

Wissenschaftsrat

# Jenseits von Ritual und Standardprozedur: Zur „guten Praxis“ institutioneller Evaluationen im Wissen- schaftssystem

Evaluationen in der Wissenschaft sind seit einigen Jahren massiver Kritik ausgesetzt – anders, als noch in den 1990er Jahren. Was muss man berücksichtigen, damit sie trotzdem erfolgreich ablaufen können? Der Wissenschaftsrat hat dazu Empfehlungen vorgelegt.

VON ANDREAS STUCKE

ICH VERMUTE, ICH BIN auch deshalb um einen Beitrag zum aktuellen Themenheft der Bayerischen Akademie der Wissenschaften gebeten worden, weil der Wissenschaftsrat als Akteur mit ausreichend Erfahrungswissen zum Thema wahrgenommen wird. Und in der Tat: Der Wissenschaftsrat hat schon zu Zeiten institutionell begutachtet, als der Begriff „Evaluation“ noch nicht gängig war. Er wurde später im Zuge der Vereinigung der beiden deutschen Wissenschaftssysteme zu der zentralen, öffentlich auch so apostrophierten „Evaluationsinstanz“ und hat danach bis in die Gegenwart großflächig ganze Teile der institutionell verfassten Wissenschaft, etwa die Einrichtungen der „Blauen Liste“ oder der Ressortforschung, begutachtet bzw. zu einzelnen Wissenschaftsorganisationen, wie der Helmholtz-Gemeinschaft, „Systemevaluationen“ durchgeführt.

## Evaluation in der Wissenschaft: Königsweg der Qualitätskontrolle oder Heilige Inquisition?

Evaluation hat ungebrochen Konjunktur im Wissenschaftssystem, sie wird zunehmend nachgefragt und angeboten – auch vom Wissenschaftsrat. Radikal verändert hat sich demgegenüber der öffentliche Diskurs über Evaluation: Hier hat sich binnen weniger Jahre ein diametraler Einstellungswandel vollzogen. Noch Ende der

*My old mother always used to say (...),  
that facts are like cows.*

*If you stare them in the face hard enough,  
they generally run away.*

(DOROTHY SAYERS)

1990er Jahre, nach der Evaluation sämtlicher Bund-Länder-finanzierten Einrichtungen der „Blauen Liste“ durch den Wissenschaftsrat, wurde, auch mit Verweis auf Entwicklungen in Großbritannien und den Niederlanden, in regelmäßigen Evaluationen der Königsweg gesehen, um verkrustete Strukturen im Wissenschaftssystem aufzubrechen. Die in Teilen schwache internationale Konkurrenzfähigkeit des deutschen Wissenschaftssystems – so die breit geteilte Wahrnehmung damals – sei auch Ergebnis einer zu großzügigen staatlichen Grundfinanzierung ohne ausreichende Qualitätskontrolle. Ausgerufen wurde deshalb eine „Kultur der Evaluation“ in Deutschland. Keine zehn Jahre später hat sich die öffentliche Bewertung praktisch umgekehrt. Institutionelle Evaluationen finden zwar weiter statt, wahrscheinlich sogar mit steigender Intensität, aber nun mit deutlich schlechterem Gewissen aller Beteiligten. In den öffentlichen Debatten steht die Befürchtung im Vordergrund, Evaluation könne wissenschaftliche Innovation nicht fördern, sondern würde sie geradezu verhindern, weil sie Forscher und Einrichtungen von deren eigentlicher Aufgabe, Forschung zu betreiben, ablenke, notwendig den wissenschaftlichen Mainstream und nicht Originalität belohne und in der Summe bestenfalls ein Ritual von Qualitätssicherung und damit Symbolpolitik sei, schlechtestenfalls aber nur bereits beschlossene Budgetkürzungen legitimieren solle. Mitunter werden Evaluationen sogar in die Nähe von peinlichen Befragungen durch die Heilige Inquisition gerückt. Zwei Welten – oder Ausdruck notwendiger Zielspannungen zwischen wissenschaftlicher Selbststeuerung und externer Kontrolle?

## Gegenstand wissenschaftspolitischer Kontroversen

Es wird niemanden verwundern, wenn ein Thema wie Evaluation, mit dem auch das Verhältnis zwischen Wissenschaft und Politik neu bestimmt wird, früher oder später Gegenstand wissenschaftspolitischer Kontroversen wird.



Insbesondere die Balance zwischen Vertrauen in die Wissenschaft und öffentlicher Rechenschaftslegung der Wissenschaft für wachsende staatliche Alimentierung ist störungsanfällig und muss stets neu hergestellt werden. Das zeigt sich auch in den Aushandlungsprozessen um das Thema Evaluation. Immer dann, wenn viel auf dem Spiel steht und Evaluationen tatsächlich Folgen zeitigen, treten Akteure mit ihren jeweiligen Interessen auf die Bühne. Diese Interessenlagen sind vielfältig, etwa zwischen Politik, die „accountability“ einklagt, und Wissenschaft, die Vertrauen und langfristige Absicherung will, zwischen Einrichtungsleitungen, die steuern wollen, und einzelnen Wissenschaftlern, die in Ruhe forschen wollen, zwischen Natur- und Ingenieurwissenschaftlern, die oftmals in kooperativen Forschungsverbänden mit großen Forschungsinfrastrukturen arbeiten, und Teilen der Geistes- und Sozialwissenschaften, die den Wert der „Individualforschung“ betonen, und – nicht zuletzt – zwischen eher disziplinär und eher inter- oder transdisziplinär arbeitenden Wissenschaftlern. Letztere beklagen seit längerem, dass mit ihnen in Evaluationen oft nicht fair umgegangen werde. Unterfüttert sind die Positionen mit vielen guten Argumenten, etwa von Geisteswissenschaftlern, die darauf hinweisen, dass weder die Zahl von Publikationen in peer-reviewed Journalen noch die Höhe von Drittmittelinwerbungen per se ein gültiger Indikator für Qualität oder gar Kreativität ist. Umgekehrt wird nicht überraschen, dass zum Beispiel Vertreter der geräteintensiven teuren Natur- und Ingenieurwissenschaften gerade diese Leistungsindikatoren hochhalten, da sie helfen, dringend benötigte Zuwächse zu begründen.

### **Dringend nötig: eine Reflexion der Evaluationspraxis**

Derartige Debatten sind als „Begleitmusik“ zu vielfältigen Evaluationsaktivitäten nicht nur unvermeidlich, sie sind geradezu notwendig. Entsprechend hat sich der Wissenschaftsrat des Themas reflexiv angenommen und die Evaluationspraxen in Deutschland, auch die eigenen, einer kritischen Untersuchung unterzogen. Am Ende standen Empfehlungen, die zwar auch zu einzelnen Methoden und Indikatoren Stellung beziehen – wie zur Bibliometrie, zum Umgang mit Publikations-, Zitations- oder Drittmittelzahlen –, deren gemeinsamer Nenner und Stoßrichtung aber gerade kein technizistisch-methodischer ist. Es ging dem Wissenschaftsrat vielmehr darum, deutlich zu machen, dass jede Evaluation in einem sozialen

Kontext steht, in dem unterschiedliche Akteure mit ganz verschiedenen legitimen Interessen beteiligt sind und es ganz wesentlich darauf ankommt, den Prozess der Evaluation so zu gestalten, dass der „Fall“, um den es geht, eine bestmögliche Herangehensweise erfährt. Es geht also um das, was man die „soziale Pragmatik“ von Evaluationen nennen könnte und was alle Prozessschritte von der Einleitung eines Evaluationsverfahrens über die Durchführung bis zum Abschluss sowie zur Umsetzung der Ergebnisse umfasst.

Evaluationskriterien und Indikatoren sind in einem solchen Verfahren Orientierungsmarken, also Hilfsmittel, die Fragen anregen und zur Strukturierung des Begutachtungsprozesses beitragen können, aber keine kontextlosen Messgrößen, die unabhängig von den Voraussetzungen und Bedingungen ihrer Anwendung wissenschaftspolitische Bedeutung haben (sollten). Die Zahl von Publikationen, Zitaten, Promotionen sowie Patenten oder die Höhe der Drittmittel von Instituten sagt für sich ebenso wenig über die relative Leistung oder Leistungsfähigkeit einer Einrichtung aus wie über die wissenschaftspolitischen Schlussfolgerungen, die daraus zu ziehen sind. So müssen Publikations- und Zitationswerte für eine seriöse qualitative Einordnung immer auf eine Bezugsgröße – die Publikationsaktivitäten im Forschungsfeld bzw. den Output anderer vergleichbarer Einrichtungen – bezogen und vor dem Hintergrund der institutionellen Rahmenbedingungen, unter denen publiziert wird, bewertet werden. Hier wird man eine Ressortforschungseinrichtung, die neben Forschung vor allem Beratung und Dienstleistungen für die Politik zu leisten hat, anders bewerten müssen als etwa ein Max-Planck-Institut. Festgestellte unterdurchschnittliche Leistungen einer Einrichtung können zudem nicht automatisch zu eindimensionalen Schlussfolgerungen, etwa zu Ressourcenentzug oder Schließung, führen. Das Gegenteil mag richtig sein, wenn die Leistungsschwäche zum Beispiel an mangelnder apparativer Ausstattung oder an einer Überfrachtung einer Einrichtung mit zu vielen externen Aufträgen durch den Zuwendungsgeber liegt. In diesem Fall kann ein niedriger Leistungsstand sogar der Anlass sein, für die Einrichtung eine bessere Ausstattung oder dauerhaft mehr Geld und Zeit für eigene Forschung zu empfehlen.

## Empfehlungen des Wissenschaftsrates

Daran wird deutlich: Der Vorgang der Bewertung ist im Kern ein interaktiver, in dem verschiedene Akteure – mindestens die zu begutachtende Einrichtung, die Bewertungsgruppe und der Auftraggeber – ihre Sichtweisen darüber austauschen, wie die Leistungen einer Einrichtung zu bewerten sind, welche Gründe dafür verantwortlich gemacht und wie mögliche Defizite behoben werden können. Je stärker dieser Tatbestand Teil der mitlaufenden Reflexion aller Beteiligten ist, desto besser sind die Voraussetzungen, die Qualität des Evaluationsprozesses zu steigern. Damit ist im Grunde auch die Prämisse beschrieben, die den Wissenschaftsrat in seinen „Empfehlungen zur Bewertung und Steuerung von Forschungsleistung“ 2011 veranlasst hat, jenseits von technizistischen Hinweisen zu Indikatorik und Methoden eine Reihe von übergeordneten Handlungsempfehlungen für die Durchführung von Evaluationen zu geben. Sie unterstreichen, dass institutionelle Evaluationen nicht als triviale Standardprozedur ablaufen dürfen, in die lediglich Zahlen und Daten eingespeist und vorgestanzte Bewertungen ausgeworfen werden, sondern in ihrem Gelingen davon abhängen, dass kontextsensitiv, d. h. unter größtmöglicher Berücksichtigung der Voraussetzungen und Besonderheiten des einzelnen Falls sowie bisheriger Erfahrungen, Bewertungen vorgenommen, Begründungen ausgetauscht, Befunde erklärt und Ratschläge ausgesprochen werden können. Der Wissenschaftsrat ist der Auffassung:

- Institutionelle Evaluationen sind umso valider und wirksamer, je breiter sie auf Berichtssysteme und Monitoring-Aktivitäten der Einrichtungen selbst aufsetzen können. Wissenschaftliche Einrichtungen sollten deshalb über *institutionelle Formen der Selbstbeobachtung* verfügen, um nach innen strategie- und entscheidungsfähig und nach außen argumentationsfähig zu sein.
- Bewertungsverfahren und Steuerungsinstrumente sollten regelmäßig auf ihre *erwünschten wie unerwünschten Wirkungen überprüft* werden, sie dürfen nicht zu „schlechter Routine“ oder zum Selbstzweck werden.
- Evaluationen verursachen Aufwand und Kosten, vor allem bei Gutachtern und den zu evaluierenden Einrichtungen. Deshalb sollte bei aller notwendigen Komplexität einer Evaluation die *Aufwandsbegrenzung* ein vorrangiges Ziel sein.
- Vielfach hat sich eine regelmäßige Rechen-schaftslegung durch externe Evaluationen etabliert. Diese sollte, wenn nicht besondere Anlässe gegeben sind, einer *angemessenen Taktung* folgen, d. h. im Regelfall eher großzügig bemessen sein (5–10 Jahre).

- Bei der Bewertung von Forschungsleistungen sollte jegliche „Tonnenideologie“ vermieden werden. „Mehr“ ist nicht gleich „besser“. Das Verfahren sollte deshalb so angelegt sein, dass vor allem die *Qualität* der Forschungsleistungen erfasst wird.
- Die verwendeten Bewertungsmaßstäbe müssen „*adäquat*“ sein, d. h. der jeweiligen Mission bzw. dem Aufgabenprofil einer Einrichtung folgen.
- In die (Weiter-) Entwicklung von Evaluationsverfahren sollten *alle relevanten Akteure*, insbesondere aber die jeweiligen Fachgemeinschaften *einbezogen werden*, um auf diese Weise disziplinäre bzw. forschungsfeldbezogene Besonderheiten zu erfassen.
- Eine *ausreichende Grundfinanzierung* von Hochschulen und außeruniversitären Forschungseinrichtungen, um „risikoreiche“, langfristige Forschung abzusichern, muss die Basis für steuernde Eingriffe durch Evaluationen sein. Evaluationen sind kein geeignetes Instrument zur Redistribution von Mitteln bei fehlender Grundfinanzierung.

Gemeinsames Merkmal dieser Empfehlungen ist, dass ihre Umsetzung nicht top-down verordnet werden kann. Es ist auch noch nicht viel geholfen, wenn Grundsätze dieser Art die Leitfäden für gute Evaluationen zieren oder einfach nur bürokratisch „abgearbeitet werden“. Die Empfehlungen stellen vielmehr eine Aufforderung an alle Beteiligten dar, gerade bei konkret anstehenden „Evaluationsfällen“ in einen Diskurs über die Voraussetzungen des Gelingens der Evaluationen zu treten. Weil Evaluationen nie im machtfreien Raum stattfinden, kommt es darauf an, Begründungslasten zu erzeugen bzw. zu verschieben und allen Akteuren Argumente zur Verfügung zu stellen, um, notfalls auch streitig, über Auftrag und Zielsetzungen, Methoden und Bewertungsmaßstäbe, Diagnosen und Therapien zu reden. Wie gut der Wissenschaftsrat in dieser Hinsicht seine eigenen Empfehlungen beherzigt, müssen andere beurteilen; der Anspruch ist jedenfalls gesetzt und darf eingefordert werden.

### DER AUTOR

*Dr. Andreas Stucke leitet die Abteilung Evaluation im Wissenschaftsrat. Der Wissenschaftsrat ist eines der wichtigsten wissenschaftspolitischen Beratungsgremien in Deutschland. Er berät die Bundesregierung und die Regierungen der Länder in Fragen der inhaltlichen und strukturellen Entwicklung der Hochschulen, der Wissenschaft und der Forschung.*

### Literatur und WWW

Wissenschaftsrat, Empfehlungen zur Bewertung und Steuerung von Forschungsleistung, November 2011. Download: [www.wissenschaftsrat.de/download/archiv/1656-11.pdf](http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf)