

**Qualitätsmessung: E. Lütkeimer: Das Forschungsrating des
Wissenschaftsrats**
von Elke Lütkeimer

Im Zuge des tiefgreifenden Wandels, in dem sich das deutsche Hochschul- und Wissenschaftssystem derzeit befindet, haben öffentliche Leistungsvergleiche wissenschaftlicher Einrichtungen erheblich an Bedeutung gewonnen. Hierzu zählen insbesondere Rankings, deren Ergebnisse von den Hochschulen mit großer Aufmerksamkeit verfolgt und, sofern sie günstig ausgefallen sind, in der eigenen Öffentlichkeitsarbeit gerne zitiert werden. Viele Hochschulen übernehmen auch Indikatoren, die für Rankings verwendet werden, in ihre internen Zielvereinbarungen, um ihre Platzierungen zu verbessern. Studienanfänger wiederum nutzen oftmals Rankings zur Orientierung bei der Suche nach einem geeigneten Studienort.

Angesichts der Folgen, die Rankings für die wissenschaftlichen Institutionen haben, und ihrer häufig unkritischen Aufnahme erscheint es jedoch umso problematischer, dass die Methodik vieler Rankings nicht transparent ist und die Wissenschaft bei ihrer Fortentwicklung keine Mitsprache hat. Der Wissenschaftsrat¹ hat sich daher eingehend mit Zielen und Methoden von Rankings befasst und im November 2004 Empfehlungen zu Rankings im Wissenschaftssystem vorgelegt. Darin hat er bestehende Rankings einer methodischen Kritik unterzogen und ein Verfahren für ein Forschungsrating entwickelt. Hierbei handelt es sich um ein neuartiges Verfahren zur vergleichenden Bewertung von Forschungsleistungen. Eine erste Pilotstudie in den beiden Fächern Chemie und Soziologie hat gezeigt, dass das Verfahren zu aussagekräftigen Ergebnissen führt und einen wesentlichen Fortschritt in der methodischen Entwicklung von Verfahren der Forschungsbeurteilung darstellt.² Im Mai 2008 hat der Wissenschaftsrat daraufhin

¹Der Wissenschaftsrat berät die Bundesregierung und die Regierung der Länder in Fragen der inhaltlichen und strukturellen Entwicklung der Hochschulen, der Wissenschaft und der Forschung.

²Die Empfehlungen zu Rankings im Wissenschaftssystem, Teil 1: Forschung (2004),

beschlossen, das Verfahren für ein Forschungsrating schrittweise weiterzuentwickeln. In einem nächsten Schritt sollen zwei weitere Fächer, und zwar eine Geistes- und eine Technikwissenschaft, in das Forschungsrating einbezogen werden, um zum einen die Vorschläge zur Weiterentwicklung des Verfahrens aus der Pilotstudie umzusetzen und zum anderen zu überprüfen, ob das Verfahren auch in diesen Wissenschaftsgebieten, die hinsichtlich der Bewertung von Forschungsleistungen als schwierig gelten, anwendbar ist.

Die für die Durchführung des Forschungsratings verantwortliche Steuerungsgruppe hat für die Geisteswissenschaften die Geschichtswissenschaft ausgewählt. Ausschlaggebend war die Überlegung, die Anpassung des Verfahrens möglichst an einer der repräsentativeren Geisteswissenschaften zu erproben. Zudem sprechen die Größe, die innere Komplexität, die interdisziplinären Bezüge sowie die Vielzahl außeruniversitärer Forschungseinrichtungen für die Wahl der Geschichtswissenschaft.

Die Entscheidung des Wissenschaftsrats, ein geisteswissenschaftliches Fach in die Weiterentwicklung des Forschungsratings einzubeziehen, stößt auf eine lebhafte Debatte zur Qualität von Forschungsleistungen in den Geisteswissenschaften und den Möglichkeiten ihrer Bewertung. Angestoßen wurde diese Debatte sicher nicht zuletzt auch durch die Empfehlungen des Wissenschaftsrats zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland³, in denen einerseits die Leistungen der Geisteswissenschaften entgegen der bis dahin verbreiteten Krisenbeschwörung eine besondere Würdigung erfahren haben. Andererseits hatte der Wissenschaftsrat in seinen Empfehlungen aber auch konkrete Forderungen erhoben, eine Verständigung über disziplinäre Standards herzustellen und angemessene und formalisierte Bewertungsverfahren und -kriterien zu entwickeln. Das

der Bericht der Steuerungsgruppe über die Pilotstudie Forschungsrating und aktuelle Informationen zur Weiterentwicklung des Forschungsratings sind unter <<http://www.forschungsrating.de>> (11.05.2009) abrufbar.

³Wissenschaftsrat, Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland, Köln 2006.

Forschungsrating des Wissenschaftsrats ist in besonderer Weise geeignet diesem Desiderat Rechnung zu tragen, da es von seiner Konzeption ausdrücklich darauf ausgerichtet ist, fachspezifische Besonderheiten zu erfassen und einer Bewertung zugänglich zu machen. Es bietet somit eine gute Möglichkeit, zur Klärung von Qualitätskriterien in den Geisteswissenschaften beizutragen. Im Folgenden sollen daher die Grundzüge und Chancen dieses Verfahrens näher beleuchtet und insbesondere die Möglichkeiten einer fachspezifischen Anpassung des Verfahrens an Beispielen der ersten Pilotstudie in den Fächern Chemie und Soziologie verdeutlicht werden.

Wissenschaftsgeleitetes Verfahren

Wissenschaft kann nur von Wissenschaftlern angemessen beurteilt werden. Zu den wichtigsten Wesensmerkmalen des Forschungsratings zählt daher, dass das Verfahren explizit von der Wissenschaft gestaltet wird. Vertreten wird die Wissenschaft im Forschungsrating durch Fachgutachter in den Bewertungsgruppen und durch die Beteiligung der jeweiligen Fachgesellschaften.

Verantwortlich für die erste Pilotstudie sowie auch für die Weiterentwicklung des Forschungsratings ist eine Steuerungsgruppe, die neben Mitgliedern der wissenschaftlichen Kommission des Wissenschaftsrats und weiteren Sachverständigen auch institutionelle Vertreter der großen Wissenschaftsorganisationen (Deutsche Forschungsgemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft, Hochschulrektorenkonferenz, Max-Planck-Gesellschaft, Leibniz-Gemeinschaft) auf Vizepräsidenten-Ebene sowie Gäste aus Länderministerien und dem Bundesministerium für Bildung und Forschung umfasst. Für die ausgewählten Fächer setzt die Steuerungsgruppe jeweils eine Bewertungsgruppe ein. Besetzungsvorschläge werden dabei von den genannten Wissenschaftsorganisationen und den jeweils einschlägigen Fachgesellschaften erbeten.⁴ In der ersten

⁴Die erste Pilotstudie wurde im Fach Chemie aktiv von der Gesellschaft Deutscher Chemiker (GDCh) sowie von dem Verband der Chemischen Industrie (VCI) und im Fach Soziologie von der Deutschen Gesellschaft für Soziologie (DGS) unterstützt.

Pilotstudie bestanden die Bewertungsgruppen aus 15 bzw. 16 Gutachtern, die in der Bewertungsphase durch weitere Sondergutachter unterstützt wurden. Bei der Besetzung der Bewertungsgruppen wurde insbesondere auf eine breite Abdeckung der wichtigsten Teilgebiete des jeweiligen Fachs sowie auf die internationale Erfahrung der Gutachter geachtet. Zu den wesentlichen Aufgaben der Bewertungsgruppen zählen die definitive Abgrenzung des Fachgebiets, die fachspezifische Umsetzung der Kriterien (das heißt die Auswahl der Informationen, die der Bewertung zugrunde gelegt werden), die Entwicklung der entsprechenden Fragebögen sowie natürlich die Vornahme der eigentlichen Bewertungen.

Mehrdimensionalität der Bewertung

In dem Forschungsrating des Wissenschaftsrats werden die Forschungsleistungen von Universitäten und außeruniversitären Instituten in drei Dimensionen (Forschung, Nachwuchsförderung, Wissenstransfer) nach insgesamt sechs Kriterien (I. Forschungsqualität, II. Impact/Effektivität, III. Effizienz, IV. Nachwuchsförderung, V. Transfer in andere gesellschaftliche Bereiche, VI. Wissensvermittlung und -verbreitung) bewertet, wobei diese Ergebnisse ungewichtet nebeneinander gestellt werden.

Jeder Einrichtung werden zwei Berichterstatter nach fachlicher Kompetenz und unter Ausschluss von Befangenheiten zugeordnet. Zunächst nehmen diese individuelle Bewertungen vor, die anschließend in mehreren Sitzungen der gesamten Bewertungsgruppe plenar diskutiert und abgestimmt werden. Die Bewertungen werden auf einer fünfstufigen Notenskala von nicht befriedigend bis exzellent angegeben. Dabei wird nicht jeder Indikator unmittelbar in eine bestimmte Note übertragen, sondern immer mehrere quantitative und qualitative Angaben zusammen betrachtet. Die Fachgutachter können aus der Übersicht der verschiedenen Indikatoren ein differenziertes Urteil geben, das somit nie allein durch Zahlenwerte bestimmt ist. Dieses Vorgehen würde seinen Sinn verlieren, wenn abschließend eine Gesamtnote über alle Kriterien ermittelt würde. Folglich findet

keine Verrechnung zu einem Gesamtwert statt und es werden auch keine Rangreihenfolgen mit Rangplätzen gebildet. Dadurch dass die Bewertungen aller Kriterien nebeneinander bestehen bleiben, kann für jede Einrichtung ein individuelles Profil erstellt werden, das ihre Leistungen in den einzelnen Bewertungskriterien unabhängig voneinander zeigt. Die verschiedenen Aufträge der einzelnen Einrichtungen werden auf diese Weise berücksichtigt und individuelle Stärken und Schwächen sichtbar gemacht.

Hinsichtlich der Dimension Forschung ist hervorzuheben, dass zwischen dem absoluten Beitrag der Einrichtung zur Entwicklung der Wissenschaft im Fachgebiet und darüber hinaus („Impact/Effektivität“) und einem auf die Größe bzw. den Ressourceneinsatz relativierten Beitrag („Effizienz“) unterschieden wird. Dadurch wird sowohl sichtbar, welche Einrichtungen im jeweiligen Fach eine wichtige Rolle spielen, als auch, welche Einrichtungen die vorhandenen/eingesetzten Mittel am besten nutzen. Eine weitere Differenzierung wird mit dem Kriterium Forschungsqualität vorgenommen, das auf Ebene sogenannter Forschungseinheiten bewertet wird. Bei der Definition dieser Forschungseinheiten zeigten sich in den Fächern der ersten Pilotstudie deutliche fachspezifische Unterschiede, die den Kooperationsgepflogenheiten der beiden Fächer Rechnung tragen: So wurden in der Chemie in der Regel größere Einheiten mit durchschnittlich sechs leitenden Wissenschaftlern gemeldet (Beispielsweise „Institut für Physikalische Chemie“, „Institut für Biochemie“), während in der Soziologie fast 75 Prozent aller Einheiten nur je eine Professur umfasste (Lehrstuhlprinzip).

Fachspezifische Operationalisierung

Die Dimensionen und Kriterien stellen den übergeordneten Rahmen des Forschungsratings dar, der möglichst für alle Fächer gelten soll. Die einzelnen erhobenen Informationen dagegen, welche die eigentliche Bewertung erst ermöglichen, sind von Fach zu Fach unterschiedlich. Sie werden von den Fachvertretern, das heißt den Mitgliedern der jeweiligen Bewertungsgruppen, erarbeitet. Hier ergibt sich für

die Fachvertreter ein erheblicher Gestaltungsspielraum, der von den Bewertungsgruppen der ersten Pilotstudie auch entsprechend genutzt wurde. Dies sei an den Kriterien Forschungsqualität und Transfer verdeutlicht: Während in der Chemie zur Bewertung der Forschungsqualität unter anderem eine Zitationsanalyse durchgeführt wurde, war dies in der Soziologie wegen der deutlich heterogeneren, neben Zeitschriftenartikeln insbesondere auch Monografien umfassenden Publikationspraxis nicht möglich. Stattdessen wurden die Wissenschaftler gebeten, ihre besten Publikationen aus dem Erhebungszeitraum einzureichen. Je nach Größe der Forschungseinheit wurde ein mehr oder weniger großer Satz an Publikationen eigener Wahl angefordert. Eingereicht werden konnten Auszüge von Monographien (von nicht mehr als 50 Seiten), Aufsätze in Zeitschriften (auch E-Journals), Beiträge zu Sammelwerken sowie Literaturberichte und Rezensionen. Diese Lektüre ermöglichte ein unmittelbar qualitatives Urteil über die Forschungsleistung einer gegebenen Forschungseinheit. Die umfangreiche Lektüre gehörte in der Pilotstudie Soziologie zu den zentralen Bestandteilen der Bewertungsarbeit und bestimmte in hohem Maße die Bewertung der Forschungsqualität. Bezüglich des Kriteriums „Transfer in andere gesellschaftliche Bereiche“ stand in der Chemie der Transfer in die Industrie im Vordergrund, der sich unter anderem an Patenten, Lizenzen, Spinoffs und Industriemitteln zeigt. In der Soziologie dagegen bezog sich die Bewertung des Transfers vor allem auf die Beratung von Politik und Verbänden im Rahmen beratender Ämter, Projekte und Gutachten.

Auch im Rahmen der Weiterentwicklung des Forschungsratings anhand eines geisteswissenschaftlichen Fachs wird die Definition der Vorgaben für die Lektüre bzw. die Einreichung von Publikationen sowie die Festlegung der weiteren Bewertungsaspekte in der Hand der einzurichtenden Bewertungsgruppe liegen. Über deren Zusammensetzung entscheidet die Steuerungsgruppe in Absprache mit dem Vorsitzenden der Bewertungsgruppe auf der Basis von Vorschlägen, die insbesondere von den jeweiligen Fachgesellschaften und den großen

Wissenschaftsorganisationen erbeten werden. Zur Erweiterung ihrer Kompetenzen kann die Bewertungsgruppe zudem noch im Verfahren Ergänzungen vornehmen und zum Beispiel Sondergutachter für Spezialgebiete bei der Bewertung hinzuziehen.

Die Ergebnisse der Operationalisierung durch die Bewertungsgruppen werden für die zu bewertenden Fächer in einer Tabelle festgehalten, in der den einzelnen Kriterien Bewertungsaspekte und diesen die verwendeten Indikatoren zugeordnet werden. Diese sogenannte „Bewertungsmatrix“ wird jeweils zu Beginn der Datenerhebung veröffentlicht, um transparent zu machen, wie die Daten verwendet werden.

Informed-Peer-Review

Anders als in den verbreiteten Rankings wird die Bewertung von Fachgutachtern vorgenommen, denen für jede zu bewertende Einrichtung umfangreiche Daten vorgelegt werden, und zwar sowohl quantitative als auch qualitative Informationen. Beispielsweise gingen in der Pilotstudie Soziologie in die Bewertung der Nachwuchsförderung als quantitative Informationen die Promotionszahlen, die Anzahl der Promotionsstipendien plus Doktorandenstellen, die Anzahl der Postdocstipendien und Nachwuchsgruppenleiter sowie die Zahl der Rufe an Nachwuchswissenschaftler ein. Als qualitative Informationen kamen hinzu die Selbstbeschreibung von Maßnahmen und Erfolgen der Nachwuchsförderung sowie Auflistungen der abgeschlossenen Dissertationen (mit Verlagsangabe), der strukturierten Promotionsprogramme sowie der Rufe an Nachwuchswissenschaftler. Bei dem Forschungsrating handelt es sich also keineswegs um eine quantitative Studie, sondern um ein kombiniertes Verfahren mit quantitativen und qualitativen Bewertungsaspekten und einer hohen Bedeutung von Peer-Bewertungen.

Chancen und Aussichten

Aufgrund seiner konzeptionellen Merkmale wie „Informed-Peer Review“, fachspezifische Operationalisierung durch Fachgutachter, Mehrdimensionalität der Bewertungen einschließlich der Differenzierung der Forschungsqualität nach Forschungseinheiten bietet das Forschungsrating einen erheblichen Mehrwert gegenüber anderen Verfahren der Qualitätssicherung. Hierzu zählen

- die Möglichkeit der beteiligten Disziplinen, sich aktiv an dem Diskurs über die Qualität von Forschung zu beteiligen, die Bewertungsprozesse aktiv mitzugestalten und bei hochschulpolitischen Entscheidungen nicht wissenschaftsfernen und fachunangemessenen Bewertungen zu unterliegen,
- die Erarbeitung validerer Informationen über wissenschaftliche Leistungen als Drittmittelstatistiken und herkömmliche Rankings und damit deren Relativierung im öffentlichen Diskurs,
- die Stärkung der Selbststeuerungsfähigkeit wissenschaftlicher Einrichtungen durch die Bereitstellung adäquater Kriterien für leistungsorientierte Mittelvergabe.

Gerade die Chance, dass die Wissenschaft im Angesicht administrativ notwendiger Steuerungsverfahren die Hoheit über den Qualitätsdiskurs der Forschung zurückgewinnt, ist für den Wissenschaftsrat ein wesentlicher Grund, die Entwicklung eines solchen Verfahrens zu erproben. Fairerweise muss man dem Forschungsrating allerdings zugestehen, dass sich dieses Verfahren noch in einer Erprobungsphase befindet und nicht jegliche Probleme antizipiert und im Vorfeld ausgeschlossen werden können. Gleichfalls müssen unerwünschte Anzeffekte von Leistungsvergleichen beobachtet und durch eine kontinuierliche Weiterentwicklung der Methode sowie durch begleitende Maßnahmen ausgeglichen werden. Ebenso gilt es sorgsam abzuwägen und zu prüfen, ob der Nutzen der Ergebnisse in einem angemessenen Verhältnis zum Aufwand des Verfahrens steht. In einem ersten Schritt haben daher Mitglieder der Steuerungsgruppe auf einem Workshop

im Dezember 2008 mit Beteiligten aus der Pilotphase und Nutzern der Ratingergebnisse diskutiert, wie die Informationen, die durch das Rating erzeugt werden, bislang genutzt werden und wie sie ggf. besser auf den Bedarf der Adressaten ausgerichtet werden können. Darüber hinaus wurden die Daten der Pilotstudie dem Zentralarchiv der GESIS übermittelt und damit weiteren Nutzern für Sekundäranalysen zur Verfügung gestellt.

Die Entwicklung angemessener Kriterien und Bewertungsverfahren für die Geisteswissenschaften stellt angesichts der Komplexität der beteiligten Fachgebiete sowie des fachspezifischen Diskursstiles und Publikationsverhaltens sicherlich eine besondere Herausforderung dar, die nicht nur erhebliche reflexive Anstrengungen, sondern auch ein großes Maß an gegenseitigem Vertrauen erfordert. Ob dies tatsächlich gelingen kann, gilt es im Rahmen einer erweiterten Pilotphase des Forschungsratings des Wissenschaftsrates zu prüfen. Dies kann selbstverständlich nur in Form eines Annäherungsprozesses vonstatten gehen, in dessen Verlauf unterschiedliche, in anderen Fachgebieten erprobte und bewährte Kriterien auf ihre Anwendbarkeit in den Geisteswissenschaften überprüft sowie neue, spezifische Kriterien entwickelt werden.

~~~~

*Elke Lütke-meier ist Projektkoordinatorin des Forschungsratings in der Geschäftsstelle des Wissenschaftsrats.*