

EVALUATION

Bewertung und Rechtfertigung von Wissenschaft in praktischer Bewährung

Umstritten sind Zeitpunkt, Methode und Gegenstand der Leistungsmessung nach wie vor



Messen, vergleichen und urteilen: Evaluationsverfahren in der Wissenschaft müssen Vertrauen gewinnen.

Foto: Universität Heidelberg

Am wissenschaftspolitischen Stichwort **Evaluation** scheiden sich gegenwärtig die Geister. Dabei ist unstrittig, dass sich eine öffentlich finanzierte Wissenschaft auch in einem vertretbaren Umfang hinsichtlich ihrer Leistungsfähigkeit und Verwendung der in sie investierten öffentlichen Gelder rechtfertigen muss. Kürzlich brachte ein Expertengespräch über „Wissenschaftsevaluation in praktischer Bewährung“ diese kontroverse Debatte auf den Punkt. Das Zentrum für Wissenschaftsmanagement hatte mit Unterstützung des Stifterverbandes für die Deutsche Wissenschaft hierzu nach Bonn eingeladen.

Zu welchen Zeitpunkten eine Rechtfertigung im Wissenschaftsprozess stattfinden soll, wie und welche Leistungen überhaupt zu rechtfertigen sind, von wem, für wen und in welchem Umfang – das bleibt weithin umstritten. Insbesondere gilt dies, wenn – wie in den meisten Fällen – mit solchen Leistungserhebungen und -bewertungen auch Entscheidungen über Mittelvergaben verknüpft sind. In Deutschland haben sich die Forschungsevaluation und an sie geknüpfte Anreiz- und Mittelverteilungssysteme seit Ende der 1990er Jahre in allen Wissenschaftsorganisationen etabliert.

Unter Evaluation lassen sich ganz allgemein die Bewertung und Prüfung von Sachverhalten, Leistungen und Konformitäten in den unterschiedlichsten Leistungszusammenhängen und Produktionsbereichen fassen. Für Stefan Hornbostel, Direktor des Instituts für Forschungsinformation und Qualitätssicherung (IFQ), Bonn, gibt es zahlreiche ideengeschichtliche Vorläufer, mit denen sich das Wissenschaftsmanagement auseinandersetzen sollte. Ein aktuelles Äquivalent von

Evaluation ist z.B. das Rating-Verfahren im Wirtschaftssystem, in dem privatwirtschaftliche Agenturen die finanzielle Lage und die Wachstumsperspektiven beispielsweise von börsennotierten Unternehmen beurteilen. Dieses Verfahren hat große Ähnlichkeiten mit Evaluationen: Üblicherweise werden Bilanzen durchleuchtet, Geschäftsrisiken geprüft, „Begehungen“ im Sinne von Interviews mit den Vorständen organisiert und ein Benchmarking mit den Konkurrenten des zu bewertenden Unternehmens durchgeführt. In einem Rating-Komitee wird dann der Projektleiter des jeweiligen Ratings einen Bewertungsvorschlag zur Diskussion und Abstimmung stellen. Da die zu diskutierenden Kennzahlen interpretationsbedürftig sind und viele Informationen auf einem den vertraulichen Interviews entnommenem Geheimwissen beruhen, ist bei der Bewertung eine Menge ‚tacit knowledge‘ im Spiel, das für eine gewisse – für notwendig erachtete – Intransparenz des Verfahrens sorgt.

Der Evaluationsaufwand im Wissenschaftssystem folgt einer Grenznutzenfunktion. IFQ-Leiter Hornbostel mahnt daher, Infrastrukturen aufzubauen, in denen die erhobenen Daten gepoolt und damit einer Mehrfachnutzung zugeführt werden können. Auch solle ein höheres Maß an Standardisierung bei den Evaluationsverfahren selbst angestrebt werden. Das änderte aber letztlich nichts an der Tatsache, dass Evaluationen wissenschaftspolitische Entscheidungen nicht ersetzen können.

Günter Hirsch, Präsident des Bundesgerichtshofs (BGH) a.D., zeigt mit einem Vergleich von richterlichen Urteilen und Entscheidungen im Evaluationsprozess Ähnlichkeiten und Differenzen beider Verfahren auf. Im Gegensatz zum Gerichtsentscheid, der nach den Prozess-

ordnungen strikten normativen Vorgaben bei der Wahl der Erhebungsinstrumente für den betreffenden Sachverhalt folgen muss, ist die Evaluation der Wissenschaft relativ frei in der Wahl der Instrumente, die zur Feststellung des jeweiligen Ist-Zustands der zu bewertenden Einrichtung anzuwenden sind. An die Erhebung des Sachverhalts schließt sich in beiden Fällen der Prozess der fachlichen Bewertung an.

In dieser Phase erfolgt im Rahmen der richterlichen Entscheidungsfindung eine Subsumtion der Fakten unter die Rechtsnorm; in der Wissenschaftsevaluation wird äquivalent die Anwendung wissenschaftlicher Kriterien zur Feststellung der Stärken und Schwächen einer wissenschaftlichen Einrichtung, also – gemäß Hirsch – „ein Qualitätsmanagement“ vorgenommen. Hier liegt die gemeinsame Schwierigkeit beider Verfahren: Sowohl die rechtliche Norm zur Regulierung des Lebenssachverhalts als auch die wissenschaftliche Leistung sind auslegungsbedürftig. Die Auslegungs- und Methodenfrage spielt dabei eine entscheidende Rolle hinsichtlich der Gewährleistung der Richtigkeit der Entscheidung. Schließlich – und dies ist wiederum ein von Hirsch hervorgehobener Unterschied zwischen beiden Praxen der Entscheidungsfindung – ergibt sich bei Evaluationen die Definition der Kriterien, an denen der Erfolg einer wissenschaftlichen Einrichtung gemessen wird, nicht aus einer normativen Vorgabe, sondern aus der Aufgabenstellung der einzelnen Einrichtung.

Dorothee Dzwonnek, Generalsekretärin der Deutschen Forschungsgemeinschaft (DFG), lenkt den Blick auf die Auslöser der Evaluationswelle im deutschen Wissenschaftssystem. Öffentliche Geldgeber und Steuerzahler fordern seit Ende der 1980er Jahre ein höheres Maß an Transparenz über die Verwendung der Mittel in den Wissenschaftseinrichtungen. Auch die DFG musste sich seit der großen Systemevaluation der deutschen Wissenschaftsorganisationen Ende der 1990er Jahre im Hinblick auf ihre Förderinstrumente und ihre internen Strukturen neu organisieren. Die

Durchführung der Exzellenzinitiative von Bund und Ländern (ExIn) und die Delegation dieser Aufgabe an die DFG und den Wissenschaftsrat (WR) ist ein weiteres aktuelles Beispiel, wie ein Leistungsmessen im Rahmen eines wettbewerblichen Verfahrens angelegt werden kann. Hiermit fiel der DFG und dem WR ab 2005 auch die Aufgabe zu, die staatlicherseits initiierte Konkurrenz deutscher Universitäten um die Finanzierung exzellenter Forschungsleistungen und innovativer Infrastrukturkonzepte in drei Förderlinien – und hierbei insbesondere den Peer-Review zur Begutachtung der eingereichten Anträge – zu organisieren.

Das ganze Verfahren hat Dzwonnek zufolge erheblich zum Wettbewerb und zum Aufbrechen der Versäulung im deutschen Wissenschaftssystem beigetragen. Gleichwohl sei zu fragen, ob die politisch gewünschten kurzen Zeitabstände solcher Evaluationen nach Auflage neuer Förderinstrumente – im Falle der ExIn kaum drei Jahre – dem Zeitraum entsprechen, ab dem man ein Forschungsprojekt von der Ergebnisseite her wirklich realistisch in seiner Qualität beurteilen könne. Im Falle der ExIn sei hierfür ein Zeitraum von fünf bis sieben Jahren anzunehmen.

Fazit der Debatte: Ein mit Evaluationen befasstes Wissenschaftsmanagement sollte durchaus die ideengeschichtlichen Fundamente verschiedener Bewertungs- und Rechtfertigungspraktiken sichten und dabei auch über den Tellerrand des eigenen Nahbereichs hinausschauen. Ähnlich gelagerte Bewertungs- und Entscheidungspraktiken in anderen Systemen sind genauso wie aktuelle Verfahren und Folgen der Wissenschaftsevaluation zu reflektieren. Jenseits der einfachen Kennzahlen und Output-Größen liegt noch eine tiefere Dimension der Vermessung der Wissenschaft, die es zu verstehen gilt, sofern man aus historischen Fehlern, vergleichbaren Problemen in anderen Bereichen und konkreten Umsetzungsschwierigkeiten im heutigen Wissenschaftsalltag lernen möchte.

Stefan Lange

Ein modernes Wissenschaftsmanagement muss mit Blick auf die Betroffenen der Evaluationen erheblich in vertrauensbildende Maßnahmen investieren. Dafür sind auch profunde Kenntnisse über Bewertungs- und Rechtfertigungspraktiken außerhalb der Wissenschaft unabdingbar.

Dr. Stefan Lange, M.A., Stiftungslehrstuhl für Wissenschaftsorganisation, Hochschul- und Wissenschaftsmanagement an der Deutschen Hochschule für Verwaltungswissenschaften Speyer, geschäftsführender Herausgeber der Zeitschrift „Hochschulmanagement“.