei. Die Bezüge, die ein Experte durch sein Wissen über die Kontextbedingungen für seine Stellungnahme und durch seine Kenntnis des Fachs unschwer herstellen kann, müssen beim Einsatz von Indikatoren mühsam in Anschlag gebracht werden.²⁶

Ohne eine bestimmte Vorstellung von »guter« Wissenschaft in einer bestimmten (Sub-)Disziplin, ohne bestimmte »benchmarks« und Zielvorgaben droht eine quantitative Erfassung schnell, zu einer »l'art pour l'art« zu werden. Diese Zielvorgaben können jedoch nur nach den spezifischen Eigengesetzlichkeiten der jeweiligen (Sub-)Disziplin ermittelt werden. Diese können, etwa hinsichtlich der Frage, ob Monographien oder Aufsätze in double blind peer re-

»Ohne eine Vorstellung von »guter« Wissenschaft in einer Disziplin droht eine quantitative Erfassung, zu einer »l'art pour l'art« zu werden.«

ment wird zunehmend auch im öffentlichen Bereich verwendet, um Methoden in Form von Kennzahlen zum Erfassen und Beurteilen wissenschaftlicher »Leistungen« zur Verfügung zu stellen. Kennzahlen sind quantitative Aussagen über Sachverhalte in verdichteter Form. Während Kennzahlen im engeren Sinne etwas über zumeist unmittelbar beobachtbare Sachverhalte ausdrücken, bilden so genannte Indikatoren Ersatzgrößen, deren Ausprägung und Veränderung Rückschlüsse auf die Ausprägung und Veränderung einer anderen interessierenden Größe zulassen.²³ Allerdings ist auch gegenüber dem Gebrauch gewichteter Indikatoren Vorsicht geboten. Denn ihre Verwendung setzt bestimmte Hypothesen über deren parallele Entwicklung zur messenden Größe voraus. Den Tatbestand und seine Veränderungen selbst können sie hingegen nicht genau abbilden. Entscheidend für die Auswahl von Indikatoren ist, dass auf ihrer Grundlage Evaluatoren in die Lage versetzt werden, die Erreichung von definierten Zielen zu beurteilen und zu kontrollieren. Dabei muss im Auge behalten werden, dass Forschungsindikatoren ihrerseits auf will-

viewed journals als die entscheidenden Publikationsformen anzusehen sind, durchaus unterschiedlich sein. Der Wissenschaftsrat hat den Unterschied der unterschiedlichen »Publikationskulturen« in Natur- und Geisteswissenschaften bereits zur Kenntnis genommen. Er zählt daher zu den »qualifizierte[n] Veröffentlichungen« im Bereich der »naturwissenschaftlichen, medizinischen sowie wirtschafts- und sozialwissenschaftlichen Forschungseinrichtungen insbesondere Aufsätze in referierten Fachzeitschriften«; bei »geisteswissenschaftlichen Forschungseinrichtungen« verlangt er hingegen »ein fachspezifisch überzeugendes Verhältnis von Monographien, Aufsätzen und referierten Fachzeitschriften, Beiträgen zu nicht referierten Zeitschriften und zu Sammel-

17 Wissenschaftsrat 2005: 7.

18 Dazu: Reissert/Carstensen 1997.

19 Behrens 1996: 230f.

20 Hombostel 2004: 175.

21 Behrens 1996: 231.

22 Füller 2004: 11.

23 Gladen 2001: 12.

24 Hombostel 2004: 176.

25 Hombostel 2004: 175.

26 Hombostel 2004: 183.

bänden.«²⁷ Daneben können auch Einladungen an Wissenschaftler des Instituts zu Vorträgen auf wichtigen nationalen und internationalen Konferenzen und die Ausrichtung national und international wichtiger Fachtagungen treten.²⁸ Der Wissenschaftsrat konzidiert somit, dass disziplinäre Standards entscheidend für die Gewinnung von Qualitätsindikatoren von Forschung sind, egal, ob diese »durch kompetente Experten oder durch bibliometrische Vergleichsdaten«²⁹ gewonnen würden.

Die verschiedenen Fachgebiete, die in Universitäten und (interdisziplinär ausgerichteten) Forschungseinrichtungen angesiedelt sind, unterliegen neben dem unterschiedlichen Zugang zu ihrem jeweiligen Forschungsobjekt noch weiteren Besonderheiten. So können Vertreter ingenieurwissenschaftlicher Fächer auf Grund der Anwendungsnähe ihrer Forschungsergebnisse Drittmittel einfacher als ihre Fachkollegen aus den Geisteswissenschaften einwerben. Dazu kommen die viel höheren

Bei der Verwendung allein der Zitierhäufigkeit bleibt die wissenschaftliche Bedeutung der Publikationen unberücksichtigt.³⁰ Deshalb werden auch hier die Zeitschriften, in denen die Zitate erscheinen, mit Qualitäts-Gewichten bewertet. Diese so genannten impact factors sollen abbilden, wie oft ein durchschnittlicher Artikel einer bestimmten Zeitschrift in einer Periode zitiert wurde.

Die Publikationsweisen sind je nach Disziplin und nationaler Wissenschaftskultur jedoch sehr unterschiedlich, so dass ein undifferenzierter Vergleich bibliometrischer Daten mit Problemen behaftet ist. Der Science Citation Index und der Social Sciences Citation Index werfen aus mehreren Gründen zur Beurteilung der Forschungsaktivität der deutschen sozial-, geistes- und rechtswissenschaftlichen Disziplinen Fragen auf. Zum einen beschränken sie sich bei der Zählung auf Fachzeitschriften, wobei geistes- und sozialwissenschaftliche Disziplinen in Deutschland vor allem über

ser« oder »schlechter« der einen gegenüber der anderen Publikationsform ausgegangen werden kann, insbesondere dann nicht, wenn neben der wissenschaftlichen Leistung an sich auch die Auswirkungen der Publikation auf die Umwelt oder die Gesellschaft (outcome) betrachtet werden.

Denn zumeist wird mittels Indikatoren nur die reine »Menge« (output) der Forschungsleistungen beschrieben, ohne dass eindeutig Bezug zur Erreichung einer gewünschten Wirkung oder eines angestrebten Ziels genommen wird. Jedoch können auch Output-Indikatoren, wie beispielsweise eine hohe Publikationshäufigkeit in bestimmten Zeitschriften, Anzeichen für eine große Auswirkung der Forschungsarbeiten auf die Umwelt enthalten und eine klare Abgrenzung zwischen Output- und Outcome-Indikatoren erschweren.

Auch über diese Problematik hinaus bereitet die Messung eher anwendungsorientierter Forschung Schwierigkeiten. Dabei ist zum einen zu beachten, dass anwendungsorientierte Institute nicht immer vollständig autonom in ihrer Themenwahl sind. Die Erwartungen der »stakeholder« dieser Institute, die in deren Aufsichts- oder Verwaltungsräten vertreten sind, sind für die Existenz der Institute von enormer Wichtigkeit. Auch wenn diese Institute nicht der so genannten »Ressortforschung« zugerechnet werden, sondern unabhängige Forschungseinrichtungen sind, ist doch zu beobachten, dass sich der öffentliche Sektor allein auf Grund ihrer Finanzierungsweise oftmals als deren wichtigster »stakeholder« erweist.³³ Nach dessen (berechtigterweise verlangten) Erwartungen werden die Verwertbarkeitsanforderungen an die wissenschaftliche Leistung derartiger Forschungsinstitute formuliert; und diese bilden auch die entscheidenden Bewertungskriterien.³⁴ Allerdings wäre eine reine Serviceleistung gegenüber dem öffentlichen Sektor und anderen »stakeholdern« auch nicht mehr mit dem Charakter eines Forschungsinstituts vereinbar. Gerade die durch die praktische Anschauung vertiefte und oftmals durch Forschungsfragen aus der Praxis angestoßene (grundlagen-)wissenschaftliche Beschäftigung mit einem Forschungsgegenstand, an deren Ende durchaus eine nicht bereits durch die Forschungsfrage vorgegebene, frei erarbeitete

»Je mehr die Grenzen zwischen Beratung und Entscheidung verschwimmen können, desto stärker stellt sich die Frage nach der demokratischen Legitimation von Evaluationsgremien.«

Anschaffungskosten für die apparative Ausstattung in einzelnen Disziplinen, die mit der Drittmittelbewilligung als notwendig für die Projektrealisierung anerkannt wird. Auch bei der Publikationshäufigkeit und der Qualitätsgewichtung dieser Publikationen bestehen gravierende Unterschiede, die fachspezifisch bedingt sind.

Zu den Indikatoren im Einzelnen

Bibliometrische Verfahren stellen auf die Publikationshäufigkeit, also die Menge der erschienenen Publikationen, ab. Hier gibt es die Möglichkeit, verschiedene Publikationsformen mit Gewichten zu versehen, die als Anzeichen für die jeweilige wissenschaftliche Qualität der Publikationsform dienen sollen. Die Zitationsanalyse beziffert die Anzahl an wissenschaftlichen Arbeiten, die auf die Veröffentlichung verweisen. Damit hat die ursprüngliche Forschungsarbeit eine Wirkung auf andere Forschungsarbeiten entfaltet. Die bedeutsamsten Instrumente der Zitationsanalyse stellen der Science Citation Index (SCI) und der Social Sciences Citation Index (SSCI) dar.

Monographien und Sammelwerke veröffentlichten. Außerdem findet dort angloamerikanische Literatur eine stärkere Berücksichtigung, und es wird nur ein Teil der insgesamt existierenden Zeitschriften erfasst.³¹ Zudem fehlt es in diesen Kontexten zumeist an einer klaren »Hierarchisierung der Publikationsmedien«.³² So ist etwa in den Rechtswissenschaften eine Publikation in einer der zumeist im 19. Jahrhundert gegründeten Archivzeitschriften (Archiv des öffentlichen Rechts, Archiv für civilistische Praxis, Verwaltungsarchiv, Goldammer's Archiv für Strafrecht und andere) mit besonderem wissenschaftlichen Prestige in den jeweiligen Subdisziplinen der Rechtswissenschaft versehen. Den größeren Einfluss auf die Rechtswirklichkeit hat jedoch eine Publikation in der von sämtlichen Juristen (Wissenschaftlern wie Praktikern des Zivil-, Straf- und Öffentlichen Rechts) gelesenen Neuen Juristischen Wochenschrift (NJW). Je nach Zuschnitt eines Forschungsinstituts (eher theoretisch oder auch für die Rechtspraxis arbeitend) können hier äußerst unterschiedliche Akzente gesetzt werden, ohne dass von vornherein von einem »bes-

27 Wissenschaftsrat 2005: 9.

28 Wissenschaftsrat 2005: 9.

29 Hombostel 2004: 175.

30 Kramer 2005: 162.

31 Kramer 2005: 162 f.

32 Hombostel 2004: 179.

33 Hener 2004: 116; Schimank 2004: 84.

34 Hombostel 2004: 174.

Anregung für die Praxis stehen kann, macht den Charakter eines anwendungsbezogenen Forschungsinstituts aus.

Der Wissenschaftsrat hat daher hinsichtlich der Erfassung der »Leistungen« anerkannt, dass für »qualitativ hochwertige Dienstleistungen für die Forschung oder Leistungen auf dem Gebiet der wissenschaftlich basierten Beratung« bei anwendungsbezogenen Instituten auch ein »angemessener Anteil an eigener Forschungsleistung notwendig« sei. Bei Instituten mit hohem Anteil an Service- und/oder Beratungsaufgaben gelten daher »die Kriterien für Forschung ebenfalls«, sie seien »aber den überwiegend andersartigen Aufgaben entsprechend anzupassen und gegebenenfalls zu ergänzen«, zum Beispiel nach »der Qualität der wissenschaftlichen Fundierung der Beratung durch eigene Forschungsarbeit«, der Aufarbeitung aktuellen Wissens, der Unabhängigkeit des Instituts, einer überzeugenden Strategie und transparenter Beratungsprozeduren und schließlich des Erreichens von geeigneten Zielgruppen.³⁵

Es kommt hinzu, dass bibliometrische Daten den Inhalt der Publikation allenfalls sehr indirekt über ihre Zitationshäufigkeit erfassen, so dass eine undifferenzierte Anwendung dieser Daten, die nicht auf ihre Kontextbedingungen abstellt, zu Fehlsteuerungen führen kann. Zugespielt könnte man sagen, dass Quantität auf Kosten der Qualität und Schnelligkeit vor der Gründlichkeit den Vorzug bekäme. So werden – unter möglichst plakativen Überschriften – in manchen Fächern inzwischen eher Forschungsprogramme als Forschungsleistungen dargestellt. Das Vorgehen, in Evaluationen beinahe ausschließlich auf Publikationen abzustellen, ist einer gründlichen Umgangsweise mit Forschungsergebnissen zumindest nicht förderlich. Zu den Folgen dieser Politik gehören jedoch nicht nur die in den vergangenen Jahren publik gewordenen Fälschungsphänomene, sondern auch die Taktik der »Salamiveröffentlichung«, also der Publikation selbst kleinteiligster Forschungsergebnisse, Mehrfachpublikationen, die Häufung von Eigenzitationen oder Gruppen-Selbst-Zitationen und nicht zuletzt das hartnäckige Weiterbestehen der problematischen Praxis der Aufnahme von »Autoren«, die im strengen Sinne nicht an der Publikation beteiligt waren.³⁶ Kritische Analysen der szientrometrischen Verfahren kommen daher zu dem Ergebnis, dass »die vielfach Kooperierenden, mehrfach und kurz Publizierenden«³⁷ bevorteilt werden.

Die Probleme potenzieren sich, wenn nicht nur disziplinäre Standards, sondern auch verschiedene nationale Standards aufeinandertreffen, wie dies zunehmend

im Rahmen der EU-Forschungsförderungsprogramme geschieht. So hat Stefan Hornbostel darauf hingewiesen, dass nach den Kriterien eines Rankings für die europäischen Wirtschaftswissenschaften,³⁸ das lediglich zehn europäische Spitzenjournals (mit einer Qualitätsgewichtung) berücksichtigt, weniger als ein Prozent der Publikationen der deutschen Volkswirte in den Jahren 1998 bis 2000 nach diesen Kategorien erfasst werden können.³⁹

Gegenüber der Erfassung bibliometrischer Daten hat die peer review den Vorteil, dass dort die Kontextbedingungen unschwer hergestellt werden können. Allerdings sind die Verfahren der Auswahl der reviewer und der Art und Weise ihrer Beurteilungen (beispielsweise der »double blind peer review« als Königsweg) noch äußerst umstritten.⁴⁰

»Der Aspekt der Darstellbarkeit der Forschung bekommt durch ihre Evaluation ein zunehmendes Gewicht.«

Die durch Wissenschaftler eingeworbenen Drittmittel werden häufig als Indikator herangezogen, weil sie in einer monetär messbaren Größe ausdrücken, welche Qualität Externe, nämlich die Drittmittelgeber, der Forschung beimessen. Für den wissenschaftlichen Rang eines Drittmittelprojektes spielt besonders das Verfahren der Mittelvergabe eine Rolle. Wenn Forscher um die Drittmittel konkurrieren und/oder schließlich Experten die Anträge bewerten, ist die wissenschaftliche Qualität als hoch einzustufen. Nach diesen Kriterien werden beispielsweise das Vergabeverfahren der DFG, der EU-Forschungsrahmenprogramme und der Volkswagen-Stiftung allgemein als besonders anspruchsvoll angesehen. Unter den Forschern herrscht Wettbewerb unter relativ marktnahen Bedingungen. Allerdings ist auch bei der Begutachtung von Drittmittelanträgen ein subjektives Element vorgeschaltet. Je nach Kultur einer bestimmten (Sub-)Disziplin kommen verschiedene Drittmittelgeber aus der Forschungsförderung oder der Forschungspraxis in Betracht. In den Rechtswissenschaften würde beispielsweise die »Nutzung des Drittmittelindikators äußerst selektiv über einen bestimmten Typ von Forschung« berichten.⁴¹

Die Förderung des wissenschaftlichen Nachwuchses über erfolgreich betreute Promotionen und Habilitationen dokumentiert im Allgemeinen Fortschritte im Be-

reich der Forschung und erfasst somit zumindest mittelbar die Forschungsleistung. Es ist allerdings zu berücksichtigen, dass die Dauer der Qualifikationsphasen je nach Fächerkultur deutlich variieren kann. Ebenso wird die Promotionshäufigkeit in den einzelnen Disziplinen unterschiedlich gehandhabt. So ist in Fächern wie Biologie oder Medizin eine Promotion in Hinblick auf spätere Beschäftigungsmöglichkeiten üblich, wohingegen diese in den Ingenieurwissenschaften eher selten ist.⁴²

Schwierigkeiten der unterschiedlichen Fächerkulturen ergeben sich insbesondere auch bei interdisziplinären Arbeiten, bei denen sich die Frage stellt, nach welchen Kategorien diese Arbeiten zu bewerten sind. So gelten für die deutsche Volkswirtschaft allen oben genannten »Rückständen« gegenüber der angloamerikanischen

Publikationskultur zum Trotz ganz andere Messkategorien (Publikation in internationalen »double blind peer reviewed journals«) als beispielsweise in den Rechtswissenschaften, in denen derartige Zeitschriften so gut wie nicht vorhanden sind und zumindest (sub-)disziplinenintern nicht das standing der zumeist durch Schriftleiter oder Herausgeber (beiräte) offen beurteilten Archivzeitschriften, der renommierten Juristenzeitung (JZ) oder der Praktikerzeitschriften haben. Schwierigkeiten haben auch diejenigen, die ihre Profession sozusagen »wechseln« und beispielsweise als Juristen vorwiegend verwaltungswissenschaftlich publizieren. Nach welchen Kategorien sollen deren Arbeiten zu beurteilen sein, nach der »Organisationszugehörigkeit« oder der studierten Profession?⁴³

35 Wissenschaftsrat 2005: 11f.

36 Bull/Mehde 2000: 651.

37 Fröhlich 2003: 62.

38 Kalaizidakis 1999.

39 Hornbostel 2004: 180.

40 Dazu Hornbostel 2004: 178.

41 Hornbostel 2004: 177; vgl. auch Oebbecke 2005.

42 Schmitzer/Isserstädt/Müßig-Trapp/Schreiber 1998: 307.

43 Hornbostel 2004: 184.

Chancen und Gefahren der bisherigen Evaluationspraxis

Die Entwicklung der Evaluation von der Qualitätskontrolle zur existenzentscheidenden Instanz hat den Alltag der Wissenschaftlerinnen und Wissenschaftler verändert. Die externe Evaluation ist zum festen Bestandteil des Forschungsprozesses geworden. Doch wenn die »Testsituation in manchem Institut den Forschungsalltag dominiert, dann beginnt das Kerngeschäft der Wissenschaft zu leiden.«⁴⁴ Institutsforschung wird weniger anhand selbst entwickelter Kriterien innovativ betrieben, dafür bekommt der Aspekt der möglichen Darstellbarkeit der Forschung ein zunehmendes Gewicht. Das führt zum einen dazu, dass im Rahmen der Außendarstellung etwa von Sonderforschungsbereichen das »Proben« von Präsentationen oder gar die Einschaltung von Werbeagenturen einen immer größeren Raum einnimmt. Auf die Probleme der bibliometrischen »Erfassung« von Forschungsleistungen war bereits hingewiesen worden. Hinsichtlich des Abstellens auf die Akquisition von Drittmitteln als Indikator für wissenschaftliche

bestimmten Fonds finden, die oftmals als »Spielwiesen«-Etats bezeichnet werden, während die Institute – auch durch eine Verschärfung der Konkurrenzsituation – ihre Akzente andernorts setzen müssen.⁴⁷ Die Ausfluggung »riskanter« Forschung als eines Objekts für »Spielwiesen« oder »Reptilienfonds« stärkt zudem die Position der Dekane oder Institutsdirektoren und sorgt tendenziell für eine Schwächung der akademischen Selbstorganisation, die – in ihren positiven wie negativen Auswirkungen – »ein Schutzwahl für unorthodoxe Perspektiven sein kann.«⁴⁸

Vor dem Hintergrund des zunehmenden Gewichts bibliometrischer Analysen in Evaluationen ist die Forderung nach einer »Rückkehr zu qualitativen Maßstäben«⁴⁹ in der Bewertung von Forschungs- und Wissenschaftsleistungen nachvollziehbar. Trotz der geschilderten Bemühungen, bibliometrische Messungen beispielsweise durch den journal impact factor qualitativ anzureichern,⁵⁰ stellen diese ihrem Charakter nach quantitative Kennzahlen dar. Sie sind zur Ableitung qualitativer Aussagen über die Leistung eines Einzelnen oder kleinerer Arbeitsgruppen quasi überhaupt nicht und

fahren, die letztlich nur der Erhöhung einer zweifelhaften Scheingenaugigkeit dienen, und ihrer Ausbreitung in den verschiedenen Wissenschaftsdisziplinen soll daher hier eher eine »Rückführung« auf ihre tatsächliche Aussagekraft über die Produktivität von Wissenschaftlerinnen und Wissenschaftlern angeregt werden. Doch selbst als Produktivitätskenngrößen können diese Zahlen nicht unkritisch verwendet werden. Die verschiedenen Publikationsformen werden in Natur-, Geistes- und Sozialwissenschaften sowie den Theologien völlig anders gewichtet; auch das Publikationsverhalten unterscheidet sich sehr stark. Elektronische und damit dynamische Publikationen schließlich werfen fächerübergreifend die Frage danach auf, was überhaupt als zählbare Publikationseinheit den entsprechenden Analysen zugrundegelegt werden kann.⁵³ All dies spricht daher dafür, sich im Rahmen von Evaluationen wieder stärker dem Publikationsinhalt zuzuwenden. An Stelle extensiver Publikationslisten – beispielsweise im Rahmen von Berufungsverfahren – sollte daher die bewusste Einschränkung und Konzentration auf eine überschaubare Anzahl von Publikationen treten. Darüber hinaus sollte in allen Formen der Evaluation der direkten »peer review« der Vorzug gegenüber indirekten Qualitätsurteilen gegeben werden.

Der zunehmende Alltagscharakter von externen Evaluationen kann nämlich auch dazu führen, dass bestimmte Aspekte – wie die Frauenförderung – zu einer Formalität abgeschliffen werden, die es durch eine positive Aussage zu »erledigen« gilt. So werden in internen Evaluationsreports mehr oder weniger routiniert »Bekanntnisse« zur Frauenförderung oder zum gender mainstreaming abgelegt, oder es wird mit Hilfe von »Textbausteinen« auf Frauenförderpläne hingewiesen, während diesem Aspekt in der Praxis kaum Beachtung geschenkt wird – und er wohl auch nicht eine entscheidende Rolle für das Evaluationsergebnis spielt. Dass die Quote von Forscherinnen in Deutschland immer noch verglichen mit anderen Ländern verschwindend gering ist und auch trotz des Einzugs des Themas in die Strategiekonzepte, Forschungsstrategien, Kriterienkataloge von

»Eine systematisch angelegte ›Evaluation der Evaluation‹ im Bereich der Wissenschaftsorganisation liegt nicht vor.«

Exzellenz besteht eine durchaus ernstzunehmende Gefahr, dass »die Professoren als erfahrene Forscher immer mehr zu »Akquisitionsprofis« werden müssen, während die Forschungsarbeiten selbst dann vor allem noch von relativ unerfahrenen und kaum angeleiteten Nachwuchswissenschaftlern erledigt werden.«⁴⁵ Von einer »Exzellenzsteigerung« von Forschung kann in diesem Zusammenhang dann nicht mehr unkritisch die Rede sein.

Durch ein ständiges »Schielen« darauf, wie potenzielle Evaluatoren den Zuschnitt und die Ergebnisse eines Forschungsprojekts beurteilen könnten, kann die Innovationshöhe von Forschung abnehmen und die »Modeabhängigkeit« zunehmen. So kann beispielsweise das »Durchhalten von längerfristigen »Forschungslinien«« schwieriger werden, was für die Grundlagenforschung nicht ohne Bedeutung ist.⁴⁶ Innovative Grundlagenforschung ohne konkrete Nutzenorientierung, so genannte »riskante« Forschung oder unorthodoxe Perspektiven können in Hinblick auf kurzfristige Evaluationszeiträume schwieriger werden und eine Förderung allenfalls in

zur Messung der Leistung größerer Kollektive nur sehr eingeschränkt geeignet: »Von einer Würdigung dessen, was die Qualität wissenschaftlicher Leistung ausmacht, nämlich ihrer Originalität, ihrer »Innovationshöhe«, ihres Beitrags zum Erkenntnisfortschritt, sind sie weit entfernt, und ihr immer häufigerer Gebrauch bringt sie in Gefahr, von Hilfsmitteln zu Surrogaten des Qualitätsurteils zu werden.«⁵¹ Die Praxis, aus bibliometrischen Daten Aussagen zur Qualität oder aber – über die Analyse der Zitationshäufigkeit von Publikationen – zum »Nutzen« einer wissenschaftlichen Veröffentlichung abzuleiten, erscheint um so problematischer, wenn man bedenkt, dass zunehmend »Sein oder Nicht-Sein« von Personen, Institutionen und ganzen Disziplinen abhängen und auf der Basis dieser Aussagen nicht nur Forschungsfelder, sondern auch Posten, Positionen, Reputation und wissenschaftliche Preise zugeteilt werden und die Existenz von Forschungseinrichtungen von diesen Daten abhängen kann.⁵²

An Stelle der vielfältigen Aktivitäten zum Ausbau szientrometrischer Messver-

44 Füller 2004: 12.

45 Schimank 2004: 83.

46 Schimank 2004: 84.

47 Schimank 2004: 84; Lange 2005.

48 Schimank 2004: 84.

49 Schneider 2003: 2.

50 Dazu Wald 2005.

51 Fröhlich 2003: 61.

52 Fröhlich 2003: 61.

53 Stock 2001: 5.

Evaluationen und anderes nicht im nennenswerten Umfang ansteigt, belegt die rein formale Repräsentanz und Relevanz dieses Themas sowie ähnlicher Aspekte. Wenn Evaluationen tatsächlich zur Weiterentwicklung von Forschung und Forschungsorganisation beitragen wollen, bedarf es daher einer neuen Diskussion über die Gewichtung einzelner Kriterien.

Schließlich dürfen Evaluationsverfahren nicht als Möglichkeit missbraucht werden, Geld zu sparen.⁵⁴ Dies gilt um so mehr, als die Mittelzuweisung nach der Leistungsbemessung ohnehin mit einem Verteilungsproblem verbunden ist: Wenn (zusätzliche) Mittelbewilligungen an die Erfüllung bestimmter »Exzellenzkriterien« gebunden werden, so werden letztlich diejenigen, die diese erfüllen, mit (staatlicher) Unterstützung in die Lage versetzt, diejenigen weiter »abzuhängen«, die – teilweise nur unwesentlich – »schlechter« waren als sie selbst. Dies kann gerade Universitäten und Forschungseinrichtungen, die sich in Veränderungsprozessen befinden, in ihren Bemühungen wieder »zurückwerfen«. Man nennt dies nach dem biblischen Gleichnis von der Belohnung derjenigen, die haben, auch den so genannten »Matthäus-Effekt«.⁵⁵

Nicht zu unterschätzen sind weiter die Kosten, die durch die Umstellung auf eine Outputsteuerung entstehen. Im Einzelfall kann dies dazu führen, dass durch die Summe der dezentral aufzubauenden Kontrollmechanismen höhere Gesamtkosten entstehen. Deshalb sollte die Kosten-Nutzen-Relation für die neuen Regelungsmechanismen genau betrachtet werden.

Für die Qualität und Aussagefähigkeit von Evaluationen ist die Qualität der Evaluatoren entscheidend. Bisher sind die Qualifikationsmerkmale von Evaluatoren noch nicht institutionsübergreifend geregelt. Vielmehr besteht ein »common sense« dahingehend, dass weitreichende Kenntnisse über die Strukturen in Forschungsorganisationen sowie eine ausgewiesene fachliche Reputation und natürlich das Fehlen eines Eigeninteresses die zentralen Voraussetzungen für die Ausübung von Evaluatorentätigkeiten sind. Doch zu der System- und durch eigene Forschungstätigkeit erworbenen Fachkenntnis muss auch methodisches Wissen über Evaluationsverfahren, Moderationsfähigkeit und Vermittlungskompetenz hinzukommen. Die Fähigkeit zur Analyse und Interpretation statistischer Daten ergänzt dieses Kompetenzprofil.⁵⁶ In diesem Zusammenhang ist auch die Frage nach einer »Evaluation der Evaluatoren« zu diskutieren. Mit Blick darauf, dass Evaluatorentätigkeiten in Deutschland bisher – mit

Ausnahme der Evaluationsagenturen – von Forscherinnen und Forschern quasi »ehrenamtlich« und nur gegen eine Aufwandsentschädigung übernommen werden, ist dem Aspekt einer systematischen Auswertung der Evaluatorentätigkeit noch wenig Bedeutung beigemessen worden. Erst langsam bildet sich ein Bewusstsein für diese Problematik.

Ebenso bedenkenswert erscheint die Frage, wie Evaluationsergebnisse letztlich umgesetzt werden beziehungsweise welche Steuerungseffekte im Hinblick auf die Organisationsentwicklung eintreten. Führt eine gute Beurteilung von Instituten oder deren Abteilungen und Fachbereichen möglicherweise dazu, dass der status quo quasi »einbetont« wird und Weiterentwicklungen eher verhindert werden? Wer garantiert, dass die nächste Evaluationskommission ähnlich zusammengesetzt ist wie die ursprüngliche und ähnliche Kategorien ihrer Leistungsbewertung zu Grunde legt?

»An Stelle des Ausbaus szientometrischer Messverfahren regen wir eine »Rückführung« der Evaluation auf ihre tatsächliche Aussagekraft über die Produktivität der Wissenschaftler an.«

Schließlich sind bisher Forschungs- und Lehrevaluationen an Universitäten noch zu wenig miteinander verzahnt.⁵⁷ Beide Bereiche werden unabhängig voneinander und von unterschiedlichen Gutachtern evaluiert. Dementsprechend gelingt der notwendige Bezug der Evaluationsergebnisse, etwa in der Frage der Forschungsorientierung der Lehre oder der Relevanz von Forschungsschwerpunkten für die Lehre eines Faches, nur eingeschränkt.

Mögliche Perspektiven für Evaluationsverfahren

Die Bewertung oder Evaluation von Wissenschafts- und Forschungsleistungen ist gerade nicht, wie vielfach vertreten wird, eine bloße Übertragung von Methoden aus der Privatwirtschaft auf die Wissenschaft. Vielmehr ist sie vom Wissenschafts- und Forschungsprozess selbst nicht zu trennen, da sie die Voraussetzung für deren Weiterentwicklung und damit einen wichtigen Motor in der Wissenschaft⁵⁸ darstellt.

Trotz aller hiermit verbundenen Schwierigkeiten sind Evaluationsverfahren

jedoch nicht grundsätzlich abzulehnen. Zwar liegt noch keine systematisch angelegte Evaluation der Evaluation im Bereich der Wissenschaftsorganisationen vor, jedoch sind positive Auswirkungen auf das Forschungsprofil der Wissenschaftseinrichtungen und das Ziel der Qualitätsförderung zu verzeichnen. Evaluationen können – wie das Bundesverfassungsgericht ausgeführt hat⁵⁹ – durchaus zu einer Transparenz der Wissenschaftspolitik beitragen, sofern die Leistungsvergleiche tatsächlich nach vergleichbaren Verfahren vorgenommen werden. Insofern ist eine an sich wünschenswerte »Umsteuerung« durch Anpassung der Evaluationsverfahren an die bisherigen Erfahrungen bei hintereinander durchgeführten Evaluationen von Instituten derselben Wissenschaftsgemeinschaft sowohl unter dem Gesichtspunkt des Gleichbehandlungs- (Artikel 3 Absatz 1 Grundgesetz) wie des »fair trial«-Grundsatzes als Ausformung des Rechtsstaats-

prinzips (Artikel 20 Absatz 1 Grundgesetz) mit großer Vorsicht zu behandeln.

Zur Benennung der Unterschiede hinsichtlich der Qualität und Relevanz von Forschung bedarf es zwingend des Blicks von »außen«. In Bezug auf die Dominanz indikatorgestützter Bewertungen ist jedoch abschließend festzuhalten, dass Kennzahlensysteme nicht »selbstregulativ« sind, sondern stets einer Interpretation bedürfen. Auch dies spricht nicht gegen eine indikatorgestützte Bewertung an sich, da die Auswahl der Indikatoren und die Diskussion über diese in den Einrichtungen selbst eine fruchtbare Debatte über die Strategie von Universitäten, Fachbereichen und Forschungsinstituten nach sich ziehen und verkrustete Strukturen aufbrechen kann. Auf Grund der aufgezeigten Problematik der einzelnen Indikatoren ist aber auf je-

54 Voßkuhle 2004: 31.

55 Merton 1968: 56.

56 Schreiber 2004.

57 Künzel 2005.

58 Füller 2004: 11.

59 BVerfG NVwZ 2005, 315.

den Fall ein mehrdimensionales Indikatoren-system einseitigen Analysen vorzuziehen.⁶⁰ Dazu muss die Definitionshoheit über die tauglichen Kriterien innerhalb eines gewissen Rahmens – der Wissenschaftsrat nennt etwa die Kohärenz des Forschungsprogramms, eine überzeugende Schwerpunktbildung, innovative Ansätze, überzeugende mittelfristige Perspektiven für die wissenschaftliche Arbeit des Instituts und die Integration in die nationale und internationale Forschungslandschaft⁶¹ – bei den Wissenschaftseinrichtungen selbst liegen und von diesen mit ihren Zuwendungsgebern ausgehandelt werden.

Indikatoren können immer nur vor dem Hintergrund einer konkreten Zielvorstellung entwickelt werden. Sie sollten zudem fach- und institutionenspezifisch differenziert sein, dass sie neben dem eingesetzten Input und dem erzieltem Output auch qualitative Dimensionen der Leistungen als Outcome abbilden können.⁶² Um gleichzeitig Manipulationen zu vermeiden, könnte ein vorgegebenes, festes Set an Kennzahlen um eigene, für wichtig erachtete, Angaben erweitert werden. Dies gilt umso mehr für interdisziplinär ausgerichtete Institute, bei denen aus unterschiedlichen Fächern sinnvolle Hilfsgrößen zusammengefasst werden müssen. Dort könnten dann auch Synergien, die sich aus der Zusammenarbeit ergeben, Berücksichtigung finden.⁶³ Damit Schwankungen der Werte im Zeitablauf bei einzelnen Indikatoren ausgeschaltet werden können, kommt auch eine Durchschnittsbildung über längere Zeiträume in Betracht.

Die Zeiträume, nach denen eine Institution evaluiert werden soll, müssen sinnvoll festgelegt werden. Erfolgt der Einsatz von Evaluationen, um damit die Umsetzung zurückliegender Empfehlungen von Gutachtern sowie die Entwicklung und Umsetzung von Forschungsprogrammen aufzuzeigen, scheinen dafür eher langfristige Bewertungszeiträume von sechs bis acht Jahren geeignet.⁶⁴ Nur dann sind die Ergebnisse aus am Forschungsprogramm ausgerichteten Projekten erkennbar und zu bewerten. Ebenso darf nicht außer Acht gelassen werden, dass die Vor- und Nachbereitung einer Evaluation einen großen Zeitaufwand bei den Wissenschaftlern verursacht und ein kurzfristiger Rhythmus die Forschungstätigkeit unverhältnismäßig beeinträchtigen kann.

Der Zusammenhang zwischen Evaluation und Steuerung muss schließlich intensiver als bisher diskutiert werden. Unübersehbar sind die »Anpassungsleistungen« von Wissenschaftlerinnen und Wissenschaftlern im Hinblick auf fragwürdige Bewertungskriterien. Insbesondere für die Geistes- und

Sozialwissenschaften gilt es, anstelle der unkritischen Übertragung der Bewertungskriterien aus der Medizin und den Naturwissenschaften solche Indikatoren zu finden, die nicht nur produktivitätssteigernd wirken, sondern gleichzeitig Exzellenz und wissenschaftliche Nachhaltigkeit fördern. Schließlich ist gerade in Bezug auf diese Bereiche über die langfristig beabsichtigten Steuerungseffekte zu beraten. Die Entwicklung von Indikatoren muss sich daher an den Entwicklungszielen dieser Wissenschaften orientieren. Anstelle einer kurzfristigen Steigerung des Outputs wären beispielsweise eine stärkere Vernetzung und Projektorientierung, eine deutlichere Schwerpunktsetzung und Programmbildung anzustreben und entsprechende Indikatoren und Kennzahlen in die Evaluationen zu integrieren. Qualitätssteigernd kann sich auch eine stärkere internationale Vernetzung auswirken, so dass die Zahl der incoming/outgoing von Wissenschaftlerinnen und Wissenschaftlern in diesen Wissenschaftsbereichen eine relevante Messgröße darstellt (im Unterschied zu den Naturwissenschaften, hier ist der Austausch in der Regel so intensiv, dass es keiner Steuerungswirkung über einen entsprechenden Indikator bedarf). Auch der Nachwuchsförderung ist besondere Aufmerksamkeit zu widmen, insbesondere den Mechanismen der gezielten Auswahl, Leistungskontrolle und Förderung des wissenschaftlichen Nachwuchses. Auch dieser Sachverhalt sollte in Indikatoren abgebildet werden. Zumindest in Bezug auf die Hochschulen und Universitäten ist darüber hinaus sicherzustellen, dass in den Geistes- und Sozialwissenschaften leistungsbezogenen Messgrößen Belastungsindikatoren gegenübergestellt werden: Denn diese Fächer haben traditionell eine stärkere Ausbildungsbelastung (deutlich höhere Studierendenzahlen, intensive Betreuung- und Prüfungskonzepte) zu bewältigen als die Naturwissenschaften. Somit sind die Ausgangsbedingungen für qualitativ hochwertige Forschung andere.

Schließlich gibt es auch in den Geistes- und Sozialwissenschaften eine Art »Wertungszusammenhang«, der jedoch deutlich schwieriger als in den Naturwissenschaften zu operationalisieren ist. Er findet seinen Ausdruck in denjenigen Aktivitäten, die dem »Import« des in diesen Disziplinen gebündelten Wissens in gesellschaftliche Zusammenhänge dienen. Mögliche Indikatoren können beispielsweise die Kooperation mit gesellschaftlichen Institutionen sein, aber auch (politik)beratende Aktivitäten, Medienpräsenz etc. können hierunter gefasst werden.

Mit der eingenommenen kritischen Perspektive auf Evaluationsverfahren und -me-

thoden soll es darum gehen, sondern die Notwendigkeit von Weiterentwicklungen aufzuzeigen. Hierzu zählt neben der Neugewichtung einzelner Indikatoren auch die Findung zeitlich angemessener Evaluationsabstände. Insbesondere müssen Evaluationen stärker als bisher mit Strategieprozessen verknüpft und auf diese bezogen werden. Dies bedeutet, dass nicht nur die Aktivitäten in Forschung, Studium und Lehre evaluiert werden, sondern die vorgenommenen Schwerpunktsetzungen selbst Gegenstand von Evaluationen sein müssen. Während Evaluationen in den genannten Bereichen an deutschen Wissenschaftseinrichtungen längst zum Alltagsgeschäft geworden sind, mangelt es gerade in Bezug auf die Evaluation der Strategieentwicklung und -umsetzung und damit im Kern der gesamten Entwicklung von Wissenschaftseinrichtungen weitgehend an Erfahrung. Hier wäre der Ansatzpunkt für eine Weiterentwicklung der bisherigen Evaluationsverfahren und -methoden gegeben.

Literatur

Arnold, Nathalie/Groß, Thomas (2005), Die Entscheidungsstrukturen der Leibniz-Gemeinschaft, *Wissenschaftsrecht* 38, 238-263.

Behrens, Thomas (1996), Globalisierung der Hochschulhaushalte. Grundlagen, Ziele, Erscheinungsformen und Rahmenbedingungen, Neuwied u. a.: Luchterhand.

Bull, Hans Peter/Mehde, Veith (2000), Reform der Hochschulorganisation – die populären Modelle und ihre Probleme, *Juristenzeitung*, 650-659.

Curtius, Carl Friedrich (2003), So etwas wollen wir bei uns nicht haben – Zur Entwicklung des Evaluationsgedankens im Hochschulwesen, in: Klaus Anderbrügge/Volker Epping/Wolfgang Löwer (Hrsg.), *Dienst an der Hochschule*, Festschrift für Dieter Leuze, Berlin: Duncker & Humblot, 109-121.

Ebel-Gabriel, Christiane (2003): Forschungs-evaluation als Basis strategischen Handelns an den niedersächsischen Hochschulen, in: *evaNet-Reihe »Positionen«*, Februar 2003, Download unter http://evanet.his.de/old_evanet/forum/pdf-position/GabrielPosition.pdf (6. Juni 2006).

Engels, Maria (2001): Die Steuerung von Universitäten in staatlicher Trägerschaft – Eine organisationstheoretische Analyse, Wiesbaden: Deutscher Universitäts-Verlag.

Fries, Marlene (2003): Evaluation von Forschung und Lehre in Bayern, in: Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung (Hrsg.), *Beiträge zur Hochschulforschung*, Heft 1, 25. Jahrgang, München: Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung.

60 Hombostel 2004: 183.

61 Wissenschaftsrat 2005: 5.

62 Schenker-Wicki 1999: 154 ff.

63 Fries 2003: 29.

64 Ebel-Gabriel 2003, 7; Viehoff 2003, 6.

Fröhlich, Gerhard: Gegen-Evaluation: Der Impact-Faktor auf dem Prüfstand der Wissenschaftsforschung. In: BUKO Info 1-4/2003, 6165. Download unter <http://www.oew.ac.at/ita/ebene5/Froehlich.pdf> (16. September 2005).

Füller, Christian (2004): Die Evaluatoren kommen! In: *duz MAGAZIN* 03/2004, 10-13.

Gladen, Werner (2001): Kennzahlen und Berichtssysteme: Grundlagen zum Performance Measurement, Wiesbaden: Gabler.

Hener, Yorck (2004), Qualitätsmanagement in Hochschulen, in: Hochschulrektorenkonferenz (Hrsg.), *Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen*, Bonn: Beiträge zur Hochschulpolitik, 111-126.

Hornbostel, Stefan (2004), Leistungsparameter und Ratings in der Forschung, in: Hochschulrektorenkonferenz (Hrsg.), *Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen*, Bonn: Beiträge zur Hochschulpolitik, 173-187.

Horstkotte, Hermann (2006): Bonner Forschungszentrum: Alzheiner statt Aquaplaning. In: *spiegel online* (<http://www.spiegel.de/uni-spiegel/jobundberuf/0,1518,419453,00.html>) (2. Juni 2006).

Hufen, Friedhelm (1995): Rechtsfragen der Lehrevaluation an wissenschaftlichen Hochschulen, Gutachten für den deutschen Hochschullehrerverband.

Kramer, Jost W. (2005): Die Vollendung des bequemen Lebens? Anmerkungen zur Forschungstätigkeit an Fachhochschulen, in: Jost W. Kramer (Hrsg.): *Hochschulen im Spannungsfeld zwischen Lehre und Forschung, Reihe Kooperation und Wettbewerb*, Band 5, München/Mering/Hamp, 133-179.

Künzel, Rainer: 10 Jahre Evaluation von Lehre und Studium – Rückblick und Perspektiven. In: Vortrag i.R. der Tagung *Qualitätsentwicklung an Hochschulen. Erfahrungen und Lehren aus 10 Jahren Evaluation. Veranstaltung des Projekts Qualitätssicherung der Lehre der HRK*, Bonn, 4.-5. November 2005.

Lange, Stefan (2005): Hochschul-Governance im Wandel. Neuere Beiträge der vergleichenden Hochschulforschung, *Soziologische Revue* 28, 309-321.

Merton, Robert K. (1968): The Matthew-Effect in Science, *Science* 159, 53-63.

Meusel, Ernst-Joachim (1999): *Außeruniversitäre Forschung im Wissenschaftsrecht*, Köln u. a.: Heymanns.

Musil, Andreas (2005): *Wettbewerb in der staatlichen Verwaltung*, Tübingen: Mohr.

Oebbecke, Janbernd (2005): *Rechtswissenschaftliche Forschung und Verwaltung, Vorträge des Deutschen Forschungsinstituts für öffentliche Verwaltung 2*, Speyer 2005.

Rau, Einhard (1988): *Evaluation von Hochschulen – Entwicklungsstand und Perspektiven der Leistungsermittlung im Hochschulbereich*, *Wissenschaftsrecht* 21, 210-224.

Reissert, Reiner/Carstensen, Doris (1997): *Praxis der internen und externen Evaluation*. Hannover: HIS Verlag.

Schenker-Wicki, Andrea (1999): *Moderne Prüfverfahren für komplexe Probleme – Evaluation und Performance Audits im Vergleich*, Wiesbaden: Deutscher Universitäts-Verlag.

Schimank, Uwe (2004): *New Governance in*

Hochschulen und die Auswirkungen auf die Forschung, in: Hochschulrektorenkonferenz (Hrsg.), *Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen*, Bonn: Beiträge zur Hochschulpolitik, 79-87.

Schneider, Christoph (2003): *Leistungsbewertung bei der Deutschen Forschungsgemeinschaft. Vortrag bei der Sitzung der Bibliometrie-Kommission der AWMF am 16. Mai 2003*. Download unter <http://www.uni-duessel-dorf.de/AWMF/bimet/lb-dfg.htm>. [21. September 2005].

Schnitzer, Klaus/Isserstadt, Wolfgang/Müßig-Trapp, Peter/Schreiber, Jochen (1998): *Das soziale Bild der Studentenschaft in der Bundesrepublik Deutschland. 15. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem*. Herausgegeben vom Bundesministerium für Bildung und Forschung. Bonn.

Schreiber, Anja (2004): *Allein die Erfahrung zählt*, *Duz Magazin* 03/2004, 16.

Stock, Wolfgang G. (2001): *Publikation und Zitat. Die problematische Basis empirischer Wissenschaftsforschung*. *Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft* Bd. 29. Februar 2001. Download unter http://www.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1078738484koelner_ar.pdf [21. September 2005].

Viehoff, Ludger (2004): *Leistungsparameter in der Forschung – Leibniz-Gemeinschaft. In: Qualität messen – Qualität managen. Leistungsparameter in der Hochschulentwicklung. Dokumentation der gleichnamigen Tagung am 26./27. Oktober 2004 in Köln. (= Beiträge zur Hochschulpolitik 6/2005), S. 64-70*. Download unter <http://www.hrk.de/de/download/dateien/Beitrag-2005-QHerbsttagung2004.pdf>.

Viehoff, Ludger (2003): *Das Evaluierungsmodell der Leibniz-Gemeinschaft*, in: *evaNet-Reihe »Positionen«*, Mai 2003, S. 6, Download unter http://evanet.his.de/old_evamet/forum/pdf-position/viehoff15.pdf (6. Juni 2006).

Voßkuhle, Andreas (2004): *Regulierte Selbstregulierung und Organisationsentwicklung*, in: Hochschulrektorenkonferenz (Hrsg.), *Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen*, Bonn: Beiträge zur Hochschulpolitik, 7-31.

Wald, Andreas (2005): *Zur Messung von Input und Output wissenschaftlicher Produktion. Daten und Ergebnisse einer Untersuchung auf der Ebene von Forschungsgruppen*, Speyer: FÖV Discussion Papers 20.

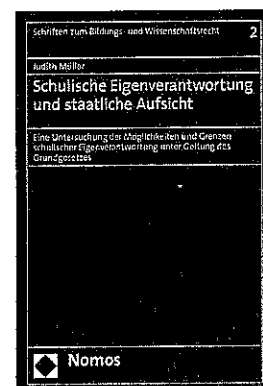
Weber, Max [1922/1996]: *Wissenschaft als Beruf*, 10. Auflage der Schrift von 1922, Berlin.

Wissenschaftsrat (2005): *Aufgaben, Kriterien und Verfahren des Evaluationsausschusses des Wissenschaftsrates*, Drs. 6966-05, Bremen, 11. November 2005/wa, download unter: <http://www.wissenschaftsrat.de/>.

Ziekow, Jan/Siegel, Thorsten (2005a): *Public-Public-Partnerships und Vergaberecht: Vergaberechtliche Sonderbehandlung der »In-State-Geschäfte«?*, *Verwaltungsarchiv*, 199-137.

Ziekow, Jan/Siegel, Thorsten (2005b): *Die Vergaberechtspflichtigkeit von Partnerschaften der öffentlichen Hand – neue Entwicklungstendenzen im Bereich der In-House-Geschäfte und der In-State-Geschäfte*, in: *Vergaberecht*, 145-156.

Neue Wege im Bildungssystem



Schulische Eigenverantwortung und staatliche Aufsicht

Eine Untersuchung der Möglichkeiten und Grenzen schulischer Eigenverantwortung unter Geltung des Grundgesetzes

Von Judith Müller

2006, 365 S., brosch., 69,-€,

ISBN 3-8329-2158-3

(Schriften zum Bildungs- und Wissenschaftsrecht, Bd. 2)

Die Verfasserin kommt zu dem Ergebnis, dass das Grundgesetz weitaus mehr Raum für eine größere schulische Eigenverantwortung lässt, als ihn die Landesgesetzgeber bislang nutzen. Insbesondere wird dargelegt, dass sowohl ein Verzicht auf die Fachaufsicht im Schulbereich als auch die Anerkennung der schulischen Eigenverantwortung als subjektiv-öffentliches Recht verfassungsrechtlich möglich sind. Ferner wird die Organisation von Schulen in anderen Rechtsformen als der bisher gebräuchlichen nichtrechtsfähigen Anstalt des öffentlichen Rechts diskutiert. Auch internationale Entwicklungen werden in den Blick genommen.

Aufgrund des großen Praxisbezuges ist das Werk sowohl für Schulpraktiker als auch für in der Schulgesetzgebung und -verwaltung Tätige von besonderem Interesse.



Nomos