

SWTR Schrift 3/2013

Leistungsmessung und Qualitätssicherung in der Wissenschaft



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Schweizerischer Wissenschafts- und Technologierat
Conseil Suisse de la Science et de la Technologie
Consiglio Svizzero della Scienza e della Tecnologia
Swiss Science and Technology Council

Der Schweizerische Wissenschafts- und Technologierat

Der Schweizerische Wissenschafts- und Technologierat SWTR berät den Bund in allen Fragen der Wissenschafts-, Hochschul-, Forschungs- und Innovationspolitik. Ziel seiner Arbeit ist die kontinuierliche Optimierung der Rahmenbedingungen für die gedeihliche Entwicklung der Schweizer Bildungs-, Forschungs- und Innovationslandschaft. Als unabhängiges Beratungsorgan des Bundesrates nimmt der SWTR eine Langzeitperspektive auf das gesamte BFI-System ein.

Le Conseil suisse de la science et de la technologie

Le Conseil suisse de la science et de la technologie CSST est l'organe consultatif du Conseil fédéral pour les questions relevant de la politique de la science, des hautes écoles, de la recherche et de l'innovation. Le but de son travail est l'amélioration constante des conditions-cadres de l'espace suisse de la formation, de la recherche et de l'innovation en vue de son développement optimal. En tant qu'organe consultatif indépendant, le CSST prend position dans une perspective à long terme sur le système suisse de formation, de recherche et d'innovation.

Il Consiglio svizzero della scienza e della tecnologia

Il Consiglio svizzero della scienza e della tecnologia CSST è l'organo consultivo del Consiglio federale per le questioni riguardanti la politica in materia di scienza, scuole universitarie, ricerca e innovazione. L'obiettivo del suo lavoro è migliorare le condizioni quadro per lo spazio svizzero della formazione, della ricerca e dell'innovazione affinché possa svilupparsi in modo armonioso. In qualità di organo consultivo indipendente del Consiglio federale il CSST guarda al sistema svizzero della formazione, della ricerca e dell'innovazione in una prospettiva globale e a lungo termine.

The Swiss Science and Technology Council

The Swiss Science and Technology Council SSTC is the advisory body to the Federal Council for issues related to science, higher education, research and innovation policy. The goal of the SSTC, in line with its role as an independent consultative body, is to promote a framework for the successful long term development of Swiss higher education, research and innovation policy.

SWTR Schrift 3/2013

Leistungsmessung und Qualitätssicherung in der Wissenschaft

Schweizerischer Wissenschafts- und Technologierat SWTR

**Zielgerichteter und vernünftiger Einsatz von Leistungsmessung und Evaluation
in der Wissenschaft – Zehn Thesen**

Vom Rat verabschiedet am 25. Juni 2013

Conseil suisse de la science et de la technologie CSST

**Pour une utilisation raisonnée de la mesure de la performance et de l'évaluation
dans le domaine de la science – Dix thèses**

Adopté par le Conseil le 25 juin 2013

Sybille Reichert

**Jenseits der Leistungsüberprüfung – Diskussionspapier zur Suche nach einem
neuen Umgang mit Qualitätssicherung an Hochschulen**

Inhaltsverzeichnis

Zusammenfassung, Résumé, Summary	4
Erster Teil – Première partie	6
Schweizerischer Wissenschafts- und Technologierat SWTR Zielgerichteter und vernünftiger Einsatz von Leistungsmessung und Evaluation in der Wissenschaft – Zehn Thesen	6
Conseil suisse de la science et de la technologie CSST Pour une utilisation raisonnée de la mesure de la performance et de l'évaluation dans le domaine de la science – Dix thèses	14
Zweiter Teil	22
Sybille Reichert Jenseits der Leistungsüberprüfung – Diskussionspapier zur Suche nach einem neuen Umgang mit Qualitätssicherung an Hochschulen	
1 Einleitung	23
2 Problemstellung	24
3 Kurze Historie der Forschungsbewertung im Kontext der Qualitätssicherung	25
4 Formen, Nutzung und Anspruchsgruppen der Forschungsbewertung	28
5 Problematik der Forschungsbewertung I: Vor- und Nachteile verschiedener Forschungsbewertungsmethoden	29
5.1 Peer Review	30
5.2 Forschungsbewertung durch quantitative Daten/Indikatoren	34
5.3 «Informed Peer Review» als Lösung und Konsens?	37
5.4 Schlussfolgerung	38
6 Problematik der Forschungsbewertung II: Zwischen Leistungsorientierung und Misstrauenskultur	38
6.1 Leistungsanreize in der Wissenschaft	39
6.2 Zwischen Nutzen und Freiheit der Wissenschaft: Brauchen wir einen neuen Contrat Social für die Wissenschaft?	41
7 Schlussfolgerungen	44
8 Literatur	46
Anhang	48
Übersicht zu Forschungsbewertung nach Nutzungsgruppen und -zielen	48
Abkürzungsverzeichnis	52

Zusammenfassung

Résumé

Summary

D

Das folgende Diskussionspapier wurde im Auftrag des Schweizerischen Wissenschafts- und Technologierats (SWTR) auf der Basis einer eingehenden Sichtung einschlägiger Forschungsliteratur zwischen Oktober 2012 und März 2013 von Dr. Sybille Reichert verfasst. Es schliesst an die Fragestellungen einer früheren Untersuchung im Auftrag des SWTR zur Quantifizierung der Forschungsbewertung (2010) an und führt die Frage nach den Grenzen der Forschungsbewertung fort, indem auch qualitative Methoden der Forschungsevaluation, insbesondere im Rahmen von Peer Review-Verfahren, ins Visier genommen werden. Dabei werden Verwendungskontexte, kognitive, psychologische und soziologische Bedingungen, Nutzen und Auswirkungen verschiedener Forschungsevaluationsmethoden auf Wissenschaftssystem und WissenschaftlerInnen untersucht. Das Diskussionspapier stützt sich dabei nicht auf eine eigene empirische Datenerhebung, sondern setzt sich kritisch-synthetisierend mit der internationalen Forschungsliteratur zu dem Thema auseinander. Da die Forschungsevaluationsmethoden international verbreitet sind, werden Untersuchungen verschiedener europäischer Wissenschaftssysteme einbezogen.

Neben einer kritischen Auseinandersetzung mit den Möglichkeiten, Grenzen und Auswirkungen gängiger Forschungsbewertungsmethoden diente die vorliegende Untersuchung dem SWTR auch als Grundlage für die Diskussion alternativer Wege der Qualitätsentwicklung und der Vertrauensbildung für die öffentliche Forschung und den sie finanzierenden Staat. Die Diskussionsergebnisse fanden Eingang in die dem Diskussionspapier vorangestellten Thesen des SWTR. Der Dank des SWTR gilt Prof. Dr. Dr. h.c. Margit Osterloh, Warwick Business School, sowie Prof. Dr. Jacques Lanarès, Universität Lausanne, für ihre wertvollen Kommentare zum vorliegenden Diskussionspapier und ihre Beiträge zur Diskussion anlässlich der Plenarsitzung des SWTR im Juni 2013.

F

Le présent document de discussion, basé sur une revue systématique de la littérature secondaire du domaine conduite entre octobre 2012 et mars 2013, a été réalisé par la Dr Sybille Reichert sur mandat du Conseil suisse de la science et de la technologie (CSST). Il fait suite à la recherche effectuée pour le CSST en 2010 au sujet de la quantification de l'évaluation de la recherche scientifique. Le document reprend la question des limites de l'évaluation, en intégrant aussi les méthodes qualitatives, comme l'évaluation par les pairs (Peer Review). L'analyse porte ainsi sur les contextes d'utilisation de différentes méthodes d'évaluation de la recherche, sur leurs conditions cognitives, psychologiques et sociologiques, ainsi que sur leurs bénéfices et effets pour le système scientifique et pour les chercheurs. Le document ne se fonde pas sur des données statistiques compilées par l'auteur, mais expose une synthèse critique de la littérature internationale à ce sujet. Compte tenu du fait que les méthodes d'évaluation de la recherche sont largement répandues, le document prend en considération un éventail de différents systèmes scientifiques en Europe.

Outre le but de fournir une analyse critique des possibilités, des limites et des effets des méthodes actuelles d'évaluation de la recherche, le présent document a également servi de base de discussion aux membres du CSST pour explorer la question des méthodes alternatives visant à assurer la qualité et à instaurer la confiance dans la recherche publique et dans son financement par l'Etat. Les résultats de cette discussion se manifestent dans les thèses formulées par le CSST, qui se situent dans la première partie de la publication. Le CSST exprime sa gratitude à la Prof. Dr Dr h.c. Margit Osterloh, Warwick Business School, ainsi qu'au Prof. Dr Jacques Lanarès, Université de Lausanne, qui ont commenté le présent document et participé à sa discussion lors de la séance plénière du CSST de juin 2013.

E

The present paper was written by Dr. Sybille Reichert on behalf of the Swiss Science and Technology Council (SSTC) and takes into account the relevant literature evaluated between October 2012 and March 2013. The paper is a follow-up to the research undertaken on behalf of the SSTC dealing with the quantification of research assessment (2010). It pursues further the question of the limits of research assessment by analysing qualitative methods of research assessment, especially in the peer-review process. The analysis takes into account the context in which they are used, cognitive, psychological and sociological conditions, as well as benefits and effects the different research assessment methods have on the science system and scientists. The author did not base her report on her own empirical data collection but critically evaluates and synthesizes the relevant international literature. As the methods of research assessment are globally applied, the paper takes into account a variety of European science systems.

The paper did not only aim at a critical analysis of the possibilities, limits and effects of the prevalent research assessment methods, but also provided the SSTC with a basis for the discussion of alternative ways of improving the quality and creating a climate of confidence within public research and the state funding this research. The discussion led to the SSTC's formulation of the theses which precede this discussion paper. Finally, the SSCT would like to express its gratitude to Prof. Dr. Dr. h.c. Margit Osterloh, Warwick Business School, and Prof. Dr. Jacques Lanarès, University of Lausanne, who participated in the council's plenary session in June 2013, commented on the discussion paper and engaged in a discussion with the SSCT's members on the subject.

Erster

Teil

Schweizerischer Wissenschafts- und Technologierat SWTR

Zielgerichteter
und vernünftiger Einsatz
von Leistungsmessung
und Evaluation
in der Wissenschaft –
Zehn Thesen

In den nachfolgenden Thesen empfiehlt der Schweizerische Wissenschafts- und Technologierat einen methodisch überlegten, sachgerechten, zielgerichteten und vernunftgeleiteten Einsatz von Leistungsmessung und Evaluation in der Wissenschaft. Zu jeder These gibt er eine kurze Erläuterung oder Begründung. Er stützt sich dabei auf den Bericht von Sybille Reichert «Jenseits der Leistungsüberprüfung – Diskussionspapier zur Suche nach einem neuen Umgang mit Qualitätssicherung an Hochschulen» vom April 2013.

1. Zeitaufwand

Die Zeit, die Forschende und Lehrende in der Schweiz für die Bewertung wissenschaftlicher Leistungen und Berichte verwenden, sollte empirisch ermittelt werden, um die Verhältnismässigkeit dieses Aufwands im Vergleich zum Zeitaufwand für die tatsächliche wissenschaftliche Arbeit in Forschung und Lehre zu beurteilen.

Nach den Beobachtungen der Wissenschaftlerinnen und Wissenschaftler sind Bewertungen wissenschaftlicher Tätigkeiten seit den 1990er Jahren immer häufiger geworden. Vermehrt müssen sie Zeit, die der Lehre und Forschung gewidmet wäre, auf die Beurteilung oder Überprüfung akademischer Leistungen verwenden. Dies hängt damit zusammen, dass immer mehr Ressourcen innerhalb wie ausserhalb der Schweiz im Wettbewerb erlangt werden. Die Entscheide über die Mittelzusprache, über Stellen, den Zugang zu Publikationsmöglichkeiten und damit auch über das Ansehen im Fach beruhen in der Regel auf *Peer Review*. Zusätzlich zum dadurch steigenden innerwissenschaftlichen Aufwand wachsen die Anforderungen, die Trägerschaften, Politik und Öffentlichkeit an das Reporting über die Leistungen der Wissenschaft stellen.

2. Aussagekraft

Die Methodenkritik und begrenzte Verlässlichkeit der Leistungsbewertungen sollten vermehrt in Wissenschaft und Öffentlichkeit kommuniziert werden. In der Regel sollte jede Bewertung wissenschaftlicher Leistungen mit einer Darlegung ihrer Datenbasis und damit ihrer Aussagekraft und Begrenztheit verbunden werden.

Sowohl die Leistungsmessung durch quantitative Indikatoren als auch die qualitative Bewertung durch auf Gutachter gestützte Evaluation (*Peer Review*) haben methodische Nachteile. Die Verlässlichkeit der Aussagen über Qualität ist beschränkt, und beide Ansätze führen zu Verhaltensanpassungen der Beurteilten, die nicht immer in eine erwünschte Richtung gehen. Im Übrigen stösst die *Peer Review* an die Grenze der Kapazität des Wissenschaftssystems. Wissenschaftsforscher und Statistiker kritisieren deshalb zunehmend die Verfahren wie auch die Verwendung der Ergebnisse solcher Beurteilungen.

3. Quantitative Leistungsindikatorik

Quantitative Leistungsindikatoren sollten stets kontextualisiert werden. In Evaluationen oder Berichten sollten sie möglichst nur als Hintergrundinformation Verwendung finden. In Fällen, in denen quantitative Bewertungen kaum Aussagekraft beanspruchen können, sollten sie gar nicht vorgesehen werden.

Die uninformierte Berücksichtigung von quantitativen Leistungsindikatoren oder von Positionen in bekannten Rankings führt zu einem verzerrten Bild wissenschaftlicher Leistungen. Wenn solche Daten direkte Auswirkungen auf die Mittelzuteilung haben, können sich daraus unerwünschte Effekte wie zum Beispiel die Vermeidung von Projekten mit unsicherem Ergebnis, die Zerstückelung der Inhalte zur Vermehrung der Anzahl Publikationen und die Bildung von Zitierkartellen ergeben.

Der Einsatz quantitativer Indikatoren vermag nur dann die wissenschaftliche Produktivität (quantitativ) zu steigern, wenn die Mittelverteilung stark von diesen abhängt. Derartige Effekte sind jedoch nicht gleichbedeutend mit gesteigerter wissenschaftlicher Qualität. Vielmehr haben bisherige Studien über die Wirkungen solcher Anreizsysteme nachteilige Einflüsse auf die Qualität aufgezeigt.

4. Grundfinanzierung und Drittmittel

Das Verhältnis zwischen der Grundfinanzierung der Forschung an Universitäten und befristeten Drittmitteln (Projektmitteln) sollte so gestaltet sein, dass sich ein Gleichgewicht zwischen langfristiger Orientierung der Forschung und dynamischer Projektentwicklung einstellt. Der Erfolg bei der Einwerbung von Drittmitteln darf nicht als Haupterfolgsmerkmal einer Person oder einer Einheit gelten, sondern muss im Zusammenhang mit anderen Erfolgsausweisen und vor dem Hintergrund der jeweiligen Disziplin und der institutionellen Rahmenbedingungen beurteilt werden.

Die Kapazität eines Wissenschaftssystems für nachhaltige Innovation hängt von der richtigen Balance zwischen frei einsetzbaren Grundbeiträgen einerseits und über den Wettbewerb eingeworbenen Drittmitteln für befristete Projekte andererseits ab. Der Anteil befristeter Drittmittel ist in den letzten Jahren zu Ungunsten der Grundbeiträge gewachsen. Zudem wird der Erfolg im Einwerben von Drittmitteln auch bei der Berechnung des Forschungsanteils an den Grundbeiträgen selbst verwendet. Dieser Drittmittelbonus verstärkt die Umlagerung der Finanzierung von längerfristig und frei verfügbaren Ressourcen zu kurzfristigen, an Projekte gebundenen Mitteln.

5. Nutzen und Frequenz von Evaluationen

Werden Evaluationen mit klar formulierten Zielen zur Identifizierung und Lösung bestimmter Probleme eingesetzt, können sie ihren spezifischen Nutzen erhalten und rechtfertigen den hiermit verbundenen Aufwand. Aber nicht alle Evaluationen sind in ihrer heutigen Form notwendig. Von einer routinemässigen, automatisierten Verwendung von Evaluationen sollte daher abgesehen werden. Es sollte stattdessen jeweils geprüft werden, ob die der wissenschaftlichen Praxis inhärente Selbstevaluation und alternative Formen der Rechenschaftsablegung (z.B. bestimmte Formen von Jahresberichten) Evaluationen ersetzen können. Jedenfalls sollte die Häufigkeit von Evaluationen an Schweizer Wissenschaftsinstitutionen dringend reduziert werden.

In empirischen Studien, aber auch von Hochschulangehörigen wird die formative Evaluation, in der *Peers* die Entwicklungs- und Verbesserungspotenziale der zu evaluierenden Einheit im Dialog mit dieser erörtern, unter allen Formen der Beurteilung am positivsten bewertet und als eigentliches Qualitätsentwicklungsinstrument geschätzt. Wird diese Art der Evaluation jedoch immer häufiger und routinemässig in von aussen auferlegten Zyklen angewandt, schwindet die Wirksamkeit und Akzeptanz auch dieses Verfahrens rasch. Der Grenznutzen zwischen Lerneffekt und Qualitätsgewinn auf der einen Seite und dem Evaluationaufwand auf der anderen sinkt mit steigender Frequenz seiner Verwendung. Nach zwei Jahrzehnten des Aufbaus und der Konsolidierung der Qualitätssicherung im Allgemeinen und der Fachevaluationen im Besonderen ist der Nutzen von Evaluationen an vielen Universitäten in der Schweiz gering geworden.

6. Nutzung von Evaluationen für die Mittelverteilung

Ein offener Dialog ist für die wirkliche Qualitätsentwicklung entscheidend. Evaluationen sollten eine offene Selbstreflexion über eigene Schwächen und Zukunftsperspektiven ermöglichen. Sie sollten deshalb nicht *automatische* Wirkungen auf die Verteilung der Mittel haben.

Als Bestandteil einer formativen Evaluation, in der Selbstevaluation und Fremdbeurteilung einen Dialog über die Potenziale einer Forschungseinheit einleiten, genießt *Peer Review* grundsätzlich immer noch ein hohes Ansehen. Wegen der begrenzten Verlässlichkeit und der möglichen nicht intendierten Effekte ist sie jedoch zu wenig tragfähig, um Mittelverteilungsentseide *unmittelbar* davon abhängig zu machen; zudem schränkt der drohende Umverteilungseffekt die Bereitschaft der Evaluierten zu einer konstruktiven Mitwirkung und zum institutionellen Lernen stark ein.

7. Intrinsische Motivation

Grundlage des Wissenschaftssystems ist eine vor allem auf intrinsischer Motivation aufbauende kulturelle Praxis. Rein kontrollierende Eingriffe sollten so weit wie möglich vermieden werden, um diese Grundlage nicht zu gefährden. Leistungsbewertungen sollten vornehmlich dann zum Einsatz kommen, wenn sie die Selbstreflexionsprozesse der Wissenschaft und ihre inhärente Qualitätsorientierung unterstützen.

Eigeninitiative, Gestaltungsfreiheit und Selbstbestimmung in den Entscheidungen über Inhalte, Ziele, Methoden und Prozesse der Forschung und Lehre sind zentrale Merkmale der wissenschaftlichen Praxis. Diese beruht wiederum auf einem hohen Mass an intrinsischer Motivation bei Wissenschaftlerinnen und Wissenschaftlern. Freiheit und Selbstbestimmung rechtfertigen für viele unter ihnen die hohen persönlichen Kosten und Unsicherheiten, die mit einer wissenschaftlichen Laufbahn verbunden sind. Ständige Berichtspflichten sowie routinisierte Begutachtungs- und Evaluationsprozesse tragen dazu bei, diese intrinsische Motivation, die hohe Leistungsbereitschaft und Risikofreudigkeit der Forschenden und Lehrenden zu erodieren. Primär auf Kontrolle ausgerichtete Formen der Leistungsbewertung in Akkreditierungen und Evaluationen schwächen deshalb die kulturelle Basis ebenso wie die eigentlichen Bedingungen wissenschaftlicher Arbeit und damit auch die Attraktivität wissenschaftlicher Laufbahnen.

8. Ex ante-Selektion und Begleitung

Hochschulen und wissenschaftsfinanzierende Institutionen sollten Methoden und Ausmass der *ex post*-Evaluation und -Leistungsbewertung als Qualitätssicherungsmassnahmen grundlegend überdenken. Vermehrt sollten sie auf effektivere Methoden der Qualitätsentwicklung abstellen. Zu diesen gehört eine verbesserte *ex ante*-Auswahl ihrer Wissenschaftlerinnen und Wissenschaftler in Berufungsverfahren, die durch Transparenz, kohärente Kriterien und Verfahren von hoher Qualität gekennzeichnet sind. Ebenso effektiv ist die langzeitige Begleitung der Lehrenden und Forschenden.

Wird die Qualität wissenschaftlicher Arbeit dadurch gesteuert, dass bei der Ausbildung und Auswahl der Wissenschaftlerinnen und Wissenschaftler *ex ante* eine gründliche Selektion gemäss den Zielen der jeweiligen Institution erfolgt, ist die Qualitätssicherung wesentlich wirksamer und das Verhältnis zwischen Aufwand und Ertrag besser als bei Versuchen, Qualität *ex post* zu kontrollieren oder Leistungen durch Anreizsysteme, die an der intrinsischen Motivation vorbeizielern, zu steigern. Zur Ergänzung der *ex ante*-Evaluation hat sich die längerfristige, kontinuierliche Begleitung der wissenschaftlichen Arbeit durch eine Gruppe von *Peers* bewährt, die einen kritischen Dialog mit den Lehrenden und Forschenden führen.

9. Vertrauen und Gestaltungsfreiheit

Leistungsförderung sollte vor allem über die Verstärkung der bereits gegebenen Motivation durch einen Vertrauensvorschuss erfolgen, d.h. Gestaltungsfreiheit sollte gewährt und vielversprechende Entwicklungen und Initiativen sollten gestärkt werden.

Ein auf Bestätigung und stetige «Produktion» ausgerichteter Wissenschaftsbetrieb droht langfristig sklerotisch zu werden. Die sich ausbreitende Misstrauenskultur und das zunehmend dichter werdende Netz von Kontrollmassnahmen fördern Mittelmass statt Exzellenz.

10. Ermöglichungskultur

Eine durch gegenseitiges Vertrauen und dichten Dialog gekennzeichnete Ermöglichungskultur sollte die zunehmende Tendenz von Staat und Gesellschaft ersetzen, die Wissenschaft einer detaillierten Kontrolle zu unterwerfen.

Staat und Politik entwickeln im Gegenzug zur Gewährung grösserer Autonomie oft einen verstärkten Kontrollwunsch gegenüber den Hochschulen. Die Lehrenden und Forschenden müssen deshalb vermehrt in der Öffentlichkeit über ihr Tun berichten, von sich aus den Dialog suchen und damit den Kulturwandel auch auf Seiten der Hochschulträger fördern.

Wenn jedoch das New Public Management bestrebt ist, auch in der Wissenschaft Effizienz über Leistungsanreize und -kontrollen zu steigern, verkennt es die Natur wissenschaftlichen Arbeitens. Unvorhersehbare Entwicklungen, Umwege, Versuche mit negativen Ergebnissen und das Risiko des Scheiterns einer Idee oder der Falsifikation einer Hypothese sind Vorbedingungen des Fortschritts und der anhaltenden Innovation.

Première

Conseil suisse de la science et de la technologie CSST

partie

Pour une utilisation
raisonnée de la mesure
de la performance
et de l'évaluation dans
le domaine de la science –
Dix thèses¹

¹ Traduit de l'allemand.

Dans les thèses ci-après, le Conseil suisse de la science et de la technologie (CSST) recommande une utilisation méthodique, pertinente, ciblée et raisonnable de la mesure des performances et de l'évaluation dans le domaine scientifique. Chaque thèse est assortie d'une brève explication. Les thèses se fondent sur le rapport présenté au CSST en avril 2013 par Sybille Reichert et intitulé «Au-delà de la *mesure de la performance*. Contribution à la recherche de nouvelles pistes pour le développement de l'assurance qualité dans les Hautes écoles».

1. Proportionnalité du temps investi

Le temps consacré par les chercheurs et les enseignants en Suisse pour l'évaluation de la performance scientifique et l'établissement de rapports doit être calculé de manière empirique afin de pouvoir estimer la proportionnalité de cette charge de travail par rapport aux travaux scientifiques effectifs dans la recherche et dans l'enseignement.

De l'avis des scientifiques eux-mêmes, les évaluations d'activités scientifiques ont pris de l'ampleur depuis les années 1990. Ils doivent consacrer de plus en plus de temps à de telles évaluations et à des contrôles au détriment des activités de recherche et d'enseignement. Cette évolution est liée au fait que l'acquisition de ressources s'effectue de plus en plus souvent par le biais d'une procédure compétitive, en Suisse comme à l'étranger. En règle générale, les mécanismes pour l'attribution de moyens et de postes, pour l'accès aux possibilités de publications ainsi que pour la définition de la renommée des chercheurs dans un domaine reposent sur la revue par les pairs (*Peer Review*). En parallèle à la charge de travail générée par cette évolution des pratiques scientifiques, les exigences des collectivités responsables augmentent, de même que celles des milieux politiques et de l'opinion publique, lesquels attendent des comptes sur la performance scientifique.

2. Pertinence des résultats

La critique méthodologique et la fiabilité relative de l'évaluation de la performance scientifique devraient davantage être communiquées au sein du monde scientifique et de l'opinion publique. D'une manière générale, chaque évaluation de prestations scientifiques devrait présenter les données sur lesquelles elle repose et fournir ainsi des indications sur sa pertinence et ses limites.

Tant la mesure de la performance scientifique fondée sur des indicateurs quantitatifs que l'évaluation qualitative par des experts (*Peer Review*) présentent des inconvénients sous l'angle méthodologique. La fiabilité de l'évaluation de la qualité est limitée et les deux approches induisent une adaptation du comportement des personnes évaluées qui ne va pas toujours dans la direction souhaitée. Le *Peer Review* se heurte par ailleurs aux limites de capacité du système scientifique. Les travaux issus de la recherche sur la science et sur la statistique remettent de plus en plus en question les méthodes et l'utilisation des résultats de telles évaluations.

3. Indicateurs quantitatifs de la performance

Les indicateurs quantitatifs de la performance scientifique doivent toujours être considérés en fonction de leur contexte de production. Dans le cadre d'évaluations ou de rapports, ils ne devraient, dans la mesure du possible, être utilisés qu'au titre d'informations relatives au contexte. Le recours à de tels indicateurs quantitatifs est en outre à bannir dans les cas où ils n'apportent que peu d'informations pertinentes.

La focalisation irréfléchie sur les indicateurs quantitatifs de la performance ou sur la position dans les rankings connus donne une fausse image des prestations scientifiques. Si de telles données ont des répercussions directes sur l'octroi de moyens financiers, des effets indésirables sont à craindre tels que, par exemple, le renoncement à des projets dont les résultats sont incertains, la dispersion des contenus en vue de multiplier le nombre de publications ou encore l'encouragement de cartels de citation.

L'utilisation d'indicateurs quantitatifs ne favorise la productivité scientifique sous l'angle quantitatif que si les moyens financiers dépendent fortement de ces indicateurs. De tels effets n'ont toutefois pas la même portée que le renforcement de la qualité scientifique en elle-même. Les études effectuées à ce jour sur les conséquences de tels systèmes d'incitation ont d'ailleurs démontré leurs influences négatives sur la qualité.

4. Financement de base et fonds de tiers

Le rapport entre le financement de base de la recherche dans les Hautes écoles et les fonds de tiers limités dans le temps (fonds de projet) devrait permettre un équilibre entre l'orientation de la recherche sur le long terme et une évolution dynamique des projets. Le succès dans l'acquisition de fonds de tiers ne doit pas constituer le critère majeur à l'actif d'une personne ou d'une unité, mais doit être considéré par rapport aux autres facteurs de succès et situé dans le contexte de la discipline concernée et des conditions-cadre institutionnelles.

L'aptitude d'un système scientifique à générer de l'innovation durable passe par un juste équilibre entre, d'une part, des financements de base dont l'utilisation est libre et, d'autre part, des fonds de tiers acquis de manière concurrentielle pour des projets de durée limitée. La part des fonds de tiers de durée limitée a augmenté au cours des dernières années au détriment des financements de base. Le succès obtenu dans l'acquisition de fonds de tiers est par ailleurs pris en compte lors du calcul de la part des financements de base en faveur de la recherche. Cette prééminence des fonds de tiers accentue encore davantage la tendance à remplacer des ressources disponibles librement sur le long terme par des moyens financiers à court terme et liés à des projets spécifiques.

5. Utilité et fréquence des évaluations

Les évaluations assorties d'objectifs clairs tels que l'identification et la résolution de problèmes précis justifient les investissements ainsi effectués. Or, dans leur forme actuelle, toutes les évaluations ne sont pas nécessaires. Il convient de renoncer à une utilisation routinière et automatisée de ces instruments et d'examiner, au cas par cas, si l'auto-évaluation inhérente à la pratique scientifique et des formes alternatives de compte rendu telles que des rapports annuels peuvent s'y substituer. La fréquence des évaluations au sein des institutions scientifiques suisses doit dans tous les cas être réduite.

Selon les études empiriques, ainsi que de l'avis du personnel des Hautes écoles, la forme d'évaluation la plus appréciée, et qui sert véritablement comme instrument de développement de la qualité, est l'évaluation formative. Celle-ci permet aux pairs de déterminer, par le biais d'un dialogue avec les personnes concernées, le potentiel de développement et d'amélioration de l'unité évaluée. Or, si l'évaluation formative est utilisée de plus en plus fréquemment et de manière routinière selon des cycles déterminés de l'extérieur, son efficacité et son acceptation en souffriront. La frontière entre les gains en enseignements et en qualité d'une part, et les investissements liés à l'évaluation, d'autre part, se réduira en fonction de la fréquence du recours à de telles évaluations. Au terme de deux décennies de mise en place et de consolidation de l'assurance qualité d'une manière générale et des expertises techniques en particulier, le bénéfice tiré des évaluations s'est réduit pour de nombreuses Hautes écoles suisses.

6. Recours aux évaluations pour l'attribution de moyens financiers

L'efficacité du développement de la qualité présuppose un dialogue ouvert. Les évaluations devraient permettre une auto-réflexion ouverte sur les propres faiblesses ainsi que sur les perspectives d'avenir. Elles ne devraient pas, de ce fait, automatiquement déterminer l'attribution des moyens financiers.

Lorsqu'elle fait partie d'une évaluation formative, au sein de laquelle l'auto-évaluation et l'appréciation par des tiers conduit à un dialogue sur le potentiel de développement d'une unité de recherche, l'évaluation formative est largement appréciée. En raison de sa fiabilité restreinte et de ses possibles effets indésirables, elle n'est cependant pas appropriée pour influencer *directement* sur des décisions d'attribution de moyens financiers. En outre, la crainte d'une redistribution des fonds restreint la disposition des personnes concernées par l'évaluation à participer de manière constructive à celle-ci et à en tirer des enseignements pour l'institution.

7. Motivation intrinsèque

Le système scientifique repose essentiellement sur une pratique culturelle fondée sur la motivation intrinsèque. Des interventions servant uniquement à effectuer des contrôles doivent, dans la mesure du possible, être évitées afin de ne pas compromettre ces fondements. Le recours à des évaluations de la performance ne devrait se faire que dans la mesure où ces évaluations encouragent le processus d'auto-réflexion sur le plan scientifique et soutiennent l'approche qualitative inhérente.

L'esprit d'initiative, la liberté de conception et d'organisation ainsi que l'auto-détermination dans les choix des contenus, des objectifs, des méthodes et des processus, aussi bien dans la recherche que dans l'enseignement, sont les caractéristiques majeures de la pratique scientifique. Une telle approche se fonde à son tour sur une forte motivation intrinsèque de la part des scientifiques concernés. Pour nombre d'entre eux, la liberté et l'auto-détermination justifient le grand investissement personnel ainsi que les incertitudes inhérentes à une carrière scientifique. L'obligation constante de devoir rendre des comptes, tout comme les processus routiniers d'expertise et d'évaluation, entraînent une baisse de cette motivation intrinsèque, de la performance et du goût du risque auprès des chercheurs et des enseignants. L'appréciation des prestations axée principalement sur le contrôle, que cela soit dans le cadre de procédures d'accréditation ou d'évaluation, érode non seulement la base culturelle, mais aussi les conditions à proprement parler du travail scientifique et, partant, réduit l'attrait des carrières scientifiques.

8. Sélection *ex ante* et encadrement

Les Hautes écoles et les agences de moyens de la science doivent entièrement reconsidérer leurs méthodes et l'ampleur des évaluations des prestations *ex post* en tant que mesures de l'assurance qualité. Elles devront de plus en plus s'appuyer sur des méthodes efficaces de développement de la qualité telles qu'une meilleure sélection *ex ante* des scientifiques lors des procédures de nomination, procédures qui doivent se distinguer par leur transparence, par le recours à des critères cohérents et par leur qualité élevée. L'encadrement longitudinal des enseignants et des chercheurs s'inscrit également dans ce contexte.

Si la qualité des prestations scientifiques passe par une meilleure sélection préalable des scientifiques, sélection qui sera fondée sur les objectifs de l'institution concernée, la qualité sera mieux assurée. Par ailleurs, le rapport coûts/bénéfices sera davantage optimisé que dans le cadre des tentatives qui visent à contrôler la qualité *ex post* ou à améliorer les prestations à travers des systèmes d'incitation qui ne tiennent pas compte de la motivation intrinsèque. En complément à l'évaluation *ex ante*, l'accompagnement à long terme et continu des travaux scientifiques par un groupe de pairs menant un dialogue critique avec les enseignants et les chercheurs a fait ses preuves.

9. Confiance et autonomie – la liberté de conception et d'organisation

L'encouragement de la performance devrait essentiellement découler d'un renforcement de la confiance envers la motivation pré-existante des scientifiques, notamment en leur garantissant la liberté de conception et d'organisation, et en encourageant les initiatives et les développements prometteurs.

Un système scientifique axé sur la «confirmation» et sur la «production» risque de se scléroser. La culture de la méfiance qui tend à se propager et le réseau de plus en plus dense de mesures de contrôle encouragent non pas l'excellence, mais la médiocrité.

10. Culture des possibles

Une culture des possibles empreinte de confiance mutuelle et d'un dialogue soutenu est appelée à remplacer la tendance croissante de l'Etat et de la société à soumettre la science à des contrôles détaillés.

En contrepartie de l'autonomie plus ou moins importante concédée aux Hautes écoles, l'Etat et les milieux politiques expriment souvent le souhait d'un contrôle renforcé. Les enseignants et les chercheurs doivent de ce fait de plus en plus rendre compte de leurs activités auprès du grand public, rechercher eux-mêmes le dialogue et, partant, promouvoir un changement de culture également auprès des collectivités responsables des Hautes écoles.

Or, si la Nouvelle gestion publique s'efforce de renforcer l'efficacité dans le domaine scientifique par des contrôles et des incitations à améliorer la performance, elle méconnaît la nature même du travail scientifique, tant il est vrai que les développements imprévisibles, les détours, les tentatives menant à des résultats négatifs tout comme le risque d'échouer et de ne pas voir aboutir une idée, ainsi que la falsification d'une hypothèse sont le terreau du progrès et de l'innovation perpétuelle.

Zweiter

Teil

Sybille Reichert

Jenseits der Leistungs-
überprüfung –
Diskussionspapier
zur Suche nach einem
neuen Umgang
mit Qualitätssicherung
an Hochschulen

«Externally and internally driven research assessment has incontestably become part and parcel of the research university enterprise. However, the explosion of various types of research assessment for a variety of users and purposes in recent years risks to create – or, one may argue, has already started to create – an obsession with measurement and monitoring, which may result in a ‹bean counting› culture detracting from the real quality of research and the boundless search for new knowledge. [...]

The demands of producing large quantities of data may create or extend a culture of regulations, instructions, lists of good practice etc., marked by an obsession with measurement and monitoring, all of which might well detract from creative freedom, flexibility and productivity.»

(The League of European Research Universities, LERU 2012)

1 Einleitung

Das folgende Diskussionspapier wurde im Auftrag des Schweizerischen Wissenschafts- und Technologierats (SWTR) auf der Basis einer eingehenden Sichtung einschlägiger Forschungsliteratur zwischen Oktober 2012 und März 2013 verfasst. Es schliesst an die Fragestellungen einer früheren Untersuchung im Auftrag des SWTR zur Quantifizierung der Forschungsbewertung (2010) an und führt die Frage nach den Grenzen der Forschungsbewertung fort, indem auch qualitative Methoden der Forschungsevaluation, insbesondere im Rahmen von *Peer Review*-Verfahren, ins Visier genommen werden. Dabei werden Verwendungskontexte, kognitive, psychologische und soziologische Bedingungen, Nutzen und Auswirkungen verschiedener Forschungsevaluationsmethoden auf Wissenschaftssystem und WissenschaftlerInnen untersucht. Das Diskussionspapier stützt sich dabei nicht auf eine eigene empirische Datenerhebung, sondern setzt sich kritisch-synthetisierend mit der internationalen Forschungsliteratur zu dem Thema auseinander. Da die Forschungsevaluationsmethoden international verbreitet sind, werden Untersuchungen verschiedener europäischer Wissenschaftssysteme einbezogen. Ziel der Untersuchung ist nicht nur die kritische Auseinandersetzung mit den Möglichkeiten, Grenzen und Auswirkungen gängiger Forschungsbewertungsmethoden, sondern auch die Schaffung einer Grundlage für die Diskussion alternativer Wege der Qualitätsentwicklung und der Vertrauensbildung für die öffentliche Forschung und den sie finanzierenden Staat.

2 Problemstellung

Im Zuge wachsender Hochschulautonomie nahmen im letzten Jahrzehnt die Rechenschaftspflichten der öffentlichen Hochschulen gegenüber ihren Geldgebern zu. Politische Entscheidungsträger und Öffentlichkeit erwarten von einer in ihren Entscheidungen mehr oder weniger frei handelnden Hochschule, dass sie diese nicht nur begründen, sondern auch deren Wirkungen nachweisen kann. In ganz Europa wurden im Rahmen der Erweiterung der Hochschulautonomie neue Steuerungsinstrumente und ein immer differenzierteres Berichtswesen entwickelt, um Parlamenten und Behörden zu beweisen, dass die Entlassung der Hochschulen aus ihrer staatsverschuldeten Unmündigkeit tatsächlich eigenverantwortliches institutionelles und individuelles Handeln zur Folge hat und gleichzeitig die Verwirklichung öffentlicher Ziele erlaubt. So entstanden in den letzten Jahrzehnten Prozesse der Strategiebildung, Entscheidungsfindung und Qualitätssicherung, um Zielsetzungen und deren Umsetzung möglichst weitsichtig, nachhaltig, nachvollziehbar und überprüfbar zu machen. Der Zuwachs an institutioneller Autonomie ist somit notwendig an neue Steuerungsformen und Berichtspflichten im Sinne des Grundsatzes der «accountability» gekoppelt, der häufig die individuelle Autonomie der Forschenden und Lehrenden einschränkt. So stellt sich zunehmend die Frage, ob und inwieweit diese Einschränkung der Wissenschaft mehr nützt als schadet oder umgekehrt.

Selbst wenn der Grundsatz der «accountability» bis heute kaum bestritten ist, scheiden sich doch die Geister hinsichtlich der Effektivität der Formen und Grade seiner Umsetzung, vor allem was die Steuerungsinstrumente und Überprüfungsverfahren angeht. Die Wunschvorstellung einer verlässlichen Darstellung und Beurteilung von wissenschaftlichen Leistungen zum Zwecke gerechter Mittelzuweisung und Qualitätsentwicklung scheint zum einen nur unzureichend realisierbar zu sein und zum anderen mit bedenklichen Begleiterscheinungen einer ausufernden Berichts- und Gutachtenspraxis einherzugehen. Daher mehren sich in verschiedenen Ländern kritische Stimmen, die eine Überprüfung der Überprüfung einfordern, um den Mehrwert der neuen Steuerungs-

instrumente abzuschätzen und gegen die materiellen und immateriellen Kosten und Auswirkungen auf die Wissenschaft und WissenschaftlerInnen abzuwägen. In Großbritannien wurde die Frage der während der letzten zwei Jahrzehnte zunehmend ausufernden Forschungsbewertung zu einer der zentralen wissenschaftspolitischen Kontroversen der letzten Jahre. Während einige Länder wie Norwegen, Deutschland und Frankreich dabei sind, national vergleichende Fachevaluationen und institutionelle Forschungsbewertungen zur Strategiebildung nach dem Vorbild der Niederlande und Großbritanniens erst einzuführen, stellt sich in Letzteren die Frage, wie man die mit Leistungsbewertung verbundenen Nebenwirkungen durch Anpassung der Methodik lindern kann. Dabei stehen sowohl die quantitativen indikatorenbasierten Leistungsbewertungen als auch die qualitativen, auf Gutachter gestützten Evaluationsmethoden (*Peer Review*) aus unterschiedlichen Gründen im Kreuzfeuer der Kritik, wie im Abschnitt 5 erläutert wird.

Jenseits dieser pragmatischen Fragen nach Aufwandsbegrenzung, Indikatorenauswahl, Evaluationsformen und -zyklen lässt sich die Diskussion über Wissenschaftsbewertung aber auch auf einer grundsätzlichen Ebene führen. Gerade in den reiferen, d.h. durch Jahrzehnte der Leistungsbewertung und Qualitätsentwicklung geprüften Systemen stellt sich die Frage, inwieweit das gewachsene und stetig weiter wachsende Instrumentarium der Bewertung und Qualitätssicherung eine Misstrauenskultur, welche die Fundamente einer innovativen, kreativen Wissenschaftslandschaft langfristig untergräbt, nicht nur ausdrückt, sondern auch verstetigt. Zu fragen ist somit nach Alternativen zu einem solchen von Misstrauen getragenen Rechenschaftssystem. Könnte sich nicht gerade ein durch international herausragende Forschungsleistungen geprägtes Hochschulsystem wie das der Schweiz eine Vertrauenskultur leisten, in der Zeit gespart und Leistungen und Kreativität gefördert würden – durch andere Formen der Rechenschaftslegung und sinnvollere Formen der Qualitätsentwicklung, die weniger kontrollierend als vielmehr entwicklungsorientiert wirkten?

Die folgenden Ausführungen sollen diese zwei Ebenen der Befragung systematisch aufbauen. Nach einer kurzen historischen Einführung folgt eine Diskussion der zentralen Aspekte, Erfahrungen und em-

pirischen Studien zu verschiedenen Formen der Wissenschaftsbewertung, aus denen jeweils Kernthesen abgeleitet werden, die als Grundlage für eine weiterführende Diskussion im SWTR dienen sollen. In dieser soll es weniger um eine Anpassung bestehender Praktiken der Forschungsbewertung gehen, als um die Suche nach neuen Ansätzen, die das Vertrauen der wissenschaftsfinanzierenden Politik und Öffentlichkeit in die Leistungsfähigkeit und Sinnhaftigkeit wissenschaftlicher Praxis nachhaltig und kontinuierlich zu festigen vermögen. Dahinter steht die Vision einer Schweizer Wissenschaftslandschaft, in der ein lebhafter inhaltsbezogener Dialog zwischen Wissenschaft und ihrem gesellschaftlichen und wirtschaftlichen Umfeld den Bedarf zeitraubender bürokratisch-invasiver Leistungskontrollen minimiert und somit den WissenschaftlerInnen mehr Zeit für Lehre, Forschung und den Austausch mit der wissenschaftsinteressierten Umwelt lässt. Noch nie war unsere Gesellschaft so wissenschaftsgetrieben und -abhängig wie heute. Und doch ist das Verhältnis der Gesellschaft und Politik zu ihren wissenschaftlichen Institutionen zunehmend von inhaltsleerer Kontrolle statt inhaltlichem Dialog bestimmt. Ein solcher Dialog sollte Bezug auf die Neuerungen und Bedingungen der Wissenschaft nehmen und dadurch auch Bedingungen der Wissenschaft mitbestimmen.

3 Kurze Historie der Forschungs- bewertung im Kontext der Qualitätssicherung

Seit den 1980er Jahren wurde die Forschung zunächst in den angelsächsischen Ländern und den Niederlanden sowie später im restlichen Kontinentaleuropa vermehrt unter dem Blickwinkel der Leistungsorientierung und Effizienzsteigerung umfassenden Prüfungen unterzogen. Diese wurden aus verschiedenen Gründen für notwendig befunden:

1. Mit der Erweiterung der Hochschulautonomie und Einführung des New Public Management in der Wissenschaftspolitik zunächst in der angelsächsischen Welt, in der Folge in den Niederlanden und seit den 1990er Jahren in der Schweiz, zogen erweiterte Berichtspflichten und leistungsorientierte Steuerungsinstrumente in die Wissenschaftslandschaft ein, in der Annahme, dass wissenschaftliche Leistungen erfasst, geprüft und belohnt werden können und so die staatliche Feinsteuerung ersetzt werden kann. Michael Powers Charakterisierung der britischen Gesellschaft als «audit society» (Power 1997) beschreibt, wie alle Bereiche der Gesellschaft mit Ritualen der Überprüfung und Evaluationen sowie mit leistungsabhängiger Ressourcenzuweisung überzogen wurden, um «value for money», «accountability» und Leistungsdynamik zu gewährleisten. Das angelsächsische Vorbild breitete sich nach und nach in Europa aus. Der Rückzug des Staates aus der Feinsteuerung des institutionellen Alltags wurde kompensiert durch die Ausweitung staatlicher Fernsteuerung durch neue Instrumente der Leistungsüberprüfung, -bewertung und -steuerung «from interventionist to «evaluative» governance» (Ferlie 2012; siehe auch Moscati 2012; Paradeise 2012; Paradeise et al. [Hrsg.] 2009). Der Zuwachs institutioneller Autonomie ging insofern Hand in Hand mit einer Einschränkung der individuellen Autonomie der WissenschaftlerInnen, deren Handlungen einer wachsenden Anzahl externer Steuerungsversuche unterlagen (Moscati 2012; Pechar 2012).

2. Im Zuge der Globalisierung gewann die Forschung als Motor der wirtschaftlichen Entwicklung in Hochlohnländern an Gewicht, da Wettbewerbsnachteile durch hohe Lohnkosten nur mit erhöhter Produktivität und somit durch Forschungs- und Technologievorsprung ausgeglichen werden können. Mit dem wachsenden Vertrauen in die Wertigkeit von Wissenschaft und Forschung schwand aber gleichzeitig das Vertrauen in ihre Selbstregulierungsfähigkeit. «Indeed, it is precisely because it [science, *Anm. der Autorin*] is such a significant policy object that its fortunes can no longer be entrusted to scientists alone. Macroeconomic strategy in many countries demands that science be made governable» (Power 2007).
3. Zur starken Gewichtung der Forschung und Betonung der Leistungsorientierung gehörte auch das erstarkte Wettbewerbsprinzip, welches sich vor allem in der relativen Zunahme der im Wettbewerb eingeworbenen Drittmittel gegenüber den über Hochschulbudgets vom Träger direkt zugewiesenen Forschungsgeldern äußerte. Dieser Anteil verdoppelte sich in einigen Ländern in nur einem Jahrzehnt, womit die Anzahl der in Begutachtung involvierten GutachterInnen ebenfalls entsprechend stieg.
4. Mit der rapiden Expansion des Hochschulsektors und der wachsenden Durchlässigkeit der Bildungselite wurden im Zuge der unzureichend finanzierten Massifizierung die Rufe nach interner Differenzierung an den Hochschulen lauter, zunächst aufseiten der Nutzer und der Öffentlichkeit, dann auch aufseiten der Träger. Die Identifizierung und Förderung besonders leistungsstarker Personen (Studierende und WissenschaftlerInnen) und Einheiten wurden mit wachsendem Interesse von Trägern, Förderorganisationen und Nutzern verfolgt. Die Situation an den Hochschulen drängte auf Reformen, allgegenwärtige Missstände zogen Leistungsüberprüfung und nach Leistung differenzierende Mittelvergabe nach sich. Die veränderten Rahmenbedingungen legten eine verstärkte Erfolgskontrolle nahe (Weingart 1995). Während Deutschland und Österreich von einer systematischen Einführung von Evaluierungsmethoden 1995 noch weit entfernt waren und sich die Information für die Öffentlichkeit und wissenschaftspolitischen Instanzen ebenso wie die Mittelzuteilung zunächst auf Inputindikatoren beschränkten, wurden schon wenige Jahre später Output-/Leistungsindikatoren-basierte Institutionenvergleiche und systematische Qualitätssicherungsinstrumente propagiert und eingeführt (Müller-Böling 1995). In der Folge setzte in Deutschland ein Bundesland nach dem anderen Evaluationskommissionen und Strukturkommissionen ein, in denen die Leistungen der Landeshochschulen durch quantitative Daten und Expertenbefragungen eingeschätzt und entsprechende Restrukturierungs- und Finanzierungsentscheide abgeleitet werden sollten. Des Weiteren wurden innerhalb von Hochschulen erste Modelle der leistungsbezogenen Mittelvergabe erprobt.
5. In der Schweiz breiteten sich Qualitätssicherung und entsprechende Evaluationsinstrumente bereits früher aus als in den anderen deutschsprachigen Ländern. So wurden systematische Studierendenbefragungen zur Lehre in den meisten Universitäten bereits seit den frühen 1990er Jahren angewendet. Zur Qualität der Forschung wurden vom Schweizerischen Wissenschaftsrat seit den 1980er Jahren systematische bibliometrische Untersuchungen durchgeführt. Institutionelle Fachevaluationen wurden zunächst im ETH-Bereich erprobt, zuerst 1989 im Departement Mathematik der ETH Zürich im Auftrag der Oberbehörde, des ETH-Rats, und dann seit Anfang der 1990er Jahre an der ETH Zürich und Lausanne im Auftrag der Hochschulleitungen für alle Departemente durchgeführt (Handbuch zum Qualitätsmanagement, ETH Zürich, 2003) – neben anderen Instrumenten der Qualitätssicherung. Die Universität Zürich folgte dem Modell und führte eine eigene Qualitätsagentur ein (2001), um Evaluationen auf allen Ebenen durch *Peer Review* und mit wissenschaftlicher Begleitung vorzunehmen. Die institutionalisierte Qualitätssicherung durch Evaluation ist an den ETHs mit dem ETH-Gesetz vom 4. Oktober 1991 und an der Universität Zürich mit dem Gesetz über die Universität Zürich vom 15. März 1998 gesetzlich verankert worden. Andere Universitäten und Träger folgten dem Modell der systematischen institutionellen Evaluation und führten koordinierende Stellen in ihren Hochschulen ein. Output- bzw. Leistungsindikatoren bei der Mittelvergabe und an Zielverein-

barungen gekoppelte Globalhaushalte wurden von verschiedenen Trägern seit der zweiten Hälfte der 1990er Jahre eingeführt; auf Bundesebene wurden sie seit 1999 auf der Grundlage des Universitätsförderungsgesetzes (UFG 1999, Art. 15) angewendet. Seit 2006 schreiben nationale Guidelines sozusagen als nachträgliche Besiegelung periodische interne Evaluationen für alle Universitäten vor (SUK 2006).

6. Im Dickicht der massifizierten Hochschullandschaft brauchten StudienbewerberInnen und andere Interessierte Orientierungs- und Entscheidungshilfen, vor allem um leistungsstarke Inseln vorteilhafter Studienbedingungen zu identifizieren. So wurden Leistungsvergleiche zwischen Hochschulen in der Öffentlichkeit populär. Dabei waren vor allem die Medien die Vorreiter, zunächst die *Spiegel*-Umfragen zur Attraktivität des Studiums (1993) und die Untersuchungen von *Focus* und *Bild der Wissenschaft* zu Forschungsleistungen ganzer Institutionen im Vergleich (1993). Später wurde vom *Centrum für Hochschulentwicklung* (CHE) die Ranking-Methodologie zur Erleichterung der Studienplatzwahl eingeführt und danach in Kooperation mit der *Zeit* vermarktet. Im Zuge der Globalisierung und erhöhten Mobilität der Studierenden, Graduierten und Forschenden kamen seit 2004 mit dem *Times Higher Education Supplement Ranking* und dem *ARWU Ranking* der Shanghai Jiao Tong University die internationalen Rankings hinzu. Inzwischen blicken wir auf zwei Jahrzehnte immer wiederkehrender nationaler und internationaler Institutionenvergleiche zurück, die der Aufklärung von Nutzern dienen sollen, dies wegen methodischer Mängel allerdings nur bedingt leisten können (Rauhvargers 2011).
7. Zur Ausweitung der Evaluationsaktivitäten im Hochschulwesen trägt zudem die Homogenisierung der Qualitätssicherung im Rahmen der Bologna-Reformen bei, die gemeinsame prozedurale Standards der Qualitätssicherung durch die *European Standards and Guidelines for Quality Assurance* und die Registrierung standardkonformer Evaluations- und Akkreditierungsagenturen im *European Quality Assurance Register for Higher Education* (www.eqar.eu) einführten. Durch den im Rahmen der Bologna-Reformprozesse erstarkten

innereuropäischen Dialog hochschulpolitischer Akteure verbreiten sich die als fortschrittlich angesehenen Evaluationspraktiken umso schneller in ganz Europa. Die «Trends»-Studien zur Umsetzung der Bologna-Reformen an Hochschulen belegen die Wichtigkeit der Qualitätssicherungsreformen in Europa: «Institutions find quality assurance reforms to have been among the most important developments that have shaped their strategy in recent years, since 63% of the respondents to Trends mentioned it among the top three developments» (Loukkola & Zhang 2010:12; EUA 2010). Ein Großteil der Universitäten berichtet, dass die internen Qualitätssicherungssysteme im letzten Jahrzehnt eingerichtet wurden und dass diese auf die Erfordernisse der nationalen Qualitäts- oder Akkreditierungsagenturen eingehen (Loukkola & Zhang 2010:23). Die Einrichtung von externen und internen Qualitätssicherungssystemen gehört neben den neuen Studienstrukturen zu den wichtigsten Systemneuerungen der Bologna-Reformen. Als externe Evaluatoren werden sowohl Fachexperten (innerhalb von Fachevaluationen) als auch Management- und QS-Experten (im Rahmen von Audits und Akkreditierungen) eingesetzt.

Dank der erwähnten hochschulpolitischen Kontexte und der mit ihnen verbundenen Begründungsdiskurse breitete sich die in den 1990er Jahren entstandene Evaluations- und Leistungsüberprüfungsbewegung nicht nur europaweit, sondern auch über eine wachsende Anzahl von Handlungsfeldern aus. Gepeist «aus einer Situation stagnierender Mittel, politisch eingeklagter Rechenschaftslegung für die Verwendung öffentlicher Mittel in der Wissenschaft und einer massenmedial transportierten Besorgnis um nachlassende Wettbewerbs- und Reformfähigkeit» (Hornbostel 2008:74) wurden Evaluationen und Leistungsüberprüfungen nicht nur der Lehre und Forschung, sondern auch in Bezug auf Transferleistungen, Internationalisierung, Kooperation mit außeruniversitären Einrichtungen und Unternehmen, Genderfragen, Marketing und Sichtbarkeit, Hochschulzugang etc. ausgeweitet. Erweitert wurde zudem der Kreis derjenigen, die mit der Sammlung und Verbreitung der Informationen befasst sind.

4 Formen, Nutzung und Anspruchsgruppen der Forschungsbewertung

Die oben aufgeführten Begründungszusammenhänge führen zu einer Vielfalt von Leistungsüberprüfungen und Begutachtungen, die den Berichts- und Begutachtungsaufwand der WissenschaftlerInnen insgesamt erheblich steigern.

Zu diesen gehören wissenschaftssystemimmanente Leistungsbeurteilungen und Begutachtungen wie:

1. die Antrags- und Berichtspflichten sowie die Begutachtung im Rahmen von Drittmittelförderung von Forschungsprojekten (Einzel- und Verbundförderung), seit einigen Jahren auch vermehrt für hochschulinterne Forschungsprojekte;
2. die Begutachtung für Veröffentlichungen in wissenschaftlichen Zeitschriften;
3. die externe und hochschulinterne Begutachtung bei Berufungen;
4. die Beantwortung von Umfragen im Rahmen von Hochschulstudien;

sowie vom Wissenschaftssystem durch externe Akteure wie Politik und Geldgeber oder andere Anspruchsgruppen eingeforderte Leistungsüberprüfungen wie:

5. die Rechenschaftslegung, z.B. Berichte zuhänden der Universitätsleitungen (Jahresberichte) oder anderer Träger oder Geldgeber als Teil der Leistungskontrolle bei Zielvereinbarungen;
6. die Begutachtung von Forschungseinheiten (Departemente oder Fakultäten) und Forschungsgruppen im Rahmen von hochschulinternen oder landesweit verordneten Fachevaluationen;
7. die Sammlung von Forschungsoutputdaten als Indikatoren zuhänden der hochschulinternen oder externen Mittelvergabe;
8. die Sammlung von Forschungsoutput- und Steuerungsprozessdaten zuhänden von externen Akkreditierungs- oder Qualitätsagenturen;
9. die Beantwortung von Umfragen zuhänden von Rankinginstitutionen;
10. die Evaluation von Forschungsprogrammen und deren Effektivität;
11. die Evaluation forschungspolitischer Maßnahmen.

Selbstverständlich werden ebenfalls über Lehr- und gelegentlich über Wissenstransferleistungen Daten erhoben und Evaluationen durchgeführt, so dass WissenschaftlerInnen auch bei der Sammlung, Berichterstattung und Begutachtung dieser involviert sind.

Evaluiert wird inzwischen auf allen Ebenen: Ganze Institutionen werden Audits und Akkreditierungen unterzogen, Fakultäten und Departemente werden evaluiert, um Mittelzuteilungen und strategische Entwicklungspläne zu rechtfertigen, Studienprogramme und Unterrichtseinheiten werden evaluiert, um die Lehre zu verbessern und Studierbarkeit zu gewährleisten, und einzelne WissenschaftlerInnen werden evaluiert, um über Projektmittel oder Berufungen zu entscheiden. An allen Evaluationen sind WissenschaftlerInnen beteiligt, z.T. mit erheblichem Zeitaufwand (Jacob & Teichler 2011).

Einen umfassenden Überblick über die verschiedenen Ziele und Nutzergruppen der Forschungsbewertung und deren Datenbedarf liefert die von der Europäischen Kommission eingesetzte Expertengruppe AUBR (Assessment of University-Based Research, 2010, Tabelle im Anhang). Dabei zeigt sich, dass der Datenbedarf auf allen Ebenen anfällt und wegen der Diversifizierung der Finanzierungsquellen für Hochschulen (EUA 2008) vor allem in der Forschung stark wächst. Der stetig zunehmende Anteil von Drittmitteln an der Forschungsfinanzierung und die Diversifizierung der Drittmittelinstrumente und -quellen führen schon bei der Einwerbung der Forschungsmittel zu steigenden Antrags-, Berichts- und Begutachtungspflichten, zu denen noch die Evaluationen oder Leistungserhebungen der Forschungseinheiten hinzukommen. Ein signifikanter Zuwachs an Begutachtungspflichten entsteht des Weiteren durch die Globalisierung der *Peer Review* im Rahmen von Fachevaluationen oder bei Begutachtungen der Forschungsförderorganisationen vieler europäischer Länder, die inzwischen immer häufiger internationale *Peers* in institutionellen oder nationalen Evaluationen oder Begutachtungen einsetzen, wie auch in den europäischen *Standards and Guidelines for Quality Assurance* und vom europäischen Zusammenschluss der Forschungsförderorganisationen (EUROHORCs 2006) empfohlen wurde.

Der AUBR-Report zeigt zudem, dass der Datenbedarf je nach Zielen und Bedürfnissen der Anspruchsgruppen nicht nur variiert, sondern sich auch zwi-

schen den Nutzergruppen z.T. stark überschneidet. Der AUBR-Bericht und das jüngste Policy Paper der *League of European Research Universities* (LERU 2012) betonen daher, dass die im Rahmen der diversifizierten Hochschulfinanzierungsquellen zunehmende Vielfalt und mangelnde Abstimmung des Daten- und Evaluationsbedarfs die Begutachtungs- und Berichtspflichten für WissenschaftlerInnen exponentiell hat in die Höhe schnellen lassen.

Die allenthalben beobachtete Zunahme dieser gutachterlichen und evaluierenden Aktivitäten in und um die Wissenschaft wird inzwischen mit großer Sorge beobachtet, tragen sie doch zu erheblichen Zeitverlusten und, aus der Sicht vieler Hochschulangehöriger, eher zu einer Effizienzminderung als zu einer Effizienzsteigerung bei. Die disproportionierte Beanspruchung der Zeitressourcen von WissenschaftlerInnen wurde in einer jüngeren Vergleichsstudie der OECD zu den Bedingungen des Akademikerberufs in 20 OECD-Staaten (*The Changing Academic Profession*) belegt: WissenschaftlerInnen verbringen zwischen 23% und 45% ihrer Zeit mit der Selbstverwaltung von Hochschule und Wissenschaft und den damit verbundenen Dienstleistungen, inkl. der Beantragung, Beurteilung und Begutachtung von und dem Berichten über wissenschaftliche Leistungen (Jacob & Teichler 2011:27). Bedenkenswert ist auch die Tatsache, dass gerade die erfolgreichsten WissenschaftlerInnen immer häufiger als Gutachter (als Editoren, Antragsbegutachtende oder Evaluatoren wissenschaftlicher Einheiten) eingesetzt und somit am meisten beansprucht werden. Während von den Berichtspflichten mehr oder weniger alle WissenschaftlerInnen in irgendeiner Weise betroffen sind, werden vor allem die erfolgreichsten und renommiertesten für externe Begutachtungen von Zeitschriftenartikeln, Berufungen, Projektanträgen und bei Fachevaluationen zu Rate gezogen, so dass ironischerweise gerade den besten Forschenden am wenigsten Zeit für Forschung übrig bleibt, es sei denn, sie werden durch andere Maßnahmen von anderen Pflichten (wie der Lehre) befreit.

5 Problematik der Forschungsbewertung I: Vor- und Nachteile verschiedener Forschungsbewertungsmethoden

Die oben aufgeführten Bewertungspraktiken sollten in zwei Richtungen hinterfragt werden:

- Sie können zum einen im Hinblick auf eine konkrete Verbesserung der gegenwärtigen Bewertungspraxis, sozusagen in pragmatischer Hinsicht, kritisch beleuchtet werden, d.h. in ihrer Auswahl, in der Häufigkeit oder dem Kontext ihrer Anwendung, in ihren Konsequenzen und ihrem Zusammenspiel.
- Darüberhinaus können die ihnen zugrunde liegenden Ansätze grundsätzlicher in Frage gestellt werden, im Hinblick auf die sie begründende Misstrauenskultur und deren Auswirkungen auf die Innovationskraft der Wissenschaft.

In diesem Abschnitt wenden wir uns der ersten, pragmatischen Ebene zu, bevor wir unter Punkt 6 die Bewertung als kulturelle Praxis und deren Prämissen ins Visier nehmen.

In pragmatischer Hinsicht spannt sich die Diskussion zur Forschungsbewertung und deren Angemessenheit vor allem zwischen den zwei Kategorien von Methoden und deren Einschätzung auf: einerseits der allgemein innerhalb der Wissenschaft verhältnismäßig gut etablierten *Peer Review* und andererseits den immer weiter verbreiteten quantitativen Leistungsindikatoren. *Peer Review* kommt vor allem im Rahmen von qualitativen *ex ante*-Bewertungen von Forschungsprojektanträgen, zur Publikation vorgeschlagenen Zeitschriftenartikeln und bei der Berufung von ProfessorInnen zur Anwendung, aber auch als formative Evaluation, die WissenschaftlerInnen und deren institutionelle Einheiten in ihrer Entwicklung berät und begleitet. *Ex post*-Bewertungen und *ex ante*-Einschätzungen fließen in dieser Art von Evaluation zusammen. Die immer auf zurückliegende Leistungen bezogenen quantitativen Indikatoren werden vor allem für Vergleiche zwischen Institutionen und Wissenschaftlergruppen verwendet. Die Vor- und Nachteile beider Arten von Forschungsbewertung werden im Folgenden kurz dargelegt.

5.1 Peer Review

Evaluationen durch wissenschaftliche Fachkollegen (*Peer Review*) haben nicht nur als Forderung externer Anspruchsgruppen, sondern auch innerhalb des Wissenschaftssystems und seiner Selbstregulierung an Bedeutung gewonnen. Der Bedeutungszuwachs ist zum einen dem wachsenden Anteil der im Wettbewerb über *Peer Review* begutachteten Drittmittel geschuldet. Zudem ist der Bedarf an *Peer Review* im Rahmen des globalisierten Wettbewerbs um qualifizierte WissenschaftlerInnen, um Sichtbarkeit und Ressourcen gewachsen, weil begutachtete Publikationen in internationalen wissenschaftlichen Journals immer wichtiger werden. Ausserdem werden bei der Mittelzuwendung Publikationen und Drittmittel, die nicht ohne *Peer Review* auskommen, als Indikatoren immer zentraler und so zu Treibern des Zuwachses. Schließlich sind sie fester Bestandteil der Evaluation von Forschungseinheiten und ganzen Einrichtungen, die an den meisten Institutionen inzwischen zyklisch betrieben wird.

Die Vorteile der *Peer Review* sind lange bekannt, gibt es diese doch bereits seit dem 17. Jahrhundert, ursprünglich als Methode der Royal Society, um die Flut von Beobachtungen, Experimenten und Theorien unklarer Validität der expandierenden modernen Naturwissenschaften einzudämmen (Hornbostel 2008). Zu den Vorteilen gehört vor allem die Tatsache, dass *Peer Review* die selbstregulierenden Kapazitäten des wissenschaftlichen Systems stärkt, indem sich WissenschaftlerInnen auf Augenhöhe gegenseitig beurteilen. Entscheidend sind dabei die Annahme des hohen intellektuellen Niveaus und vor allem der Fachkompetenz der beurteilenden *Peers* und der damit verbundene Konsens unter WissenschaftlerInnen, dass *Peer Review* der bestmögliche Weg zur Fremdbeurteilung der eigenen Leistungen darstellt. Solange Ausmaß und Aufwand der *Peer Review* nicht ausufern und die Auswahl der GutachterInnen oder der im Fokus der Evaluation stehenden Fragen und Themen als angemessen empfunden wird, stellt die *Peer Review* eine weitgehend akzeptierte Methode der Forschungsbewertung dar.

Trotz ihrer anerkannten Rolle als Selbstregulativ der Wissenschaft werden gegen die *Peer Review* inzwischen viele kritische Einwände ins Feld geführt, die auf Beobachtungen von Betroffenen (GutachterInnen und Begutachteten) im Rahmen der Ausweitung der *Peer Review* und zum Teil auf empirischen Studien fusen (so z.B. Peters & Ceci 1982; Moxham & Anderson 1992; Daniel 1993; Röbbcke & Simon 2001; Wenneras & Wold 1999; Lamont 2009). Kritisiert wird die Verlässlichkeit von *Peer Review* in mehrerlei Hinsicht:

- Die jeweiligen Urteile der *Peers* widersprechen sich häufig. Der aufsehenerregendste Nachweis der mangelnden intersubjektiven Verlässlichkeit erfolgte 1983 in einem Feldexperiment, in dem Peters und Ceci vor 18 bis 32 Monaten bereits veröffentlichte Artikel von namhaften Wissenschaftlern bei denselben Zeitschriften unter anderen Namen und Adressen nochmals einreichten. Weniger als 20% wurden erneut angenommen (Peters & Ceci 1982). Von einer Reproduzierbarkeit der *Peer Review*-Urteile kann insofern kaum gesprochen werden.
- Ebenfalls problematisch ist die Begutachtung durch *Peers* bei der Beurteilung von sich neu herausbildenden Fachgebieten. Angesichts der starken Zunahme solcher Fachgebiete sowie zunehmender Interdisziplinarität sei es für die *Peers* nicht einfach, sich eine valide Meinung zu neuen, gerade interdisziplinären Forschungsgebieten zu bilden (van Raan 1995:88), nicht zuletzt, da GutachterInnen häufig aufgrund ihres Erfolgs in einer Disziplin ausgewählt und zumeist in disziplinären Panels eingesetzt werden, so dass sich in diesen Beurteilungskontexten ein disziplinärer Mainstream verstärkt (Lamont 2009; Martin & Whitley 2009). Dies wird auch durch Fehlurteile der *Peers* über bahnbrechende Forschungen verschiedener Nobelpreisträger oder anderer herausragender WissenschaftlerInnen bestätigt. Beispiele sind Urteile über Enrico Fermi, Sir Hans Krebs, Rosalyn Yalow, Gerd Binnig, Hans Rohrer, Alfred Wegener, Alan Turing, Konrad Zuse, Hermann Oberth, Francis Peyton Rous, Mitchell Feigenbaum, Frank Rosenblatt, Stanley Prusiner, Günter Blobel, Andrei Linde, Noam Chomsky oder Karl Popper (zusammengestellt von Fischer 2003:29).
- Mitunter können Interessenskonflikte der Gutachtenden zu unfairen oder sogar falschen Bewertungen führen.

- Beobachtet wird des Weiteren, dass *Peers* die Leistungen jüngerer WissenschaftlerInnen leichter übersehen und häufig einen negativen Bias gegenüber Newcomern im Fachgebiet mitbringen (Fischer 2003).
- Einige empirische Studien zu den Auswirkungen der *Peer Review* auf Antrags- und Publikationsverhalten beobachten Mainstreaming-Effekte und den Hang zur Vermeidung risikoreicher und paradigmengreifender Forschung. Durch die zunehmende Betonung von Publikationen in auf *Peer Review* basierenden wissenschaftlichen Zeitschriften wird daher immer mehr die nicht riskante Forschung gefördert (Henkel 1999; Horrobin 1990; Lee 2007; Overbeke & Wager 2003).

Als Grundlage für die Beurteilung von wissenschaftlicher Leistung einzelner WissenschaftlerInnen ist die *Peer Review* bei der Einschätzung der Leistungsschwachen deutlich verlässlicher als bei der Beurteilung der Leistungsstarken. Außerdem wird das Risiko der Fehleinschätzung umso größer, je größer die zu beurteilende Einheit und damit die unzureichende Feldkenntnis der GutachterInnen ist. Je kleinteiliger der Begutachtungsfokus, desto größer ist allerdings das Risiko des disziplinär verengten Blicks und damit der Fehleinschätzung neu entstehender wissenschaftlicher Felder.

Ein denkbare Korrektiv der subjektiven Urteilsfindung wäre die Einbettung der Einzelurteile in einen Gruppenbegutachtungszusammenhang. Doch empirische Untersuchungen zeigen, dass der Gruppeneffekt als Korrektiv kaum in Frage kommt. So beobachten Langfeldt (2004) und Olbrecht & Klein (2011), dass der Entscheidungsfindungsprozess der Gruppenbegutachtung durch eine eindeutige Aufgabenteilung sowie durch wenig Interaktion zwischen den Gutachtenden gekennzeichnet war, mit Ausnahme der Fälle, in denen sich die Kompetenzbereiche von Gutachtenden überschneiden. Langfeldt fordert daher: «Two experts assessing each object under review and some time for discussing the results would be the minimum needed if expert panel evaluations are to have some function exceeding individual review reports when it comes to assessing the quality of research» (Langfeldt 2004:60). Neben Untersuchungen zu den kognitiven und epistemologischen Grenzen der Verlässlichkeit von *Peer Review*-Beurteilungen, die auf individuelle und psycho-

logische Bedingungen der beurteilenden Individuen eingehen, gibt es seit wenigen Jahren auch soziologische Analysen der *Peer Review* als intersubjektive Praxis. Vor allem die Arbeit von Michèle Lamont macht auf die Tatsache aufmerksam, dass *Peer Review* als gemeinschaftliche Praxis in Teams oder im Kontext von Wissenschaftsgemeinschaften erfolgt und damit jenseits der individuellen Haltungen und Voreingenommenheit einzelner Personen gemeinsamen Normen, Verhaltensregeln und intersubjektiven Konventionen unterworfen ist, die ihre Ergebnisse legitimieren. Sie untersucht «how panelists justify their judgments, but also how the processes and rules of practice set the stage for a sense of legitimacy» (Lamont 2009:246). Gerade die Spannung zwischen individuellen Positionen und dem Konsens der *Peer Panels* gehört wesentlich zur Funktionsweise der *Peer Review* als zentraler Evaluationspraxis. Evaluation ist immer auch kulturell verhaftet. Die individuellen und intersubjektiven Voreingenommenheiten gehören zur qualitativen Evaluation wesentlich dazu, so dass die Erwartung einer universalen Rationalität der Evaluationsergebnisse und einer entsprechenden meritokratischen Gerechtigkeit fehl am Platze ist.

«What is presented as expertise may sometimes be merely preference (taste), described in depersonalized language. The reciprocal recognition of authority is central to the process, but it may lead to explicit horse-trading, which produces suboptimal results. Despite these potential hazards, however, panelists think the process works, in part because they adopt a pragmatic conception of truth (or at least of what constitutes a fair evaluation) as something inevitably provisional and defined by the best standards of the community at the time» (Lamont 2009:240).

Aus Lamonts Sicht ist damit die positive Funktion und Legitimität der *Peer Review* im Wissenschaftssystem nicht in Frage gestellt, sondern nur die überhöhten Erwartungen an ihre Belastbarkeit (Lamont 2009:241). Empirische Untersuchungen zu den Folgen von Evaluationsverfahren sind erst in wenigen europäischen Ländern durchgeführt worden und dort primär zu Evaluationsverfahren der Lehre, die bereits länger etabliert sind, so z.B. in Belgien (Hulpiu & Waeytens 2003), in Dänemark (Bjørnkilde & Bason 2000; Thune & Kristoffersen 1999; The Danish Evaluation Institute 2004), in Großbritannien (Brennan/Frederiks/Shah 1997; Brennan/Shah/Williams 1996) und in den Nie-

derlanden (Frederiks/Westerheijden/Weusthof 1994; Jeliaskova 2002). Was die Forschungsevaluation angeht, wurden Studien zur Akzeptanz von Evaluationen der Forschung nur von der Wissenschaftlichen Kommission Niedersachsen, in begrenztem Maße vom deutschen Wissenschaftsrat als Auswertung seines Pilotprojekts beim *Forschungsrating* sowie von verschiedenen Akteuren in Großbritannien im Rahmen der Diskussion um das Evaluationsverfahren *Research Assessment Exercise* (RAE) und das dieses ablösende *Research Excellence Framework* (REF) durchgeführt. Erste nicht systematisch ausgewertete oder veröffentlichte Ergebnisse zur Akzeptanz von Evaluationsergebnissen wurden auch von Catherine Paradeise bei einem Vortrag in Lausanne vorgestellt (Paradeise 2011).

Das britische RAE ist die umfassendste systematische *Peer Review*-basierte Forschungsevaluation Europas und im Hinblick auf finanzielle und systemische Folgen sicherlich auch die tiefgreifendste. Ihre Auswirkungen auf das britische Forschungssystem wurden eingehend untersucht, allerdings vor allem im Hinblick auf veränderte Governancestrukturen, Machtverhältnisse, institutionelle Differenzierung und relative Kosten. Auf den Mehrwert des *Research Assessment Exercise* für die Qualität der Wissenschaft wurde allerdings in der jüngsten Evaluation des RAE (PA Consulting Group 2009) nicht eingegangen. Bei den Befragungen wurde lediglich deren Mehrwert für die Sichtbarkeit von Institutionen und Forschungsgruppen und für ihre Fähigkeit, zusätzliche Mittel einzuwerben, hervorgehoben. Zu den Auswirkungen auf wissenschaftliche Themenstellung, Orientierung und Qualität der Publikationen gibt es in dieser Evaluation der Evaluation keine Aussagen. Andere jüngere Studien wie Martin & Whitley (2010) untersuchen die Auswirkungen des RAE auf Machtverhältnisse («authority relations») und verzeichnen dabei eine Stärkung staatlicher Macht gegenüber den Universitäten (via Funding Councils und Ressourcenallokation) sowie eine Stärkung wissenschaftlicher Eliten und traditioneller disziplinärer Grenzen, die durch die Organisation der Evaluationspanels stabilisiert werden (Martin & Whitley 2010:65). So werden bei der Auswahl der Gutachter diejenigen bevorzugt, die im Mainstream disziplinärer Forschung erfolgreich sind, wodurch das RAE zur Betonung konventioneller For-

schung beiträgt: «The RAE has reinforced the emphasis on conventional mainstream research, discouraging new developments and interdisciplinary research» (Martin & Whitley 2010:65). Durch die Betonung der *Peer Review* gewinnen die «Wächter» (*gatekeeper*) der Wissenschaft als «arbiters of excellence» an Einfluss auf die wissenschaftliche Produktion und mittelfristig sogar auf die Auswahl der WissenschaftlerInnen, wie Lee & Harley nachwiesen (1998). Eine Auswirkung auf die Qualität der wissenschaftlichen Outputs ist ansonsten nicht Gegenstand der empirischen Untersuchung.

Bei einer der wenigen empirischen Untersuchungen über die Akzeptanz der Evaluation und ihrer Ergebnisse unter betroffenen WissenschaftlerInnen und GutachterInnen, welche im Anschluss an die Forschungsevaluation aller Hochschulen in Niedersachsen von der wissenschaftlichen Kommission vorgenommen wurde, fallen die Urteile über den Evaluationsprozess und den Nutzen der Evaluation unter den betroffenen WissenschaftlerInnen eher kritisch aus. So findet weniger als die Hälfte, dass die Gespräche effektiv waren (44%) oder dass alle relevanten Stärken und Schwächen des Faches behandelt wurden, nur 28% beurteilen die Gespräche auch unter wissenschaftlichen Gesichtspunkten als interessant und anregend, obwohl 72% der Hochschulangehörigen die GutachterInnen als angesehene VertreterInnen ihres Fachs ansahen. Nur die Hälfte der WissenschaftlerInnen bewertet die gutachterlichen Empfehlungen als hinreichend präzise und handlungsleitend formuliert (im Gegensatz zu 90% der GutachterInnen, WKN 2006:19 und 23). Lediglich 43% sind der Ansicht, dass die Empfehlungen im Fach akzeptiert wurden (im Gegensatz zu 55% der GutachterInnen). Selbst wenn 56% der WissenschaftlerInnen die Umsetzung der Empfehlungen für wichtig halten, findet ein noch größerer Anteil (62%), dass es keine ausreichende finanzielle oder strukturelle Unterstützung bei der Umsetzung der Empfehlungen gibt. 55% beklagen außerdem, dass man das Gutachten gar nicht oder kaum mit der Gutachterkommission diskutieren konnte. 48% waren der Auffassung, dass der Aufwand der Evaluation nicht dem Nutzen entspricht. Entsprechend vorsichtig sind die Einschätzungen zur optimalen Nutzung der Evaluation unter den WissenschaftlerInnen. So erkennt zwar etwas über die Hälfte (56%) der durch die *Infor-*

med Peer Review Niedersachsens begutachteten WissenschaftlerInnen Forschungsevaluation als ein geeignetes Instrument der Qualitätssicherung an, aber fast zwei Drittel (64%) finden gleichzeitig, dass die Ergebnisse der Evaluation ausschließlich Empfehlungscharakter hätten und nicht für die Mittelzuweisung genutzt werden sollten.

Auffällig ist dabei, wie stark die Rückmeldungen der WissenschaftlerInnen von denen der GutachterInnen abweichen: Insgesamt äußern sich die WissenschaftlerInnen deutlich kritischer als die GutachterInnen über Prozess und Ertrag der Evaluation. Die GutachterInnen urteilen insgesamt bedeutend positiver, was ihre eigene Tätigkeit angeht, und schreiben den eigenen Empfehlungen eine wesentlich höhere Bedeutung zu. Aber sie merken ebenfalls kritisch an, dass die Ursachendiagnose zur Varianz der Forschungsleistungen in Abhängigkeit unterschiedlicher Rahmenbedingungen zu schwach ausfällt.

Eine ähnliche Diskrepanz zwischen der Akzeptanz der Evaluation durch die GutachterInnen und derjenigen durch die begutachteten WissenschaftlerInnen offenbart die Gesamteinschätzung der Evaluationsverfahren von Studium und Lehre durch ZEvA und Nordverbund, in der Fachangehörige und Begutachtende im Rahmen der schriftlichen Befragung nach der Bewährung des Evaluationsverfahrens, der Zielerreichung von Qualitätssicherung und -verbesserung, ihrer allgemeinen Zufriedenheit mit dem Ablauf der Evaluation, der Angemessenheit des Verhältnisses zwischen Aufwand für das Evaluationsverfahren und Ergebnissen der Evaluation sowie dem persönlichen Nutzen durch die Beteiligung an der Evaluation gefragt wurden. Die Einschätzung der Evaluation durch die begutachteten WissenschaftlerInnen fällt insgesamt zwischen 13% und 32% negativer aus als die Einschätzung durch die GutachterInnen, aber insgesamt positiver als in der oben erwähnten zeitgleichen und ebenfalls mehrstufigen Forschungsevaluation in Niedersachsen (Mittag/Bornmann/Daniel 2006).

Gemäß Catherine Paradeise (Paradeise 2011) hängt die Variabilität der Akzeptanz der Evaluationen nicht nur von deren Verwendungen, sondern auch von der Art des Engagements der WissenschaftlerInnen in Hochschule und Fachgemeinschaft und den damit verbundenen Erwartungen an die Institution ab. Auf der Basis einer Befragung evaluierter Wissenschaft-

lerInnen entwickelt Paradeise eine Typologie der Aufnahme der Evaluation, gemäß welcher mit entscheidend ist, wie stark oder schwach die Identifikation des oder der Wissenschaftlers/in mit der Institution oder der Fachgemeinschaft ist. Die sich aus dieser Identifikation ergebenden Erwartungen an die Institution bestimmen Umgang und Akzeptanz der Forschungsbewertung.

Angesichts der begrenzten Verlässlichkeit und der möglichen nicht intendierten Effekte der *Peer Review* wird deren Einsatz je nach Anwendungskontext sehr unterschiedlich eingeschätzt. Während *Peer Review* als reine Leistungsbewertung bisweilen auf Kritik trifft, wird sie als Bestandteil einer formativen Evaluation, bei der die Entwicklungsperspektiven einer Forschungseinheit eingeschätzt werden, im Großen und Ganzen eher positiv bewertet.

Unabhängig vom Verwendungskontext wird angesichts der wachsenden Anzahl, Vielfalt und internationalen Verbreitung der *Peer Review* die Überbeanspruchung einer zunehmenden Anzahl von GutachterInnen beanstandet, die deren Auswahl immer schwieriger werden lässt. Ebenso wird der bei wachsender Routinisierung abnehmende Grenznutzen der Evaluationen kritisch gesehen (Hornbostel 2008). So warnt Wilhelm Krull (Krull 2011) vor den Gefahren der eskalierenden Evaluationsfrequenz, in der selbst formativ angelegte Evaluationen ihren formativen Charakter verlieren und zu «einem sinnentleerten Ritual» verkommen.

Eine Anpassung der Evaluationspraktiken in Umfang, Fokus, Frequenz und Intensität an die bisherigen Evaluationserfahrungen und Leistungsbilanzen einer Forschungseinheit läge insofern nahe.

5.2 Forschungsbewertung durch quantitative Daten/Indikatoren

Wegen des erheblichen Aufwands qualitativer *Peer Review*-Evaluierungen für die als Gutachtende oder Beutachtete beanspruchten WissenschaftlerInnen erscheint der Rekurs auf quantitative Indikatoren als leicht handhabbares Mittel der Leistungsüberprüfung attraktiv. Im Gegensatz zu den durch *Peer Review* ermittelten qualitativen Daten stehen die quantitativen Forschungsdaten, die zumindest zum Teil als Indikatoren für Forschungsqualität verstanden werden, im Ruf, deutlich weniger aufwendig zu sein. Die erhitzten Diskussionen um die Reform des RAE wurden nicht umsonst durch das über den Evaluationsaufwand besorgte Finanzministerium ausgelöst. Die quantitativen Methoden der Forschungsbewertung bieten den wissenschaftsexternen Steuerungsorganisationen zudem den Vorteil, eine Externalisierung der Bewertungsprozesse zu erlauben. Inhaltliche Beurteilungen werden ausgeklammert. Mit Daten, die zwar vom Wissenschaftssystem selbst produziert werden und zum Teil Ergebnis wissenschaftsinterner *Peer Reviews* sind, wird von außen mit der Ressourcenzuweisung über Erfolgsmessung in die autonomen Selbststeuerungsprozesse der Wissenschaft eingegriffen.

An quantitative Indikatoren knüpft sich vor allem die Hoffnung auf eine dank ihrer vermeintlichen Objektivität gewährleistete Vergleichbarkeit zwischen verschiedenen Forschungskontexten (van Raan 1995, 2005). Entsprechend werden Indikatoren vor allem dann eingesetzt, wenn eine große Anzahl von Institutionen oder Forschungsgruppen miteinander verglichen werden sollen (wie bei der Mittelzuweisung über indikatorenbasierte Leistungsbewertung oder bei Rankingvergleichen von Institutionen) oder wenn möglichst objektive Daten als Hintergrundinformation und Ergänzung zu anderen intersubjektiven Einschätzungen (von GutachterInnen) dienen und Vergleiche zwischen Forschungsstandorten zulassen sollen (wie beim RAE oder REF in Großbritannien). Bei der Mittelzuweisung an und innerhalb von Hochschulen sind indikatorengestützte Methoden inzwischen in den meisten europäischen Ländern die Norm, wenngleich die Auswahl und Gewichtung der verwendeten Indikatoren heftige Kontroversen auslösen.

Die Kritik an der Verwendung quantitativer Indikatoren bei der Leistungsbewertung und vor allem bei ihrer Verwertung bei Rankings oder bei der Mittelzuweisung richtet sich zum einen auf die starken Verzerrungen bei der Beurteilung von Leistungen und zum anderen auf die unerwünschten Verhaltensanpassungen bei den WissenschaftlerInnen.

Verzerrte Einschätzungen von Forschungsleistungen entstehen zunächst durch die Reduktion der Mehrdimensionalität von Hochschulaktivitäten auf messbare Forschungsleistungen. Die mangelnde Verfügbarkeit und Vergleichbarkeit der Forschungsdaten stellen immer noch eine der größten Herausforderungen bei der Forschungsbewertung dar (AUBR 2010; LERU 2012). Viele Daten sind entweder gar nicht bzw. nur durch erheblichen Aufwand durch hochschulinterne Umfragen oder Analysen oder nur mit bestimmten kontextabhängigen Spezifikationen verfügbar, so dass Vergleiche über Hochschulen und Länder hinweg unmöglich werden oder mit einem riesigen Datensammlungsaufwand aufseiten der Hochschulen verbunden sind, der in keinem Verhältnis zum Nutzen des Leistungsvergleichs steht, wie in dem von der EU geförderten *Multirank*-Projekt. Der Mehraufwand bei der Datenerhebung muss durch den Mehrwert der Aussagekraft der Daten aufgewogen werden. Beschränkt man sich aber bei Forschungsbewertungen nur auf die verfügbaren und vergleichbaren Datensätze, ergibt sich ein verzerrendes Bild der bewerteten Forschung, da wichtige Forschungsoutputs nicht erfasst sind. So sind z.B. bibliometrische Daten zu *peer-reviewed* Zeitschriftenartikeln verfügbar bzw. kommerziell erhältlich, aber andere Arten von Forschungsergebnissen oder -veröffentlichungen, die eventuell für einzelne Disziplinen noch wichtiger sind (wie Conference Proceedings oder Monographien), sind nicht international zusammengefasst und über Suchfunktionen identifizierbar.

Kritisiert wird zudem die Bevorteilung publikations- und drittmittelintensiver Fachbereiche, deren Leistungen sich eher in den international registrierten Zeitschriften und wettbewerblich eingeworbenen Drittmitteln spiegeln. So räumen Bibliographieexperten wie van Raan ein, dass sich nicht alle Disziplinen in den bibliometrischen Indikatoren mit Aussagekraft abbilden lassen: «Bibliometric indicators are very well applicable in the natural and life sciences. However, in

the applied and engineering sciences as well as in the social and behavioural sciences (and even more in the humanities) international journals are often not the primary communication channel» (van Raan 2005:9). Inzwischen hat die Bibliometrie zwar ihre Methoden der Publikationserfassung auch in diesen Disziplinen verfeinert, kann aber immer noch nicht den Anspruch einer zufriedenstellenden Forschungsoutputerfassung erheben. Verliesse man sich allein oder zu stark auf Indikatoren, würde also mittel- und langfristige ein Abbau der Geistes- und Sozialwissenschaften resultieren. Diese Tendenz ist (auch aus anderen Gründen) bereits in einigen europäischen Ländern zu verzeichnen (Henkel 1999; LERU 2012).

Ein dritter wichtiger Einwand gegen die Überbewertung quantitativer Leistungsindikatoren betrifft deren Fokussierung auf Forschungsleistungen und damit die schleichende langfristige Unterbewertung der weniger messbaren Lehrqualität. So gibt es z.B. keine national oder international vergleichbaren Daten zu den Kompetenzen der Absolventen (trotz des AHELO Projekts der OECD) oder über deren Berufserfolge.

Die starke Betonung von Leistungsindikatoren, vor allem der Drittmittel und Publikationsdaten, bei Berufungen und internen Mittelzuweisungen führt nachweislich zu unerwünschten Verhaltensanpassungen. Diese lassen sich z.B. bei der Auswahl und Planung von wissenschaftlichen Projekten beobachten, aber auch als Vermeidung von ergebnisoffenen Projekten mit unsicherem Ergebnis und Publikationspotenzial sowie in der Veröffentlichungspraxis, z.B. bei der fragmentierenden Veröffentlichung der Ergebnisse in möglichst zahlreichen Einzelpublikationen («smallest publishable units»; Henkel 1999; Weingart 2005). Eine weiterführende Zusammenfassung der Studien zu den Auswirkungen der quantifizierten Forschungsevaluation wurde bereits im Bericht «Zur Quantifizierung wissenschaftlicher Leistungen in Forschungsevaluation und Hochschulrankings» von S. Reichert, im Auftrag des SWTR, August 2010 (intern, unveröffentlicht), vorgenommen.

Zur Verwendung einzelner Indikatoren werden ebenfalls kritische Einwände vorgebracht: So müssen z.B. bei der Erhebung und Nutzung der bibliometrischen Leistungsindikatoren, welche die Produktivität über Publikationen und Forschungswirkung über Zitationen messen und wegen ihrer größeren Aussagekraft

am beliebtesten sind, bestimmte Bedingungen gegeben sein, damit die Daten nicht durch die Darstellung zu falschen Analysen und verzerrten Wahrnehmungen führen:

- Das Aggregationsniveau darf nicht zu niedrig sein.
- Wissenschaftliche Felder dürfen nicht auf der Basis absoluter Häufigkeit verglichen werden, gebietsbezogene Normierungen sind Voraussetzung für Vergleichbarkeit.
- Alle Autoren müssen berücksichtigt werden und ihr Beitrag entsprechend gewichtet sein, und Namen und Adressen sollten vereinheitlicht werden.

Darüber hinaus muss berücksichtigt werden, dass Zitationen, die als Hauptindikator für Forschungswirkung und damit indirekt für Forschungsqualität angesehen werden, nicht tatsächlich wissenschaftliche Qualität abbilden. So kann z.B. ein Review-Artikel mit einer Übersicht über ein Feld besonders häufig zitiert werden, obwohl er keinen Anspruch auf Originalität oder eigene Forschungsqualität erhebt. Conference Proceedings sind ein weiteres Beispiel.

Kritische Einwände werden gegen die Verwendung der Drittmittel einer Einheit oder Person als Indikator ihrer Forschungsintensität oder ihres Erfolgs im Forschungswettbewerb vorgebracht. Die Verwendung der Drittmittel als Indikator wird zum einen damit gerechtfertigt, dass Drittmittel Forschungsprojekte erst ermöglichen, so dass mehr Drittmittel auch mehr Forschungsprojekte widerspiegeln. Zum anderen wird auf die *Peer Review* hingewiesen, die zumindest bei öffentlichen Drittmitteln über Erfolg im Wettbewerb zwischen verschiedenen Bewerbern entscheidet. Drittmittel werden somit als indirekter Indikator der Anerkennung unter *Peers* angesehen. Kritisiert wird diesbezüglich, dass Drittmittelacquire als Gradmesser von Forschungsintensität für verschiedene disziplinäre Kontexte nicht gleichermaßen aussagekräftig ist. In den immer noch primär auf die individuelle Leistung angelegten Geisteswissenschaften sind Drittmittel nicht unbedingt ein Gewinn für die eigene Forschungsarbeit. Über Drittmittel Doktorierende zu finanzieren, trägt bei GeisteswissenschaftlerInnen nur in Ausnahmefällen zur *eigenen* Forschung bei, obwohl hier ebenfalls verstärkte Vernetzungspraktiken und drittmittelgestützte Verbundinitiativen zu verzeichnen sind. Hingegen kämen in den experimentellen Na-

turwissenschaften, in denen die Forschungsarbeit von der Verfügbarkeit (extern) finanzierter NachwuchswissenschaftlerInnen und wissenschaftlicher Geräte abhängt, die meisten Forschungsprojekte kaum ohne Drittmittel aus. Irreführend ist zudem, dass an einigen Forschungsinstituten die Verfügbarkeit großzügig bemessener interner Forschungsgelder die Acquire von Drittmitteln überflüssig macht, so dass die Drittmittel nicht als Abbild der Forschungsintensität angesehen werden können.

Bedenkenswert ist jedoch nicht nur die begrenzte Aussagekraft, sondern auch die wachsende Bedeutung der Drittmittel als vermeintlicher Indikator wissenschaftlichen Erfolgs. In Deutschland schauen DFG und Wissenschaftsrat mit Sorge auf den wachsenden Drittmittelanteil. Demnach sind die laufenden Grundmittel der Hochschulen in Deutschland zwischen 1998 und 2010 nur um 23% angewachsen (von 12,6 auf 15,5 Milliarden Euro), während die von den Hochschulen im Wettbewerb eingeworbenen Drittmittel im selben Zeitraum um mehr als 100% (von 2,5 auf über 5,3 Milliarden Euro) anstiegen. Damit hat sich die «Drittmittelquote», also der Anteil der Drittmittel an der Gesamtfinanzierung der Hochschulen und ihrer Forschungen, in nur gut einem Jahrzehnt von 16 auf 26% erhöht. In England bestimmte daher die «Full costing»-Diskussion mehrere Jahre die Wissenschaftspolitik, da der gestiegene Drittmittelanteil die Grundversorgung weitgehend ausgehöhlt hat. Dort kann man inzwischen leichter ein teures neues wissenschaftliches Gerät beantragen als ein etwas älteres für einen Bruchteil des finanziellen Aufwands wieder instandzusetzen. Bedenklich ist der steigende Drittmittelanteil vor allem aus zwei Gründen:

- Zum einen wird dadurch eine erhöhte Anzahl von Gutachten und damit von Gutachterzeit beansprucht. Im Zuge der wachsenden Anzahl von Gutachten für Drittmittelprojekte und -instrumente wird es zusehends schwieriger, gute GutachterInnen für Evaluationen zu gewinnen. Ebenso nehmen nach Aussagen der WissenschaftlerInnen der Zeiteinsatz der GutachterInnen bei einer einzelnen Begutachtung und damit die Solidität der Beurteilung stetig ab.
- Zum anderen ist die Ausweitung der Drittmittelgelder zuungunsten der institutionellen Mittel bedenklich, weil diese mit einer schleichenden Aus-

höhlung der laufenden Mittel für Infrastruktur, Personal, Gerätwartung und Instandsetzung einhergeht. Diese Entwicklung wird nicht nur durch die externe Mittelzuweisung verordnet, sondern auch durch die zunehmende Betonung von Eigenbeiträgen der Institutionen bei Drittmittelprojekten. Aufschlussreich ist in diesem Zusammenhang die Selbstwahrnehmung der kürzlich vom iFQ befragten Professorinnen und Professoren: Sie werben Drittmittel ein, weil Forschungsfragen sonst nicht zu bearbeiten sind, weil Mitarbeitende nicht beschäftigt werden könnten und weil der Einwerbungserfolg Konsequenzen für die Zumessung von Grundmitteln im Rahmen der leistungsorientierten Mittelvergabe hat. Fragt man sie hingegen, welche Maßnahmen tatsächlich wissenschaftliche Exzellenz befördern, so antworten die Forschenden vor allem: verlässliche Karriereperspektiven und leistungsunabhängige Grundfinanzierung (Böhmer 2011).

5.3 «Informed Peer Review» als Lösung und Konsens?

Um die möglichen Verzerrungen bei der Verwendung quantitativer Indikatoren zu unterbinden und Kontextabhängigkeiten entsprechend einordnen zu können, müssen quantitative Wissenschaftsindikatoren «in die systematische Reflexion der Wissenschaft eingebettet sein [...]», gab Peter Weingart schon 1995 zu bedenken (Weingart 1995) und fügte hinzu, «die Entwicklung und Anwendung von Kombination externer Indikatorenenerstellung und interner Evaluierungskommissionen erscheinen richtungweisend.» Auch der Bibliometriexperte van Raan sah schon damals die aussagekräftigeren bibliometrischen Indikatoren vor allem als Hilfsmittel für das *Peer Review*-Verfahren an und hielt Letzteres trotz seiner Mängel immer noch für das beste Verfahren zur Forschungsbewertung (van Raan 1995). Während Weingart damals auf die Vorbilder des amerikanischen *National Science Board* und des Schweizerischen Wissenschaftsrats verwies, könnte man heute auf die Forschungsbewertungspraxis des britischen *Research Assessment Exercise* und seines Nachfolgers, des *Research Excellence Framework*, der niederländischen Forschungsfachberichte oder der australischen und französischen Fachvergleiche hinweisen. 15 Jahre später scheint sich nach erhitzten Debatten (vor allem in England und Australien) für und wider Indikatoren vs. *Peer Review*-Evaluationen die *Informed Peer Review* als konsensfähige Grundmethode der Forschungsbewertung und Kompromiss zwischen den obigen Hauptmethoden der Forschungsbewertung etabliert zu haben (AUBR 2010). Während die quantitativen Indikatoren eine erste Grundlage für Fragen und eine mögliche Korrektur für *Peer Bias* bieten, leistet die *Peer Review* eine Beurteilung des wissenschaftlichen Potenzials durch Experten des Feldes. Allerdings kann die *Informed Peer Review* bzw. quantitativ gestützte *Peer Review* zwar als Versuch der Linderung möglicher Verzerrungen bei der Beurteilung von Forschungsleistungen, nicht aber als Lösung aller oben erwähnten methodischen Probleme der Beurteilung angesehen werden. Zentrale Probleme der *Peer Review* bleiben nicht nur bestehen, sondern werden durch neue Beurteilungskontexte und -ziele sogar verstärkt:

Zum einen wird wegen der Zunahme der Anzahl und des Umfangs der Evaluationen, die immer häufiger nicht nur einzelne Forschungsanträge, sondern ganze Forschungseinheiten betreffen, der Rekurs auf zeitsparende Überblicksoutputdaten unter den *Peers* vermehrt genutzt, um die Fülle an Einzelinformationen überhaupt bewältigen zu können (Torka 2011; Kieser 2010). Somit nehmen die mit Indikatoren verbundenen Verzerrungseffekte zu.

Zum anderen müssen sich *Peers* immer häufiger mit Evaluationsgegenständen jenseits ihrer eigentlichen disziplinären Expertise beschäftigen, entweder weil das zu beurteilende Wissenschaftsfeld inter- oder multidisziplinär aufgespannt ist, oder weil organisatorische Rahmenbedingungen zunehmend ins Visier der Evaluation geraten. Gerade Letzteres ist wissenschaftspolitisch immer häufiger gewollt, wie die jüngsten Entwicklungen im Rahmen der Exzellenzinitiative und des vom Wissenschaftsrat in Deutschland betriebenen Wissenschaftsratings belegen. Von der Verbesserung der Bedingungen für NachwuchswissenschaftlerInnen, der Unterstützung von Innovation und Transferleistungen bis zur strategischen Steuerung der Departemente werden den *Peers* immer häufiger Fragen außerhalb ihrer eigentlichen disziplinären Autorität gestellt. Selbst wenn sie als WissenschaftlerInnen diesbezüglich Erfahrungswissen mitbringen, ist ihre Autorität in Bezug auf derartige Fragen eventuell weniger unumstritten als ihre Forschungsautorität im Feld ihrer Spezialisierung. Soweit aber eine wirkliche Beurteilung und nicht etwa ein Dialog über Perspektiven bezweckt wird, müsste die Evaluation eigentlich im engeren Expertenkreis gefasst sein (wie in der klassischen Forschungsförderung bei Gutachten zu Einzelprojektanträgen üblich), damit der Prozess tatsächlich eine «*Peer*» *Review* darstellt. Eine wirkliche Expertise, in der Experten über Experten urteilen können, weil sie das gleiche Forschungsumfeld überblicken, müsste sich auf das verhältnismäßig enge Feld derselben oder zumindest eng verwandte Spezialisierungen beschränken. Dies ist aber nur bei der Beurteilung von Forschungsanträgen und Zeitschriftenartikeln gegeben. Selbst dann kann die *Peer*-Beurteilung unter Umständen zu den erwähnten Mainstreaming-Effekten beitragen und das Potenzial neuer vorgeschlagener Forschungswege unterschätzen.

5.4 Schlussfolgerung

Die Lösung dieses Dilemmas, selbst in der *Informed Peer Review* keine wirklich zufriedenstellende, nachhaltige Form der Leistungsoffenlegung zu finden, liegt daher in der Anpassung der Frequenz, Form und Zielsetzung der Evaluationsverfahren. Selbst wenn die Leistungsbeurteilung einer hohen Fehlermarge unterworfen ist, können Evaluationen dennoch zur Qualitätsentwicklung beitragen, allerdings nur unter bestimmten kommunikativen Bedingungen, in denen ein offener Dialog und die Offenlegung der eigenen Schwächen als Teil eines Versuchs, diese zu beheben und die Stärken besser nutzen zu können, gelingen. Gerade bei evaluationsgeprüften Forschungseinheiten bringen weniger routinisierte, d.h. seltenere Evaluationen mit einem Fokus auf Fragen, die sich die Einheit tatsächlich auch selbst stellt, mehr Hoffnung auf einen anregenden, wirklich auf Qualitätsverbesserung zielenden Evaluationsprozess. Evaluationen sollten daher «auf der Ebene ganzer Fächer oder Institutionen sparsam eingesetzt» und «anlass- und strukturbezogen» konzipiert werden (Krull 2011).

Vor allem ist wichtig, dass die Interaktion zwischen Begutachtenden und Begutachteten einen wirklichen Dialog erlaubt, indem Stärken-Schwächen-Analysen einer Einschätzung der Zukunftspotenziale dienen. Damit eine Evaluation eine «produktive Verunsicherung» oder einen «Impuls zur Pfadkorrektur» bewirken kann (Matthies et al. 2008:340; Wiesenthal 2008:324), müssten Evaluationsverfahren von «Gratifikations- und Sanktionsmaßnahmen» entkoppelt werden (Matthies et al. loc.cit.).

6 Problematik der Forschungsbewertung II: Zwischen Leistungsorientierung und Misstrauenskultur

Grundsätzlich betrachtet lassen sich die Vervielfältigung der Bewertungspraktiken und deren Prämissen im Wissenschafts- und Hochschulsystem auch als kulturelle und normative Praxis hinterfragen. Zunächst stellt sich die Frage, was für Annahmen heutige Praktiken der Leistungsbewertung über die Funktionsweisen der Wissenschaft, über ihren Motor und mögliche Leistungsanreize implizit machen und inwieweit diese zutreffen. Der neue Umgang zwischen Staat und Universitäten spiegelt einen Wandel der Annahmen über den Sinn und Zweck der Wissenschaft als menschliches Unterfangen und über ihre Rolle und ihren Nutzen in der Gesellschaft wider, die kritisch betrachtet werden sollten.

6.1 Leistungsanreize in der Wissenschaft

Was ist neu an den oben beschriebenen Bewertungspraktiken, mit denen die Wissenschaft ihre Leistungs- und Anpassungsfähigkeit beweisen oder zumindest einschätzen soll? *Peer Review* gibt es so lange wie die etablierte Wissenschaftspraxis. Begutachtungen von Forschungsprojekten sind seit Langem Teil des wissenschaftlichen Alltags. Neu sind nicht nur der Zuwachs an zu begutachtenden Projektmitteln und der größere Zuschnitt der Evaluationseinheiten (ganze Forschungseinheiten und Institutionen), neu ist vor allem die Allgegenwart der Leistungsüberprüfung, die alle Aspekte des wissenschaftlichen Alltags erfasst und mit zyklischer Regelmäßigkeit jede Einheit des Wissenschaftssystems einscannt, um die Angemessenheit ihrer Leistung unter die Lupe zu nehmen, oft in Verbindung mit einem künstlich aufgeblasenen Wettbewerb. Wurde mit dem «Zurückdrängen der politischen Steuerungs- und Kontrollansprüche die Mobilisierung <unternehmerischer> Aktivitäten und insgesamt eine Effizienzsteigerung erwartet» (Hornbostel 2011:59), so entartet dieser Dynamisierungswille jetzt zu einer Überprüfungs-kultur, in der Leistungen erst ins Gewicht fallen, wenn sie evaluiert und gemessen werden. Am Ende interessiert nicht mehr die Leistung selbst, sondern die Fähigkeit einer Einheit, die eigene Leistung zu überprüfen. «Management attention shifts from individual scientists as experts to the capacity of scientific organisations to self-regulate» (Power 2008). Michael Powers Charakterisierung der britischen Gesellschaft als «Audit Society», die hemmungslos alle Bereiche mit ritualisierten Verifikationen und Evaluationen überzieht (Power 1997), trifft immer mehr auch auf das europäische Festland zu. Selbst wenn durchaus Leistungssteigerungen erzielt werden, stellt sich doch die Frage, inwieweit eine Dynamisierung wirklich erreicht wird oder eher die oben erwähnten Fehlsteuerungen und nicht intendierten Anpassungseffekte erzeugt werden (Henkel 1999; Frey 2008; Kieser 2010), die zwar Produktivitätssteigerungen bewirken, aber wissenschaftliche Innovation eher unterbinden.

Es gilt also zu fragen, inwiefern ständige Leistungsbewertung geeignet ist, Leistung zu fördern, d.h. als Anreiz zu dienen. Eine der zentralen Begründungen der systematischen Bewertung von Forschenden und Forschungseinheiten ist die Idee, gute Leistung zu belohnen. Inwiefern beeinflussen solche Anreize, sei es durch die Zuweisung zusätzlicher Mittel oder durch andere Formen der Belohnung, tatsächlich die Leistungen der WissenschaftlerInnen?

Richtungsweisend in diesem Zusammenhang sind organisations- und sozialpsychologische empirische Studien zur Wirkung von Anreizen auf Kreativität, die in den letzten Jahrzehnten durchgeführt wurden.

Auf der Basis empirischer Untersuchungen zeigt Teresa Amabile in ihrer Monographie über die Soziopsychologie der Kreativität (Amabile 1996), dass die Evaluation (und mitunter sogar schon die Erwartung der Evaluation), aber auch andere als kontrollierend empfundene «Constraints» negative Auswirkungen auf die Kreativität von Erwachsenen (und von Kindern) haben können. In späteren Studien differenziert sie diesen Effekt, z.B. in Abhängigkeit von persönlichen Eigenschaften oder von der Form der Evaluation. So korrelieren «prospects of threatening critical evaluation» mit nachweisbar niedrigerer Kreativität, während «informative, constructive feedback and evaluation that conveys positive recognition of creative work often co-occur with high levels of creativity». Somit wirkt nicht jede Form der Leistungsbewertung, sondern nur die als kontrollierend, überprüfend und einschränkend empfundene Evaluation kreativitätshemmend:

«[...] in real world settings, creativity may be undermined by evaluation that conveys incompetence or threatens self-determination, but creativity may be supported by evaluation that is work-focused and constructive (even when negative), that provides information about competence improvement, or that conveys positive recognition of competence and valued work. These positive effects may be due to the motivational synergy mechanisms (extrinsics in service of intrinsics)» (Amabile 1996:152).

Ähnlich differenziert Bruno Frey die potenziell anreizmindernde Wirkung von Evaluationen und unterscheidet zwischen einem Verstärkungs- oder Verdrängungseffekt auf die intrinsische Motivation der WissenschaftlerInnen (Frey 1997; Frey 2008). Der em-

pirisch von Psychologen überprüfte Verdrängungseffekt besagt, dass intrinsische Motivation durch extrinsische Motivation, wie z.B. externe monetäre Anreize, verdrängt wird (Deci & Ryan 1999). Dies geschieht immer dann, wenn die von außen kommenden Eingriffe als kontrollierend angesehen werden, so dass durch die Einschränkung des Handlungsspielraums die Selbstbestimmung und durch die Nicht-Würdigung der intrinsischen Motivation eventuell sogar die Selbsteinschätzung gemindert wird (Frey 1999:23–25). Empirische Studien weisen darauf hin, dass leistungsabhängige Belohnung der Arbeit sich sogar nachteilig auf Motivation und Leistung auswirken kann (Fehr & Falk 1999). Auf die Kreativität intrinsisch motivierter Akteure hat vor allem die Einschränkung des Handlungsspielraums nachteilige Effekte: «The characteristic most likely to kill creativity is not inadequate pay or tight deadlines, but a lack of freedom in deciding what to do or how to accomplish a task, lack of sense of control over one's own work and ideas» (Kohn 1993). Anreize hingegen, die auf zusätzliche Gestaltungsfreiheit bei zukünftigen Projekten abzielen, wirken motivationsverstärkend (Amabile 1996:175–177). «The creativity of work groups within organizations correlates positively with the degree to which members of those work groups report feeling that they have freedom in their work, a sense of autonomy and control over their own work and their own ideas» (Amabile 1996:177). Das genaue Zusammenspiel von extrinsischer und intrinsischer Motivation und die Bedingungen der gegenseitigen Verdrängung oder Verstärkung werden erst seit Kurzem empirisch erforscht, so dass diesbezüglich in der Zukunft noch konkretere Anregungen für institutionelle Bedingungen zu erwarten sind.

Fest steht, dass die aus der Wirtschaft übernommenen Vorstellungen der Effizienzsteigerung durch leistungsbezogene Anreize, Sanktionen und regelmäßige Leistungsbewertungen als Grundlage der Mittelverteilung an der Realität sozialpsychologischer Gegebenheiten des Wissenschaftsbetriebs vorbeizielten. Dieser zeichnet sich durch einen hohen Anteil intrinsisch motivierter Mitglieder und kreativer, d.h. nicht strengen Vorgaben folgender Tätigkeiten aus. Gerade diese eignen sich aber schlecht für externe kontrollierende Steuerung. Nur Evaluationen, die auf die selbst gegebenen Interessen und Fragen der Mitglieder eingehen und nicht mit als kontrollierend empfundenen

Konsequenzen verbunden sind, haben eine Chance auf leistungssteigernde Wirkung.

Was heisst dies für die Steuerung des Wissenschaftssystems?

Zunächst folgt unmittelbar, dass Leistungsanreize so konzipiert werden müssen, dass sie die Gestaltungsfreiheit und Selbstverantwortung verstärken. Ansonsten versprechen externe Leistungsanreize kaum positiven Einfluss auf die Kreativität der wissenschaftlichen Leistungen, auch wenn eventuell eine Produktivitätssteigerung im Sinne einer Anpassung an externe Steuerungsanreize erzielt wird. Ebendies entspricht der gängigen Argumentation gegen das RAE, dessen Produktivitätsgewinne unumstritten sind, während sich die Kritik gegen die erhöhte Kurzfristigkeit der Forschung, die verminderte Risikofreude der Forschenden und damit die nachlassende Innovationskraft der Forschungsrichtung wendet (Evidence Ltd 2005; Henkel 1999; May 2004).

Eine zweite Schlussfolgerung lautet, dass die Steuerung sich nicht auf die Beeinflussung bereits im System etablierter Akteure konzentrieren, sondern sich stattdessen primär auf die *ex ante*-Entscheidungen fokussieren soll, d.h. auf die Auswahl und Ausbildung der WissenschaftlerInnen (Frey & Osterloh 2006; Krull 2011). Diese Einschätzung, die seit Langem die Grundlage von Eliteinstitutionen wie der Ivy League-Universitäten, der ETHs oder der Max Planck-Institute ist, scheint langsam auf vermehrtes Interesse zu stoßen (so in den Ansätzen der Exzellenzinitiative in Deutschland). Keine Evaluation kann eine derartige *ex ante*-Qualitätssicherung ersetzen. Durch eine qualitativ anspruchsvolle Ausbildung und strikte Auslese der WissenschaftlerInnen wird die *ex post*-Kontrolle im Sinne einer Leistungsangemessenheitsüberprüfung weitgehend überflüssig. Schon jetzt werden in diesen Forschungskontexten maximale Autonomie und Gestaltungsfreiheit gewährt, wenngleich das Vertrauen der staatlichen Geldgeber in die Leistungsausweise ihrer WissenschaftlerInnen nicht groß genug ist, um auf Evaluationen verzichten zu können.

Zu erwähnen ist der Einwand, dass der Ersatz von *ex post*-Kontrollen durch *ex ante*-Selektion nur für SpitzenwissenschaftlerInnen und auf Exzellenz ausgerichtete Institutionen sinnvoll sei. Dies ist nicht notwendig der Fall, kann sich doch gerade eine *ex ante*-Auswahl durchaus an unterschiedlichen Zielen der

Institution und damit an unterschiedlichen Kriterien ausrichten. Denn neben dem Ziel, sich im internationalen Wettbewerb der Spitzenwissenschaft zu behaupten, sind auch andere Ziele für die Wissenschaft wie ihre Leistungen für die Innovation in Wirtschaft und Gesellschaft und ihr Beitrag zur Bildung und Ausbildung zukünftiger BürgerInnen und hochqualifizierter Arbeitskräfte entscheidend für den Erfolg eines nationalen Systems. Schließlich zeichnet sich ein Wissenschaftssystem nicht nur durch die Leistungsfähigkeit seiner Spitze, sondern ebenso durch die seiner Breite in all ihrer Vielfalt aus. WissenschaftlerInnen werden nicht nur für die international wettbewerbsfähige Spitzenwissenschaft gebraucht. Sie versorgen eine Vielzahl anderer wichtiger Wissenschaftsmärkte mit Ausbildungs-, Forschungs- und Dienstleistungen, die eventuell andere Auswahlkriterien benötigen.

Neben der notwendigen Diversität der Zielsetzungen und Ausrichtungen ist auch eine gewisse Diversität der wissenschaftlichen Qualität unvermeidlich. Bruno Frey, der das Modell der *ex ante*-Auswahl befürwortet, konzediert, dass Fehleinschätzungen bei der Auswahl unvermeidlich sind. Als Reibung im System seien sie immer noch verträglicher als die Transaktionskosten und kulturellen Auswirkungen einer *ex post*-Kontrolle. Es geht aber nicht nur um einzelne «Irrtümer», sondern ebenso um die Variabilität der Leistungsverteilung in jedem System, wozu die breite Masse der guten Leistungen gehört, die dennoch nicht zur internationalen Spitzenleistung gezählt werden. «Ein breites Fundament qualitativ hochwertiger wissenschaftlicher Arbeit ist unverzichtbar und Basis für Spitzenleistungen im Wissenschaftssystem» (Wissenschaftsrat 2011). Das sogenannte 80/20-Pareto-Prinzip, wonach 80 % der Leistungen auf 20 % der Akteure entfallen, mag zwar für die Wissenschaft eine Übertreibung sein; eine gewisse quantitative und qualitative Ungleichverteilung gehört aber zu jedem Leistungssystem, egal wie leistungsstark es insgesamt ausfällt. Deswegen sollte der Nachweis einzelner weniger leistungsstarker Personen oder Einheiten nicht zu einem allgemeinen Misstrauen gegenüber der Leistungsfähigkeit und den Bedingungen des Systems führen.

6.2 Zwischen Nutzen und Freiheit der Wissenschaft: Brauchen wir einen neuen Contrat Social für die Wissenschaft?

Wenn Vertrauen in die intrinsische Motivation der WissenschaftlerInnen und die Gestaltungsfreiheit wichtige Bedingungen für ihre Kreativität und Schaffenskraft darstellen, stellt sich schließlich die Frage, unter welchen Bedingungen die staatlichen Geldgeber in der Lage sind, ein solches Vertrauen zu gewinnen und damit auf aufwendige Rechenschaft und die mit ihr verbundenen kontrollierenden, die Wissenschaftskultur schädigenden Instrumente zu verzichten. Unter welchen Bedingungen und mit welchen Erwartungen könnten die Geldgeber den Verzicht auf bestimmte Formen der Kontrolle rechtfertigen? Welche Art von Kommunikation zwischen Wissenschaft und Öffentlichkeit muss gegeben sein, um ein solches Vertrauen zu erlauben?

Die Hochschulforschung ordnet den Aufstieg neuer Leistungsbewertungspraktiken in den Entwurf neuer Governance-Modelle ein, die ein gänzlich verändertes Verständnis von Wissenschaft und deren Institutionen in der Gesellschaft und damit auch von der Rolle des Staates gegenüber der Wissenschaft widerspiegeln bzw. mitbefördern (Power 2008; Neave 2012). Vier tiefgreifende Veränderungen sind in diesem Zusammenhang von entscheidender Bedeutung:

1. *Wissenschaft rechtfertigt sich zunehmend über ihren wirtschaftlichen Nutzen.* Mit der Betonung der Wichtigkeit ihres Beitrags zur «Knowledge Economy» werden Wissenschaft und ihre Institutionen politisch und wirtschaftlich zentraler. Gleichzeitig wird der Beitrag der Wissenschaft immer mehr nur über seinen Nutzen für die Wirtschaft definiert. Damit wird der Beitrag privater Institutionen und das Zusammenspiel zwischen Wissenschaft und Privatwirtschaft gewollt und gefördert. Die Nützlichkeitsrhetorik verstellt den Blick auf andere Beiträge der Wissenschaft: zum menschlichen Drang, sich und seine Welt hemmungslos zu untersuchen, nach dem Sinn und der physischen und historischen Bedingtheit seiner Existenz zu fragen, oder die Bedingungen seiner gesellschaftlichen Praktiken im Hinblick auf andere denkbare

Formen des Zusammenlebens zu durchleuchten. Die neuen Wertschöpfungsdiskurse zielen eher auf materiell greifbare Werte und Veränderungen. Die Anpassungsfähigkeit der Wissenschaft soll sich vor allem in ihrer Fähigkeit, auf die Bedürfnisse der wertschöpfenden Wirtschaft einzugehen, erweisen. Nicht mehr der Abstand der Universität zu Wirtschaft und Gesellschaft erhöht deren Wert, sondern vielmehr deren enges Zusammenspiel.

2. *Wissenschaftliche Institutionen verlieren ihren Sonderstatus.* Verbunden mit der neuen Nähe der Wissenschaft zu ihren externen Partnern lösen sich die institutionellen Grenzen zwischen rein wissenschaftlichen Institutionen und wissenschaftsbasierten Unternehmen auf. Die rein wissenschaftlichen Institutionen wie die Universität verlieren damit zunehmend ihren Sonderstatus («Mode 2 Science», Nowotny/Scott/Gibbons 2001). So werden sie etwa zunehmend den Leistungskriterien und Steuerungsprozeduren der anderen Institutionen unterzogen.
3. *Wissenschaft und ihre Institutionen gehören nicht mehr allein zum öffentlichen Interesse.* Die Verzahnung wissenschaftlicher Interessen mit den Bedürfnissen externer, häufig auch privatwirtschaftlicher Partner lässt eine Vielzahl neuer Mischformen zwischen privaten und öffentlichen Belangen entstehen: Public-Private-Partnerships werden gefördert, IP-Rechte gesichert, Cluster-Strategien für Wissenssektoren entworfen, in denen Universitäten mit privaten Akteuren gemeinsame Ziele verfolgen und Infrastrukturen aufbauen. Stiftungsprofessuren, private Drittmittel, innovationsfördernde Dienstleistungen und Fundraising werden zu Symbolen universitärer Fortschrittlichkeit. Damit sind Universitäten und ihre WissenschaftlerInnen aber nicht mehr ausschließlich im Dienste des öffentlichen Interesses angesiedelt. Das Misstrauen gegenüber «Wettbewerbsverzerrungen» und privater Vorteilsnahme unter dem Dach einer öffentlichen Institution wächst mit der Zunahme privater Hochschul- und Forschungsfinanzierung.
4. *Wissenschaftliche Institutionen gewinnen an Autonomie.* Mit erhöhter institutioneller Autonomie verliert der Staat einen Teil seines steuernden Zugriffs und damit auch die alltägliche Identifikation mit ihren Belangen. Gleichzeitig suchen die

autonomeren Institutionen nach neuen Methoden der Selbststeuerung und greifen zunächst auf Modelle aus anderen institutionellen Kontexten zurück. Häufig werden diese ebenso von staatlicher Seite eingefordert. Die Entwicklung wissenschaftskompatibler Selbststeuerungsformen und Führung ist noch lange nicht abgeschlossen, da die alten Modelle konsensorientierter Kollegialität nicht einfach abgeschafft werden können, sondern an die neuen Entscheidungsanfordernisse angepasst werden müssen.

Je wichtiger Wissenschaft für die Gesellschaft wird und je mehr sie alle Bereiche durchdringt, desto weniger wird ihr ein eigener geschützter Raum zuerkannt. Ironischerweise scheint ihre Allgegenwart und wachsende Bedeutung sie gleichzeitig entthront zu haben. Selbst die Spuren des alten Vorbilds eines fast monastischen Raums kontemplativen Daseins mit seinen eigenen materiellen Entbehrungen, Riten, Formen, Pflichten und Freiheiten sind verwischt und mit ihnen die Anerkennung gesonderter Bedingungen für die wissenschaftliche Arbeit. Angesichts der in Kapitel 4 beschriebenen Entwicklungen liegt der Verdacht nahe, dass das Kind hier mit dem Bade ausgeschüttet wurde, dass nämlich mit der Betonung der Öffnung der Wissenschaft gegenüber gesellschaftlichen und wirtschaftlichen Bedürfnissen die Eigenheit einiger Ziele und Rahmenbedingungen ausgeblendet wurde. Der Wert zweckfreier Räume frei zu gestaltenden Nachdenkens und hemmungslosen Prüfens überkommener Annahmen besteht in ebendieser Unbestimmtheit, durch die sie sich aber auch unternehmerischen Nutzenrechnungen, kurz gesteckten Effizienzzielen und messbarer Überprüfbarkeit gänzlich oder zumindest großenteils entziehen. Die Eigenheit wissenschaftlichen Erkenntnisinteresses mit seinem Wechselbad von mühevoller Kleinstarbeit, unvorhersehbaren Rückschlägen und plötzlichen Heimsuchungen von Begeisterung über neue Einsichten unterliegt eigenen Motivationen, Rhythmen und Belohnungen, die sich kaum durch die Maßeinheiten anderer Arbeitswelten abbilden lassen – es sei denn, diese sind nach dem Vorbild der Wissenschaft entstanden.

So wie große Unternehmen beginnen, ihre R&D-Einheiten anders zu besetzen und zu behandeln als andere Unternehmenseinheiten, sollte auch die finanzielle

rende Politik und Öffentlichkeit die besonderen Rahmenbedingungen und eigenen Funktionsweisen wissenschaftlicher Institutionen verstehen, schützen und fördern. Der enge Dialog zwischen Wirtschaft und Wissenschaft, der in den vergangenen Jahrzehnten so viel Dynamik freigesetzt hat, sollte nicht dazu führen, die Besonderheiten beider Bereiche zu verkennen, aus der ihre Zusammenarbeit Innovationsleistung schöpft. Der Verlust des steuernden Zugriffs des Staates auf die Wissenschaftsinstitutionen, den er durch erhöhte Autonomie erfahren hat und der vor allem als Gewinn an Flexibilität und Anpassungsfähigkeit gemeint war, sollte nicht durch neue Formen der Detailsteuerung entwertet werden. Vielmehr sollte sich das ureigenste Interesse der finanzierenden Öffentlichkeit an ihren wissenschaftlichen Institutionen in einem verdichteten Dialog über die Inhalte, Ziele, Werte und Produkte der Wissenschaft, die unseren Alltag heute oder morgen so stark mitbestimmen, äußern. Diesen zu entwerfen, sollte eine vorrangige Aufgabe der Schweizer Wissenschaftspolitik der nächsten Jahre sein.

Gerade für das Schweizer Wissenschaftssystem drängt sich diese Aufgabe aus mehreren Gründen besonders auf:

1. Der Schweiz wird seit der Existenz der Wissenschaftssystemvergleiche eine besonders starke Wettbewerbsfähigkeit, Leistungsstärke und -dichte im internationalen Vergleich nachgewiesen. Sie hätte ein enges Netz von Kontrollinstrumenten somit weniger nötig.
2. Das Wissenschaftssystem ist nicht so stark unterfinanziert wie viele andere Wissenschaftssysteme, in denen vorteilhafte Ressourcenzuteilungen einem noch größeren Rechtfertigungsdruck unterliegen, da Grundbedürfnisse noch nicht abgedeckt sind.
3. Evaluationen sind in der Schweiz bereits lange etabliert, das System hat insofern eine Reife in seinen Qualitätssicherungssystemen erreicht, die nahelegt, dass der Grenznutzen der Leistungsüberprüfung bei neuen Evaluationen stetig abnimmt.
4. Die Kleinheit des Landes erlaubt eine vertrauensbildende Kommunikationsdichte zwischen Öffentlichkeit, Wirtschaft und Wissenschaft, die traditionell immer wieder hervorragend praktiziert wurde.
5. Die Schweiz pflegt seit Langem eine Kultur der Nachhaltigkeit und wagt traditionell immer wieder eher als viele andere Industrieländer, in langfristigen Perspektiven zu denken. Die Wissenschaft sollte es hier leichter haben, als eigene kulturelle Praxis mit langfristigen Frage- und Problemstellungen verstanden zu werden.

Es wäre insofern denkbar, dass sich die Schweiz international als ein Wissenschaftssystem positioniert, das optimale Rahmenbedingungen für gute Wissenschaft anstrebt. Dazu soll ein neuer Dialog zwischen Wissenschaft und Gesellschaft/Politik/Öffentlichkeit die Überzeugung stärken und verbreiten, dass Vertrauen die zugleich sachadäquate und effiziente Basis des Verhältnisses zwischen beiden Welten darstellt. Konsequente *ex ante*-Selektion von WissenschaftlerInnen, denen nach der Eingangsüberprüfung Vertrauen entgegengebracht wird, soll das auf Misstrauen beruhende Prinzip der dauernden Kontrollevaluation ablösen. Evaluation wird dann richtig eingesetzt, wenn sie situativ statt routinemässig erfolgt, durch wissenschaftliche Inhalte strukturiert ist, formative Ziele verfolgt und Selbstreflexion vorsieht, die den Dialog mit Öffentlichkeit und Wirtschaft sucht und pflegt.

7 Schlussfolgerungen

1. Im letzten Jahrzehnt ist die Vielfalt und Häufigkeit von wissenschaftlichen Leistungsbewertungen nach Einschätzung der WissenschaftlerInnen stark gestiegen, so dass inzwischen ein großer und weiter wachsender Teil ihrer Zeit durch die Bewertung oder Überprüfung von wissenschaftlichen Leistungen statt durch Wissenschaft oder deren Vermittlung beansprucht wird. Dieser Zuwachs an Leistungsüberprüfung ist zum einen die Folge eines wachsenden Anteils der auf *Peer Review* beruhenden Entscheidungen in einer Vielfalt von nationalen und internationalen Wettbewerben um Mittel, Stellen für WissenschaftlerInnen, Veröffentlichungen, Renommée und Sichtbarkeit. Zum anderen folgt er aus den immer aufwendiger werdenden Berichtspflichten zuhanden von Staat, Förderorganisationen und Öffentlichkeit.
2. Beide Hauptarten von Leistungsbewertungen, d.h. sowohl die quantitativen Leistungsbewertungen durch Indikatoren als auch die qualitativen gutachtergestützten Evaluationsmethoden (*Peer Review*), werden als Grundlagen für leistungsorientierte Steuerungsinstrumente verwendet. Die Grenzen dieser Bewertungsmethoden finden aber in hochschulpolitischen Diskussionen über den Ausbau entsprechender leistungsorientierter Steuerungsinstrumente kaum Erwähnung, obwohl beide Arten von Leistungsbewertungen durchaus hinsichtlich ihrer methodischen Mängel, der Verlässlichkeit ihrer Aussagen über wissenschaftliche Qualität und ihrer negativen Auswirkungen auf das Verhalten von WissenschaftlerInnen im Kreuzfeuer der Kritik stehen. Die Grenzen der Leistungsbewertungsmethoden werden zu wenig thematisiert oder transparent kommuniziert.
3. Die Verwendung quantitativer Leistungsindikatoren bei der Mittelzuteilung oder in einflussreichen Forschungsrankings verzerrt das Bild wissenschaftlicher Leistungen, da die Aussagekraft der quantitativen Bewertungen sich nur durch eine Kontextualisierung in der entsprechenden wissenschaftlichen Praxis und deren Rahmenbedingungen erschliesst. Außerdem führt sie zu unerwünschten Verhaltensanpassungen aufseiten der von diesen abhängigen WissenschaftlerInnen. Die Anreizwirkung von quantitativen Leistungsindikatoren auf gesteigerte Produktivität, die nur bei erheblichen Mittelverteilungskonsequenzen verzeichnet werden kann, sollte nicht als gesteigerte Qualität der Wissenschaft missverstanden werden. Soweit der Effekt auf wissenschaftliche Qualität bisher empirisch untersucht wurde, konnten eher nachteilige Effekte nachgewiesen werden.
4. Die Balance zwischen frei und flexibel einsetzbaren, über den Grundauftrag an die Universitäten vergebenen Forschungsmitteln und den über den Wettbewerb eingeworbenen Drittmitteln für befristete Projekte bestimmt die nachhaltige Innovationskapazität eines Wissenschaftssystems. Im letzten Jahrzehnt hat sich der Anteil der befristeten Drittmittel bedenklich zuungunsten der flexibel und langfristig einsetzbaren Forschungsgrundmittel verschoben. An den Fachhochschulen ist die Forschung ganz von der Verfügbarkeit externer Mittel abhängig. Drittmittel haben in ihrer Bedeutung für die Forschungsfinanzierung sowohl über Forschungsprojektmittel als auch bei der Bemessung der Forschungsbeiträge an die Hochschulen deutlich zugenommen. Der Drittmittelbonus in der indikatoren gestützten Mittelverteilung an die Universitäten verstärkt diesen Effekt. Drittmittel sind zwar wichtig für den Wettbewerb und die Anpassung der Forschung an externe Bedürfnisse. Allerdings unterminiert ein zu hoher Drittmittelanteil der Forschungsfinanzierung die langfristige Orientierung der Forschung, mindert die Wahrscheinlichkeit bahnbrechender Forschung und beansprucht einen disproportionalen Anteil an Wissenschaftlerzeit im Rahmen der *Peer Review*.
5. Von allen Formen der Leistungsbewertung wird die formative Evaluation, in der *Peers* die Entwicklungs- und Verbesserungspotenziale der zu evaluierenden Einheit im Dialog mit dieser beleuchten, von WissenschaftlerInnen, Hochschulangehörigen und in empirischen Untersuchungen am positivsten bewertet sowie als Qualitätsentwicklungsin-

strument anerkannt. Allerdings beginnt deren zunehmende Häufigkeit und Routinisierung im Zuge kurzer, von außen festgelegter Evaluationszyklen, die Effektivität dieser Methode auszuhöhlen. Mit steigender Frequenz der Evaluationen nimmt der Grenznutzen zwischen Lerneffekt und Leistungs-optimierung auf der einen Seite und dem Aufwand der Evaluation auf der anderen stetig ab. Nach zwei Jahrzehnten des Aufbaus und der Konsolidierung der Qualitätssicherung im Allgemeinen und der Fachevaluationen im Besonderen ist der Grenznutzen von Evaluationen an vielen Universitäten in der Schweiz gering.

6. Wissenschaftliche Praxis lebt von einem hohen Maß an Gestaltungsfreiheit und Selbstbestimmung in den Entscheidungen über Inhalte, Ziele, Methoden und Prozesse der Forschung, die wiederum ein entsprechend hohes Maß an intrinsischer Motivation unter WissenschaftlerInnen zur Folge haben. Diese rechtfertigen für viele WissenschaftlerInnen die hohen persönlichen Kosten, die mit einer wissenschaftlichen Laufbahn verbunden sind (dazu gehören die im Verhältnis zu Qualifikationen und Zeiteinsatz geringe Entlohnung und die im Vergleich zu anderen Professionen lange Phase der beruflichen Unsicherheit).

Gerade diese Grundbedingungen des wissenschaftlichen Lebens, die intrinsische Motivation, die hohe Leistungsbereitschaft und Risikofreudigkeit der WissenschaftlerInnen, werden durch ständige Berichtspflichten, Begutachtungs- und Evaluationsprozesse langsam erodiert. Kontrollierende Formen der Leistungsbewertung in Akkreditierungen und Evaluationen unterminieren somit die kulturelle Grundlage und Attraktivität wissenschaftlicher Laufbahnen.

7. Eine *ex ante*-Steuerung der Wissenschaft über Ausbildung und Auswahl der WissenschaftlerInnen gemäß den Zielen einer Institution ist wesentlich effektiver als jegliche Versuche, Leistungen *ex post* zu kontrollieren oder scheinbar anzureizen. So kann die Vielfalt institutioneller Ziele und verschiedener Wissenschaftsbereiche sich in entsprechenden Auswahlkriterien widerspiegeln statt durch einheitliche Bewertungsstandards über ver-

schiedene Hochschulen hinweg die externe und interne Differenzierung in und zwischen Hochschulen, die eigentlich zu fördern wäre, zu unterbinden. Die Qualität der Auswahlprozesse bemisst sich daher nicht nur an der Orientierung an internationalen Höchststandards einer Fachdisziplin, sondern auch an der Sensibilität gegenüber den Ausrichtungen der Institution, der jeweiligen wissenschaftlichen Einheiten und den an sie gestellten Anforderungen.

8. Die zentrale Zielvorstellung des New Public Management, in der Wissenschaft Effizienz über Leistungsanreize und -kontrollen zu steigern, verkennt das Wesen wissenschaftlicher Prozesse, in denen unvorhersehbare Entwicklungen, Umwege, gescheiterte Versuche und das Risiko des Scheiterns einer Idee oder der Falsifikation einer Hypothese zur Lebensader des Fortschritts und der langfristigen Neuerung gehören. Ein auf Bestätigung und stetige Produktion ausgerichteter Wissenschaftsbetrieb droht langfristig zu sklerotisieren und eine mittelmäßige Wissenschaftsqualität zu befördern. Ebendiese Mittelmäßigkeit wird durch die sich ausbreitende Misstrauenskultur und das immer dichter werdende Netz von Qualitätskontrollmaßnahmen verstärkt.

8 Literatur

- AUBR (2010). *Assessing Europe's University-Based Research*. Expert Group on Assessment of University-Based Research, Europäische Kommission.
- Böhmer, Susan/Neufeld, Jörg/Hinze, Sybille/Klode, Christian/Hornbostel, Stefan (2011). *Wissenschaftler-Befragung (2010). Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten*. iFQ Working Paper No. 8. Bonn.
- Bornmann, Lutz/Daniel, Hans-Dieter (2003). «Begutachtung durch Fachkollegen in der Wissenschaft», in: Schwarz, Stefanie/Teichler, Ulrich (Hrsg.), *Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung*. Frankfurt: Campus, 211–230.
- Csikszentmihalyi, Mihaly (1996). *Creativity. Flow and the Psychology of Discovery and Invention*. New York: Harper Collins.
- Curaj, Adrian/Scott, Peter/Vlasceanu, Lazăr/Wilson, Lesley (2012). *European Higher Education at the Crossroads. Between the Bologna Process and National Reforms. Part V: Higher Education Governance in the European Higher Education Area*. Heidelberg: Springer.
- Daniel, Hans-Dieter (1993). *Guardians of Science. Fairness and Reliability of Peer Review*. Weinheim: VCH.
- Daniel, Hans-Dieter/Mittag, Sandra/Bornmann, Lutz (2007). «The Potential and Problems of Peer Evaluation in Higher Education and Research», in: Cavalli, Alessandro (Hrsg.), *Quality Assessment for Higher Education in Europe*. London: Portland Press, 71–82.
- Deci, Edward/Koestner, Richard/Ryan, Richard (1999). «A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation», in: *Psychological Bulletin* 125 (6), 627–668.
- Deci, Edward/Ryan, Richard (1985). *Intrinsic Motivation and Self-determination in Human Behaviour*. New York: Plenum.
- EUA (2008). *Financially Sustainable Universities. Towards Full Costing in European Universities. An EUA Report*. Brussels: Publications of the European University Association.
- EUA (2009). *Improving Quality, Enhancing Creativity: Change Processes in European Higher Education Institutions. Final Report of the Quality Assurance for the Higher Education Change Agenda (QAHECA) Project*. Brussels: Publications of the European University Association.
- EUA (2010). *Trends 2010: A Decade of Change in European Higher Education*. Brussels: Publications of the European University Association.
- EUROHORCS (2006). *Strengthening European Research Cooperation through Partnership. EUROHORCS Perspective on Networking and Coordination of National Programmes*. www.eurohorcs.org/SiteCollectionDocuments/Networking_and_Coordination_of_National_Programmes.pdf (letzter Zugriff 27.5.2013).
- Evidence Ltd (2005). *Impact of Selective Funding of Research in England and the Specific Outcome of HEFCE Research Funding*. Report to HEFCE and the Department for Education and Skills.
- Fehr, Ernst/Falk, Armin (1999). «Wage Rigidity in a Competitive Incomplete Contract Market», in: *Journal of Political Economy* 107 (1), 106–134.
- Ferlie, Ewan/Musselin, Christine/Andresani, Gianluca (2009). «The Governance of Higher Education Systems: A Public Management Perspective», in: Paradeise, Catherine/Reale, Emanuela/Bleiklie, Ivar/Ferlie, Ewan (Hrsg.), *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer, 1–19.
- Fischer, Klaus (2003). «Soziale und kognitive Aspekte des Peer Review-Verfahrens», in: Fischer, Klaus/Parthey, Heinrich (Hrsg.), *Evaluation wissenschaftlicher Institutionen: Wissenschaftsforschung Jahrbuch 2003*, Berlin: Gesellschaft für Wissenschaftsforschung, 23–62.
- Frey, Bruno S. (2006). «Evaluitis – Eine neue Krankheit», *IEW Working Papers* No. 293, Institute for Empirical Research in Economics, University of Zurich. www.fh-dortmund.de/de/studi/fb/2/personen/lehr/wilhelm/medien/evalu_doku/Evaluitis_BrunoFrey.pdf (letzter Zugriff 27.5.2013).
- Frey, Bruno S. (2007). «Evaluierungen, Evaluierungen... Evaluitis», in: *Perspektiven der Wirtschaftspolitik* 8 (3), 207–220.
- Frey, Bruno S./Osterloh, Margit (2006). «Evaluations: Hidden Costs, Questionable Benefits and Superior Alternatives», *IEW Working Papers* No. 32, Institute for Empirical Research in Economics, University of Zurich. www.iew.unizh.ch/wp/iewwp302.pdf (letzter Zugriff 17.12.2012).
- Henkel, Mary (1999). «The Modernisation of Research Evaluation: The Case of the UK», in: *Higher Education* 38, 105–122.
- Hornbostel, Stefan (1997). *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westdeutscher Verlag.
- Hornbostel, Stefan/Schelling, Anna (Hrsg.) (2011). *Evaluation: New Balance of Power*. iFQ Working Paper No. 9.
- Horrobin, David F. (1990). «The Philosophical Basis of Peer Review and the Suppression of Innovation», in: *Journal of the American Medical Association* 263, 1438–1441.
- Jacob, Anna-Katharina/Teichler, Ulrich (2011). *Der Wandel des Hochschullehrerberufs im internationalen Vergleich. Ergebnisse einer Befragung in den Jahren 2007/08*. Bonn und Berlin: BMBF. www.uni-kassel.de/wz1/pdf/BMBF_Hochschullehrerstudie2011_Druck.pdf (letzter Zugriff 20.12.2012).
- Kieser, Alfred (2010). «Unternehmen Wissenschaft?», in: *Leviathan. Berliner Zeitschrift für Sozialwissenschaft* 38, 347–376.
- Kohn, Alfie (1993). *Punished by Rewards*. New York: Plenum Press.
- Krull, Wilhelm (2011). «Bewertung, Begutachtung und Evaluation in Wissenschaft und Forschung», in: Hornbostel, Stefan/Schelling, Anna (Hrsg.), *Evaluation: New Balance of Power*. iFQ Working Paper No. 9, 15–23.
- Lamont, Michèle (2009). *How Professors Think: Inside the Curious World of Academic Judgement*. Cambridge: Harvard University Press.
- Lamont, Michèle (2011). «Pragmatic Fairness: Production of the Sacred while Observing the Rules», in: Hornbostel, Stefan/Schelling, Anna (Hrsg.), *Evaluation: New Balance of Power*. iFQ Working Paper No. 9, 47–56.
- Langfeldt, Liv (2004). «Expert Panels Evaluating Research: Decision-making and Sources of Bias», in: *Research Evaluation* 13 (1), 51–62.
- League of European Research Universities (LERU) (2012). *Research Universities and Research Assessment*. Brussels: Publications of the League of Research Universities.
- Lee, Frederic S. (2007). «Research Assessment Exercise, the State and the Dominance of Mainstream Economics in British Universities», in: *Cambridge Journal of Economics* 31, 309–325.
- Lepper, Mark/Greene, David (Hrsg.) (1978). *The Hidden Cost of Reward: New Perspectives on the Psychology of Human Motivation*. New York: John Wiley.
- Loukkola, Tia/Zhang, Thérèse (2010). *Examining Quality Culture: Part 1 – Quality Assurance Processes in Higher Education Institutions*. Brussels: Publications of the European University Association.
- Matthies, Hildegard/Simon, Dagmar (Hrsg.) (2008). *Wissenschaft unter Beobachtung. Effekt und Defekt von Evaluationen*. *Leviathan: Zeitschrift für Sozialwissenschaft*. Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften.
- May, Robert (2004). «The Pathologies of the RAE», in: *The Journal of the Foundation for Science and Technology* 18 (4), June 2004, 3–4.
- Mittag, Sandra/Bornmann, Lutz/Daniel, Hans-Dieter (2006). «Qualitätssicherung und -verbesserung von Studium und Lehre durch Evaluation. Akzeptanz und Folgen mehrstufiger Evaluationsverfahren», in: *Beiträge zur Hochschulforschung* 28 (2), 6–27.

→ 8 Literatur

- Moed, Henk F./Luwel, Marc/Nederhof, Anton J. (2002). «Towards Research Performance in the Humanities», in: *Library Trends* 50, 498–520.
- Moxham, Harriet/Anderson, Joe (1992). «Peer Review. A View from the Inside», in: *Science and Technology Policy*, 7–15.
- Neave, Guy (2012). *The Evaluative State, Institutional Autonomy and Re-engineering Higher Education in Western Europe: the Prince and his Pleasure*. Houndmills/Basingstoke/Hampshire: Palgrave Macmillan.
- Nowotny, Helga/Scott, Peter/Gibbons, Michael (2001). *Re-thinking Science. Knowledge and the Public in an Age of Uncertainty*. Cambridge: Blackwell.
- Olbrecht, Meike/Bornmann, Lutz (2010). «Panel Peer Review of Grant Applications: What Do We Know from Research in Social Psychology on Judgment and Decision-Making in Groups?», in: *Research Evaluation* 19 (4), 293–304.
- Olbrecht, Meike/Klein, Thamar (2011). «SFB-Begutachtung: Entscheidungsfindung in Gruppen», in: Hornbostel, Stefan/Schelling, Anna (Hrsg.), *Evaluation: New Balance of Power. iFQ Working Paper No. 9*, 33–45.
- Overbeke, John/Wager, Elizabeth (2003). «The State of the Evidence: What We Know and What We Don't Know about Journal Peer Review», in: Godlee, Fiona/Jefferson, Tom (Hrsg.), *Peer Review in Health Sciences*. London: BMJ Books, 45–61.
- PA Consulting Group, (2009). *RAE 2008 Accountability Review*. HEFCE.
- Paradeise, Catherine (2011). «La réception des démarches d'évaluation», Vortrag anlässlich des Colloque CRUS *Mesurer les performances de la recherche: enjeux et perspectives*, Lausanne, 7. Oktober 2011.
- Paradeise, Catherine (2012). «Tools and Implementation for a New Governance of Universities: Understanding Variability Between and Within Countries», in: Curaj, Adrian/Scott, Peter/Vlasceanu, Lazăr/Wilson, Lesley (Hrsg.) (2012), *European Higher Education at the Crossroads. Between the Bologna Process and National Reforms. Part V: Higher Education Governance in the European Higher Education Area*. Heidelberg: Springer, 573–598.
- Paradeise, Catherine/Reale, Emanuela/Goastellec, Gaële (2012). «A Comparative Approach to Higher Education Reforms in Western European Countries», in: Paradeise, Catherine/Reale, Emanuela/Bleiklie, Ivar/Ferlie, Ewan (Hrsg.), *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer, 197–226.
- Peters, Douglas P./Ceci, Stephen J. (1982). «Peer-review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again», in: *The Behavioral and Brain Sciences* 5 (2), 187–195.
- Power, Michael (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Power, Michael (2008). «Research Evaluation in the Audit Society», in: Matthies, Hildegard/Simon, Dagmar (Hrsg.), *Wissenschaft unter Beobachtung. Effekt und Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft*, Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, 15–24.
- Rauhvargers, Andrejs (2011). *Global University Rankings and their Impact*. Brussels: EUA Publications.
- Röbbecke, Martina (2007). «Evaluation als neue Form der (Disziplinierung) – ein nicht intendierter Effekt?», in: Matthies, Hildegard/Simon, Dagmar (Hrsg.), *Wissenschaft unter Beobachtung. Effekt und Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft*. Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, 161–177.
- Röbbecke, Martina/Simon, Dagmar (2001). *Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten*. Berlin: edition sigma.
- Schweizerische Universitätskonferenz (SUK) (2006). *Richtlinien für die Qualitätssicherung an den schweizerischen universitären Hochschulen (Qualitätssicherungs-Richtlinien) vom 7. Dezember 2006*.
- Torka, Marc (2011). «Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen», in: Hornbostel, Stefan/Schelling, Anna (Hrsg.), *Evaluation: New Balance of Power. iFQ Working Paper No. 9*, 69–81.
- van Raan, Anthony (1995). «Bewertung von Forschungsleistungen: Fortgeschrittene bibliometrische Verfahren», in: Müller-Böling, Detlef (Hrsg.), *Qualitätssicherung in Hochschulen. Forschung – Lehre – Management*. Gütersloh: CHE der Bertelsmann-Stiftung, 85–102.
- van Raan, Anthony (2005). «Measurement of Central Aspects of Scientific Research: Performance, Interdisciplinarity, Structure», in: *Measurement* 3 (1), 1–19.
- Weingart, Peter (1995). «Forschungsindikatoren: Instrumente politischer Legitimierung oder organisatorischen Lernens», in: Müller-Böling, Detlef (Hrsg.), *Qualitätssicherung in Hochschulen. Forschung – Lehre – Management*. Gütersloh: CHE der Bertelsmann-Stiftung, 73–84.
- Weingart, Peter (2005a). «Impact of Bibliometrics upon the Science System: Inadvertent Consequences?», in: *Scientometrics* 62 (1), 117–131.
- Weingart, Peter (2005b). «Das Ritual der Evaluierung und die Verführbarkeit der Zahlen», in: Ders. (Hrsg.), *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit*. Weilerwist: Velbrück, 102–122.
- Wenneras, Christine/Wold, Agnes (1999). «Bias in Peer Review of Research Proposals», in: Godlee, Fiona/Jefferson, Tom (Hrsg.), *Peer Review in Health Sciences*. London: BMJ Books, 79–89.
- Whitley, Richard/Engwall, Lars/Glaser, Jochen (Hrsg.) (2010). *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation*. Oxford: Oxford University Press.
- Wissenschaftsrat (2011). *Neuere Entwicklungen der Hochschulfinanzierung in Deutschland. Bericht des Vorsitzenden zu aktuellen Tendenzen im Wissenschaftssystem vom 8. Juli 2011*.

Anhang

Übersicht zu Forschungsbewertung nach Nutzungsgruppen und -zielen

Quelle: AUBR-Report, Assessing Europe's University-Based Research. Expert Group on Assessment of University-Based Research, Europäische Kommission 2010.

User Group	Why Is Research Assessment Data Required?	What Research Assessment Data Is Required?
HE Management and Governance		
Governing Bodies/Councils	<ul style="list-style-type: none"> • Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value for money and efficiency • Quality assurance 	<ul style="list-style-type: none"> • Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-à-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally
HE Executives/Management	<ul style="list-style-type: none"> • Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value for money and efficiency • Quality assurance • Publicity • Student and academic recruitment • Improve and benchmark performance and quality 	<ul style="list-style-type: none"> • Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-à-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of partnerships (academic, public/private sector, NGOs, research organisations, etc.)
HE Research Groups	<ul style="list-style-type: none"> • Strategic positioning • Research strategy development/management • Investor confidence/value for money and efficiency • Student and academic recruitment 	<ul style="list-style-type: none"> • Discipline data re. level of intensity, expertise, quality and competence benchmarked against peer institutions • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of partnerships (academic, public/private sector, NGOs, research organisations, etc.)
Government and Government Agencies		
EU and National Governments	<ul style="list-style-type: none"> • Define policy and inform decisions about HE system and HEIs • Determine national/international competitiveness • Quality, sustainability, relevance and impact of research activity • Investor confidence/value for money and efficiency • Improve performance and quality • Improve system functionality 	<ul style="list-style-type: none"> • System and institutional data re. level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency
Ministries of Education/Higher Education or Enterprise and Employment	<ul style="list-style-type: none"> • Policy and planning • Strategic positioning of HE institutions • Quality, sustainability, relevance and impact of research activity • Research strategy development/management • Investor confidence/value for money and efficiency • Quality assurance 	<ul style="list-style-type: none"> • Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Indicator of national competitiveness • Performance of HE system and individual institutions • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency

Local and Regional Governments	<ul style="list-style-type: none"> • Define local/regional policy and competitiveness • Quality, sustainability, relevance and impact of research activity • Improve integration/collaboration between universities, government and private sector • Improve attraction capacity 	<ul style="list-style-type: none"> • Benchmarking performance and quality of HE system/ institutions nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-à-vis funding
HE Agencies	<ul style="list-style-type: none"> • Define policy and inform decisions about HE system and HEIs • Quality, sustainability, relevance and impact of research activity • Determine national/international competitiveness • Investor confidence/value for money and efficiency • Improve performance and quality • Improve system functionality 	<ul style="list-style-type: none"> • System and institutional data re. level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency
Other Government Agencies	<ul style="list-style-type: none"> • Improve and benchmark performance and quality • Aid resource allocation • Investor confidence/value for money and efficiency 	<ul style="list-style-type: none"> • Benchmarking performance and quality of HE system institutions nationally and worldwide
Academic Organisations		
	<ul style="list-style-type: none"> • Benchmark professional and academic performance and quality • Student and academic recruitment 	<ul style="list-style-type: none"> • Academic and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Quality of academic staff and PhD students
Individuals		
Academics and Researchers	<ul style="list-style-type: none"> • Identify career opportunities • Identify research partners • Identify best research infrastructure and support for research 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Employment conditions • Impact of research on teaching, staff/student ratio • Institutional research support
Students	<ul style="list-style-type: none"> • Inform choice of HEI • Identify career opportunities 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Research capacity of institution and research team, e.g. graduate students/academic ratio, age of PhD students, time to completion, structure/characteristics of PhD programme and support • Graduate career and employment trends • Quality of the research infrastructure • Staff/student ratio

Peer HEIs		
	<ul style="list-style-type: none"> • Identify peer HEIs and best research partners 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institutions and researchers benchmarked against peers in field of interest • Research capacity of institution and research team • Potential for partnership
Industry Partner Organisations		
Private Firms and Entrepreneurs	<ul style="list-style-type: none"> • Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Public Organisations	<ul style="list-style-type: none"> • Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Employers	<ul style="list-style-type: none"> • Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees 	<ul style="list-style-type: none"> • Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Civic Society and Civic Organisations		
	<ul style="list-style-type: none"> • Identify specific expertise and information • Identify potential collaborator • Identify consultancy, technology transfer and knowledge transfer partners 	
Ministries of HE in Developing Countries		
	<ul style="list-style-type: none"> • To help determine which foreign higher education institutions are applicable for overseas scholarships studies • To help determine research partnerships for knowledge and technology transfer 	

Sponsors and Private Investors

Benefactors/ Philanthropists	<ul style="list-style-type: none"> • Determine institutional performance vis-à-vis national and international competitors • Investor confidence/value for money and efficiency • Quality, sustainability, relevance and impact of research activity • Quality of academic staff and PhD student • Contributor to own brand image 	
Alumni	<ul style="list-style-type: none"> • Determine institutional performance vis-à-vis national and international competitors • Investor confidence/value for money and efficiency • Quality of academic staff and PhD students activity • Quality of academic staff and PhD student • Reflect pride and career aspirations/reputation 	<ul style="list-style-type: none"> • Institutional data re. level of quality and international competitiveness • Benchmarking between nationally and worldwide • Quality, sustainability, relevance and impact of research

Public Opinion

	<ul style="list-style-type: none"> • Determine institutional performance quality and international competitors • Quality, sustainability, relevance and impact of research • Investor/parental confidence and value for money 	<ul style="list-style-type: none"> • Institutional data re. level of intensity, expertise and competence • Benchmarking against peer institutions, nationally and activity worldwide • Student choice and career opportunities • Indicator of national competitiveness • Performance of HE system and individual institutions • Efficiency level: how much output vis-à-vis funding
--	--	---

Abkürzungsverzeichnis

AHELO	Assessment of Higher Education Learning Outcomes
ARWU	Academic Ranking of World Universities
AUBR	Assessment of University-Based Research
BMBF	Bundesministerium für Bildung und Forschung
CHE	Centrum für Hochschulentwicklung
CRUS	Conférence des recteurs des universités suisses
CSST	Conseil suisse de la science et de la technologie
DFG	Deutsche Forschungsgemeinschaft
ETH	Eidgenössische Technische Hochschule
EU	Europäische Union
EUA	European University Association
EUROHORCS	European Heads of Research Councils
HE	Higher Education
HEFCE	Higher Education Funding Council for England
HEI	Higher Education Institution
IEW	Institut für Empirische Wirtschaftsforschung
iFQ	Institut für Forschungsinformation und Qualitätssicherung
IP	Intellectual Property
LERU	League of European Research Universities
NGO	Non-governmental Organization
OECD	Organization for Economic Co-operation and Development
QAHECA	Quality Assurance for the Higher Education Change Agenda
QS	Qualitätssicherung
R&D	Research and Development
RAE	Research Assessment Exercise
REF	Research Excellence Framework
SFB	Sonderforschungsbereiche
SUK	Schweizerische Universitätskonferenz
SWTR	Schweizerischer Wissenschafts- und Technologierat
UFG	Universitätsförderungsgesetz
UK	United Kingdom
WKN	Wissenschaftliche Kommission Niedersachsen
ZEVA	Zentrale Evaluations- und Akkreditierungsagentur Hannover

Impressum

Schweizerischer Wissenschafts-
und Technologierat SWTR
Hallwylstrasse 15
CH-3003 Bern
T 0041 (0)31 323 00 48
F 0041 (0)31 323 95 47
swtr@swtr.admin.ch
www.swtr.ch

ISBN 978-3-906113-04-3

Bern 2013

Lektorat: Doris Tranter, Stéphane Gillioz
Gestaltung: VischerVettiger, Basel
Titelfoto: Mélanie Roullier

Schweizerischer Wissenschafts- und Technologierat SWTR
Hallwylstrasse 15
CH-3003 Bern
T 041 31 323 00 48
F 041 31 323 95 47
swtr@swtr.admin.ch
www.swtr.ch