

Vertrauen? Kontrolle?

Gratwanderung: Evaluationen, Rankings und Ratings in der Wissenschaft

Evaluationen werden in der Wissenschaft immer stärker zur Qualitätskontrolle- und Steuerung eingesetzt. In der kritischen Diskussion über verschiedene Bewertungsinstrumente reicht das Spektrum von völliger Ablehnung bis zu differenzierter Zustimmung – im Grundsatz.

Audits, Zertifizierungen, Evaluationen und andere Bewertungs- und Kontrollinstrumente werden zwar häufig kritisch diskutiert, aber nirgendwo zum Teil so vehement abgelehnt wie in der Wissenschaft. Hier stößt vor allem die flächendeckende Einführung von Evaluationen, Rankings und Ratings bei vielen Wissenschaftlern und Wissenschaftlerinnen mit der Begründung auf Ablehnung, diese Instrumente seien dem Gegenstand nicht angemessen und könnten zu Fehlurteilungen und wissenschaftspolitischen Fehlsteuerungen führen. Das jüngste Beispiel aus dem Messinstrumentenkasten ist die „Pilotstudie Forschungsrating“ des Wissenschaftsrats im Fach Chemie, die mit hohem Aufwand betrieben wurde: 42.000 Publikationen und mehr als 300.000 Zitationen wurden erfasst und bibliometrisch ausgewertet und – das ist das Neue am Verfahren – etliche Artikel von der Bewertungsgruppe auch gelesen. Im April sollen die Ergebnisse der Pilotstudie im Fach Soziologie bekanntgegeben werden.

Die Prüfung und Bewertung wissenschaftlicher Leistungen ist so alt wie die organisierte Wissenschaft selbst. Neu ist, dass dieses Geschäft in den letzten zehn Jahren zunehmend als Grundlage für wissenschaftspolitische Entscheidungen genutzt wurde. Dass staatliche Akteure nun eine größere Rolle spielen, kann als Versuch moderner Gesellschaften gewertet werden, die bislang eher informelle und unsystematische Weise der Qualitätssicherung formalisierter und transparenter zu gestalten, und zwar durch die Einführung von systematisierten Indikatoren und Bewertungsverfahren.

In den letzten Jahren hat sich dabei aber gezeigt, dass die an Evaluationen geknüpften Erwartungen, wissenschaftliche Leistungen nach allgemein anerkannten und nachvollziehbaren Kriterien zu bewerten, durchaus problematisch ist. Es ist fraglich, ob wissenschaftliche Qualität sich überhaupt messen lässt. Und schon der Gegenstand der Bewertung ist unscharf: Sollen beispielsweise wissenschaftliche Erkenntnisse im Hinblick auf

ihre wissenschaftliche Güte oder ihre gesellschaftliche Relevanz bewertet werden? Geht es um die effiziente Organisation der Erkenntnisproduktion, also um Fragen der Steuerung von wissenschaftlichen Institutionen und den effizienten Einsatz von Ressourcen?

Ebenso unklar ist oft, wofür Evaluationsergebnisse verwendet werden. Sollen daraus Empfehlungen zur Weiterentwicklung von Forschungsprogrammen generiert werden? Oder soll auf ihrer Grundlage über die Existenzberechtigung oder Förderfähigkeit von Forschungseinrichtungen entschieden werden? Schließlich geht es um die Frage, inwieweit sich die mittlerweile stark ausdifferenzierten wissenschaftlichen Disziplinen überhaupt noch nach einheitlichen Maßstäben beurteilen lassen. Gelten etwa die gleichen Gütekriterien für eine sozialwissenschaftliche Forschungseinrichtung und für ein Institut der Atmosphärenphysik? Und können an ein Max-Planck-Institut gleiche Maßstäbe angelegt werden wie an eine Ressortforschungseinrichtung des Bundes?

Warum nimmt die Evaluation von Wissenschaft trotz der bestehenden Bedenken jetzt und in diesem Ausmaß zu? Wissen ist zu einer entscheidenden Innovationsressource im wirtschaftlichen Wettbewerb der Industriestaaten geworden. Wissenschaft muss daher ihre gesellschaftliche Bedeutung unter Beweis stellen, zumal der Staat weniger zuversichtlich ist, dass die Wissenschaft ihre Aufgaben durch Selbststeuerung gut genug bewältigen kann.

Historisch betrachtet ist dies kein neues Phänomen. Dies zeigen die Analysen von Rüdiger vom Bruch und Stefan Hornbostel: Wenn die Wirtschaft prosperierte, genoss auch die Wissenschaft mehr Autonomie. Bei Rezessionen oder gar Krisen hingegen nahmen die staatlichen Kontrollbedürfnisse gegenüber der Wissenschaft immer zu. Bruno Frey nennt das Bedürfnis, alles und jedes in der Wissenschaft zu evaluieren, „Evaluitis“. Diese zeige, dass der Staat sich wieder einmal unter Handlungsdruck befindet und vor diesem Hintergrund auch die Zügel gegenüber der Wissenschaft straffer ziehen möchte.

In der Diskussion über Evaluationen wird aber auch ein grundlegendes Problem deut-

Summary

Quality control in academia

Evaluations, ratings and rankings are widely used and critically discussed in academia. The debate is highly polarized. While some researchers completely dismiss the use of these instruments, others accept their use in general – but voice reservations concerning its scope.

lich, nämlich die Notwendigkeit eines „ausgewogenen“ Verhältnisses zwischen Vertrauen und Kontrolle. Jede Form von Steuerung wird mit der Eigendynamik und Eigenwilligkeit gesellschaftlicher Akteure konfrontiert, die der Regulierung und Kontrolle quasi natürliche Grenzen setzen. Steuerungsprozesse, wo auch immer sie stattfinden, benötigen insofern ein Mindestmaß an Vertrauen, an gemeinsamen Identifikationen und an geteilten Werten und Zielen. Dies gilt umso mehr in der Wissenschaft, die sich traditionell eher über Diskurs- und Konsensbildungsprozesse verständigt. Sobald die wissenschaftspolitischen Akteure ihr Kontrollbedürfnis gegenüber der Wissenschaft ausdehnen wollen, müssen sie ein Steuerungsinstrument finden, welches das Verhältnis von Vertrauen und Kontrolle in der erforderlichen Balance hält.

Evaluationen sind vor diesem Hintergrund deshalb interessant, weil sie keine direkten Eingriffe in die Wissenschaft vornehmen, sondern im Wesentlichen Informationen als Entscheidungsgrundlage für mögliche politische Korrekturen der Rahmenbedingungen liefern, etwa bei der Vergabe von finanziellen Zuwendungen. Das eigentliche Kontroll- und Bewertungsverfahren verbleibt hingegen in der Wissenschaft. Dies geschieht durch das Peer Review, die Beurteilung durch die Fachkollegen.

Vor Kritik ist dieses Instrument jedoch nicht gefeit. Man denke nur an die zahlreichen kritischen Studien zur Zuverlässigkeit des Peer Review oder an die Kritik, nur der wissenschaftliche Mainstream werde befördert. Dennoch hat sich das Peer Review seit langem als Verfahren zur internen Qualitätsbeurteilung von Wissenschaft, etwa von Zeitschriftenartikeln, bewährt. Es genießt großes Vertrauen und dementsprechend hohe Anerkennung und weltweite Verbreitung.

So nimmt es nicht Wunder, dass auch die Wissenschaftspolitik in den von ihr initiierten Evaluationen auf die Bewertungsmechanismen der *scientific communities* setzt, also auf das Peer Review. Eine Alternative ist bislang nicht in Sicht. Parallel eingeführte Bewertungsverfahren wie Rankings und Ratings beruhen letztlich ebenfalls auf dem Peer Review; als Bewertungsindikatoren gelten dabei nämlich wissenschaftliche Produkte, die ein Peer Review erfolgreich überstanden haben, wie zum Beispiel Zeitschriftenartikel.

Am Peer Review wird häufig kritisiert, dass seine Beurteilungskriterien intransparent sind, nicht zuletzt deshalb, weil bei der Be-

wertung wissenschaftlicher Güte in hohem Maße implizit bleibende Kriterien angewandt werden. Das „Implizite“ macht aber gerade die Prominenz des Peer Review aus: Es ist das „Geschäft von Personen, denen man nichts vormacht, da sie über intime Kenntnisse der Arbeiten, der Forschungslinien und der Reputation verfügen“, wie Robert Salais es formuliert. Für die evaluierten Einrichtungen verhindert jedoch nicht zuletzt diese Intransparenz, dass sie aus den Evaluationsergebnissen lernen könnten.

Um Evaluationen nachhaltiger und effektiver für die Forschungseinrichtungen zu gestalten, fordern Michael Power und Helmut Wiesen-thal deshalb, sie stärker mit internen Verfahren der Qualitätsentwicklung zu verbinden. Damit wäre das Geschäft nicht allein den Peers überlassen. Den Evaluierten würde vielmehr ein aktiver Part zugebilligt in dem Sinne, dass die von ihnen selbst gesetzten Kriterien zu ihrer Beurteilung ausreichend Beachtung finden.

Die Wissenschaft muss sich gegenüber der Gesellschaft legitimieren und um Anerkennung bemühen. Deshalb sollte auch die Außenwelt stärker in die Verfahren einbezogen werden. Das wäre besonders dann wichtig, wenn es um die Bewertung der Anwendbarkeit von wissenschaftlichem Wissen geht. Wim Blockmans bezieht sich hierbei auf ein radarförmiges Modell, das für die Agrar- und pharmazeutischen Wissenschaften entwickelt und angewandt wurde. Dabei geht es um die zielgruppenspezifische Bewertung folgender Leistungen: science and certified knowledge; education and training; innovation and professionals; public policy and societal issues; collaboration and visibility.

Dieser Vorschlag ließe sich dahingehend weiterentwickeln, dass die Peers in Evaluationsverfahren im Wesentlichen für die Bewertung der wissenschaftlichen Qualität zuständig wären, während die gesellschaftliche Relevanz der Forschungsleistung durch Akteure aus Politik, Verwaltung und anderen Bereichen beurteilt werden würde. Erst die Zusammenschau der unterschiedlichen Perspektiven und Einschätzungen ergäbe ein vollständiges Bild der Aufgaben und Leistungen des Forschungsinstituts.

So könnte möglicherweise auch das Verhältnis von Vertrauen und Kontrolle bei der Bewertung wissenschaftlicher Leistungen und



Hildegard Matthies (links) und Dagmar Simon [Foto: David Ausserhofer]

Hildegard Matthies, geboren 1952 in Soltau (Niedersachsen), studierte Soziologie und Sozialökonomie in Hamburg. 1990 war sie Fellow am Minda de Gunzburg Center for European Studies an der Harvard University. 1999 wurde sie promoviert, seit 1998 ist sie wissenschaftliche Mitarbeiterin am WZB, heute in der Forschungsgruppe „Wissenschaftspolitik“.

matthies@wzb.eu

Dagmar Simon, geboren 1954 in Frankfurt a.M., studierte Politikwissenschaften und Germanistik in Frankfurt a.M. und Berlin, wo sie auch promoviert wurde. Von 1994 war sie Referentin für Forschungsplanung und –koordination am WZB, seit Januar 2008 leitet sie kommissarisch die neue WZB-Forschungsgruppe „Wissenschaftspolitik“, die aus der Projektgruppe „Wissenschaftspolitik“ hervorgegangen ist.

dsimon@wzb.eu

ihrer gesellschaftlichen Bedeutung neu justiert werden. Die Politik wäre zwar stärker in das Verfahren einbezogen, aber die Bewertung der wissenschaftlichen Leistungsfähigkeit bliebe in den Händen der Fachgemeinschaften. Die Forschungseinrichtungen selber wären ein aktiver Part des Evaluationsprozesses.

Hildegard Matthies, Dagmar Simon
Forschungsgruppe „Wissenschaftspolitik“

Literatur:

Hildegard Matthies, Dagmar Simon (Hg.): Wissenschaft unter Beobachtung – Effekte und Defekte von Evaluationen. Leviathan Sonderheft 24/2007, Wiesbaden: VS Verlag für Sozialwissenschaften 2008