

Die Krankheit der Wissenschaft

Der Forschungsbetrieb leidet am Übermaß falsch ausgerichteter Evaluationen / Von Margit Osterloh und Bruno S. Frey

Eine Krankheit hat die Wissenschaft befallen: die Evaluitis. Heute werden in immer kürzeren Abständen ganze Universitäten, Fakultäten, Fachbereiche, Institute, Forschungsgruppen und einzelne Forschende begutachtet. Evaluationen und daraus abgeleitete Ranglisten sind im Wissenschaftsbetrieb allgegenwärtig. Unter „Evaluation“ wird hier die nachträgliche Einschätzung der Leistung einer Organisation oder Person durch von außen kommende Experten verstanden. Evaluationen haben verborgene und damit gewöhnlich vernachlässigte Kosten. Ihr Nutzen wird überbewertet. Insoweit diese Effekte nicht berücksichtigt werden, wird der Nettotonnenwert dieses Instruments systematisch überschätzt. In diesem Falle werden mehr Evaluationen durchgeführt, als gesellschaftlich sinnvoll wäre.

Evaluationen sind zwar in manchen Fällen notwendig, aber sie verbessern nicht immer das Wissenschaftssystem. Gegenwärtig auftretende Schwierigkeiten können nicht einfach durch noch sorgfältigere Evaluationen beseitigt werden, sondern können sogar einige Probleme noch verschlimmern. Denn Evaluationen verzerrten Anreize. Sie verändern das Verhalten der davon betroffenen Personen in systematischer und unbeabsichtigter Weise, unabhängig davon, wie sorgfältig sie durchgeführt werden.

Der Nutzen der Bewertung wird überschätzt

Für qualifizierte Tätigkeit lassen sich nicht alle relevanten Aspekte im Vorhinein definieren und im Nachhinein messen. Eine Evaluation nach vordefinierten Kriterien veranlasst die Bewerteten, sich vorwiegend nach diesen Kriterien zu richten. Häufig legen Evaluationen die Anzahl der Publikationen als Maßstab zugrunde. In diesem Falle werden die Forschenden neue Ideen oder interessante Datensätze so dünn wie Salamischeiben aufschneiden und zu möglichst vielen – dann aber gehaltenen – Publikationen verarbeiten. Eindeutige empirische Evidenz hierfür gibt es für Australien. Mitte der neunziger Jahre wurden dort die Bezahlung der Wissenschaftler und die Finanzierung der Universitäten an die Zahl der Veröffentlichungen in referierten Zeitschriften gekoppelt. Wie zu erwarten war, ist die Zahl der Veröffentlichungen dramatisch gestiegen – aber die Qualität (gemessen anhand der Zahl der Zitierungen) entsprechend gesunken. Sie fiel unter den Durchschnitt der übrigen OECD-Länder.

Die Publikationszahl lässt sich auch steigern, indem an der Forschung unbeteiligte Wissenschaftler als Autoren mit aufgenommen werden, vorausgesetzt, sie revanchieren sich. Die Publikationsflut bewirkt darüber hinaus, dass ein Heer von Gutachtern aufgeboden werden muss, um die Publikationen zu beurteilen. Die Belastung führt dazu, dass immer oberflächlichere Gutachten geschrieben werden oder dass die Gutachten von Assistenten verfasst werden. Auch die Messung von Forschungsleistungen mittels Zitierungen führt zu Verzerrungen. Es werden Zitationskartelle gebildet, und der Anreiz wächst, sich modischen Themen mit hohem Aufmerksamkeitswert zu widmen. Hingegen wird die Übertragung wissenschaftlicher Erkenntnisse in die Praxis oder in andere Disziplinen vernachlässigt, weil Publikationen in allgemeinverständlichen Büchern, Vorträgen vor einem Nichtfachpublikum sowie Beratungs- und Schulungstätigkeiten keine Zitate in wissenschaftlichen Veröffentlichungen bringen.

Wird schließlich die Anzahl der betreuten Doktoranden als Leistungsmaß verwendet, werden die Anforderungen gesenkt. Als ein Ausweg zur Leistungsmessung häufig die Anwerbung von Drittmitteln. Damit wird allerdings weder der Sinn noch die Produktivität der damit finanzierten Forschung erfasst. Dennoch ist dieses Kriterium beliebt, weil Geldströme besonders leicht messbar sind. Wird eine wissenschaftliche Institution danach beurteilt, ist sie gezwungen, sich um Drittmittel zu bemühen und gleichzeitig weniger gut messbare Forschungs- und Lehrleistungen zu vernachlässigen. Dieses Kriterium ist aber erstens für zahlreiche Fächer unsinnig, insbesondere im Bereich der Geisteswissenschaften. Zweitens entsteht, wenn die Höhe der Drittmittel zählt, ein Anreiz zur Beantragung überhöhter Forschungsmittel und zu ineffizienter Forschung.

Die Beispiele ließen sich beliebig erweitern. Das Ergebnis sind immer aufwendigere Evaluationsprozesse, die einem Wettlauf von Hase und Igel gleichen. Es entsteht ein „Rattenrennen“, das die Wissenschaft nicht verbessert, sondern nur zu höheren Kosten führt. Der bürokratische Aufwand für Evaluationen und die sich etablierende Evaluationsindustrie ist heute schon gewaltig und wird durch eine leistungsorientierte Bezahlung nochmals gesteigert. Darüber hinaus lassen sie den Forschenden immer weniger Zeit für ihre eigentlichen Aufgaben. Sie werden gezwungen, permanent entweder zu evaluieren oder evaluiert zu werden.

Hinzu kommt, dass gerade bahnbrechende Forschung oft gegen den Konsens des wissenschaftlichen Hauptstroms verstößt, deshalb in einer Evaluation zunächst schlecht beurteilt wird und sich erst nach Jahrzehnten durchsetzt. Gute oder gar revolutionäre wissenschaftliche Forschung zeichnet sich dadurch aus, dass sie neue Maßstäbe setzt und sich oft erst

Ob staatlich oder privat finanziert, um an knappe Forschungsgelder zu kommen, müssen sich Universitäten immer mehr und immer aufwendigeren Bewertungen von außen unterziehen. Doch taugen die von Gutachtern erstellten Ranglisten

kaum, um das Geld dorthin zu lenken, wo es den größten wissenschaftlichen Nutzen stiftet. Albert Einstein und Max Planck in den Naturwissenschaften, John Maynard Keynes und John Hicks in der Ökonomie wären im System permanen-

ter Evaluationen nicht sehr erfolgreich gewesen, warnen die Autoren dieses Beitrags. Spitzenforschung brauche Zeit, um zu bewertbaren Ergebnissen zu führen. Dauerkontrollen begünstigten Mittelmaß und verführten zu Manipulationen. Wo-

nach sollten sich die Geldgeber dann richten? Ganz ohne Evaluationen gehe es nicht, räumen Osterloh und Frey ein. Ziel müsse aber in erster Linie eine Bewertung der Qualität des Forschungsprozesses sein, nicht der Ergebnisse. (hig.)

ger Bedeutung: Es bleibt völlig offen, was daraus zu schließen ist. Sollten den für schlecht befundenen Institutionen und Forschenden die Mittel gekürzt werden? Oder sollten ihnen nicht zusätzliche Mittel bewilligt werden, damit sie ihre Qualität erhöhen können? Sollten umgekehrt den für gut befundenen Institutionen und Forschenden die Mittel gekürzt werden, weil sie ja ohnehin erfolgreich sind? Führen bei ihnen zusätzliche Mittel möglicherweise zu einem abnehmenden Zusatzbeitrag? Eine sinnvolle Evaluation müsste den marginalen Effekt einer Änderung der Mittel erfassen. Was würde geschehen, wenn einer Institution oder einem Forscher mehr oder weniger Mittel zur Verfügung stehen? Diese Frage ist äußerst schwierig zu beantworten, denn sie hängt von einer großen Zahl von Bedingungen ab. Außerdem bleibt offen, wie die Ergebnisse einer derartigen Evaluation in der politischen Auseinandersetzung aufgenommen werden. Dies zeigen die Reaktionen auf die Exzellenzinitiative des Bundes und der Länder.

Bei aller Fragwürdigkeit von Evaluationen könnte argumentiert werden, dass es keine Alternative gäbe. Sinnvolle Alternativen gibt es durchaus.

Sinnvolle Alternativen gibt es durchaus

Mit einer veränderten institutionellen Ausgestaltung des Wissenschaftsbetriebs können die permanenten Evaluationen zurückgedrängt und teilweise sogar ersetzt werden: Werden Universitäten einem stärkeren Wettbewerb unterworfen, ist keine staatliche Evaluation mehr nötig. Die Studierenden wählen diejenige Universität, die ihrer Ansicht nach die besten Leistungen bietet. Die Universitäten haben die Freiheit, sich diejenigen Studierenden auszuwählen, die ihre Anforderungen am besten erfüllen und ihre Reputation steigern.

Die Studierenden mögen bei ihrer Auswahl Evaluationen und Rankings zugrunde legen. Es gibt eine Vielfalt existierender Ranking-Angebote, die alle mehr oder weniger für begrenzte Fragestellungen taugen, die Forschungsleistung aber schlecht oder gar nicht messen. Auch hier ist ein Wettbewerb zwischen den verschiedenen Rankings besser als der Versuch, das von der Politik gewünschte Super-Ranking zu erstellen, wie dies der Deutsche Wissenschaftsrat gegenwärtig tut. Auch ein mit größter Sorgfalt durchgeführtes Ranking kann nicht verhindern, dass sich in der Folge die geschilderten Anreizverzerrungen einstellen.

Die heute üblich gewordene nachträgliche Evaluation wissenschaftlicher Einrichtungen lässt sich zu einem guten Teil umgehen, wenn die Forschenden und Lehrenden sorgfältig ausgeselen werden. Diese Strategie setzt die Ressourcen zukunftsorientiert ein, indem das Gewicht auf die Selektion gelegt wird. Dabei müssen durchaus die herrschenden Kriterien wie die Anzahl und Qualität der Publikationen zur Anwendung kommen. Sie sichern, dass Standards der Wissenschaftlichkeit erfüllt sind, und geben Hinweise auf das Potential der Kandidaten. Ist eine Person aber einmal als Professor für ein bestimmtes Wissensgebiet nach strengen Kriterien ernannt, muss ihr vertraut werden. Deshalb sind Berufungsverfahren das mit Abstand wichtigste Geschäft einer wissenschaftlichen Institution. Aufgrund der sorgfältigen Auslese kann damit gerechnet werden, dass die berufene Person die erwarteten Leistungen auch ohne die Knete ständiger Evaluationen erbringt. Man lässt sie in Ruhe arbeiten. Dabei ist durchaus mit einer gewissen Varianz zu rechnen. Manche unter den Ausgewählten werden in ihrer Leistung nachlassen, andere hingegen werden durch den gewählten Freiraum beflügelt und zu Spitzenleistungen motiviert. In der Wissenschaft sollten Letztere zählen.

Unwillig und Versager müssen als notwendiges Übel betrachtet werden, damit das Wissenschaftssystem als Ganzes Höchstleistungen erbringen kann. Hingegen können dauernde Evaluationen, insbesondere Ergebnis-Evaluationen, nur ein Durchschnittsniveau sichern. Die als Kontrolle erlebten fortwährenden Beurteilungen begünstigen eine „normale“ Wissenschaft ohne Höchstleistungen. Forscher wie Albert Einstein oder Max Planck in der Naturwissenschaften und John Maynard Keynes oder John Hicks in den Wirtschaftswissenschaften wären im derzeitigen System permanenter Evaluationen wahrscheinlich nicht sehr erfolgreich gewesen.

Evaluationen von Forschungsinstitutionen können gleichwohl nicht ganz vermieden werden, weil sonst keine Kriterien der Mittelzuweisung existieren. Aber sie sollten in erster Linie prozess- und nicht ergebnisorientiert erfolgen, wie dies der Wissenschaftsrat ansonsten bisher praktiziert hat. Wichtigste Kriterien dabei sind, ob ein sorgfältiger Prozess der Personalauslese gewährleistet und ein hoher Grad an Autonomie im Forschungsprozess gesichert ist. Damit wird nicht nur den besonderen Problemen der Leistungsermittlung im Wissenschaftssystem Rechnung getragen, sondern es werden zugleich die für Forschenden wichtigsten Anreize – Autonomie sowie ein anregendes wissenschaftliches Umfeld – geschaffen.

Ein nach diesen Kriterien gestaltetes System hat der deutschsprachigen Wissenschaft in der Vergangenheit Weltgeltung verschafft. Es besteht nach wie vor an akademischen Leuchttürmen wie der Harvard University, die doch sonst so gerne als Vorbild genommen werden.



Absurder Wettlauf: Wie beim Rennen zwischen Hase und Igel suchen Wissenschaftler ihren Bewertern ein Schnippchen zu schlagen.

Foto bpk

gegen das herrschende Paradigma durchsetzen muss. Zahlreiches Anschauungsmaterial findet sich in der Literatur im Anschluss an Thomas Kuhn über „wissenschaftliche Revolutionen“ oder an Ludwig Fleck über „wissenschaftliche Denkkollektive“.

Die Leistungskriterien können manipuliert werden

Spitzenforschung braucht Zeit, bis sie zu bewertbaren Ergebnissen führt, und noch länger, bis ihre Bedeutung innerhalb des Hauptstroms der Wissenschaft erkannt ist. Eine an kurzfristigen Publikations- und Zitationsraten orientierte Leistungsermittlung hätte vielen bahnbrechenden Forschern wenig Chancen gegeben. Ohnehin gibt es zahlreiche Beispiele dafür, dass der Hauptstrom der Scientific Community Neuerungen oft eher hindert als fördert. Ein besonders dramatisches Beispiel ist Ignaz Semmelweis, der Entdecker des Kindbettfiebers. Er musste über dreißig Jahre auf die Anerkennung und Anwendung seiner Forschungsergebnisse warten, die Tausenden Frauen das Leben gerettet hätte.

Wenn ein Indikator für die eigene Position wichtig wird, wird ein starker Anreiz ausgeübt, diesen Indikator zu den eigenen Gunsten zu beeinflussen. Schulleitungen können die Beurteilung ihrer Schule anheben, indem sie die Schüler auf die Examensaufgaben hin trimmen, schlechte Schüler unter Vorwänden von den entsprechenden Tests ausschließen und damit die Ergebnisse ihrer Schule künstlich verbessern. Manche Manager beeinflussen die Leistungsindikatoren, sobald ihr Einkommen davon abhängig ist. Sie treiben kurzfristig die Aktienkurse in die Höhe, wenn sie einen Teil ihres Gehaltes in Form von Aktienoptionen erhalten.

Derartige Manipulationen haben sich auch in der Wissenschaft verbreitet, seit im Zuge von Evaluationen die Forschungsleistung anhand quantitativer Kriterien gemessen wird. So werden gern Wissenschaftler mit entsprechenden Leistungsausweisen an eine Universität verpflichtet, damit diese in Evaluationen und Rankings gut abschnitten. Streit bricht aus zwischen Universitäten, wer einzelne Nobelpreisträger für sich verbuchen darf. Für die Wissenschaftskultur besonders schädlich ist das Hochjubeln von Ergebnissen in der Forschung. So besteht ein verstärkter Anreiz, nur noch erfolgreiche

Tests zu publizieren, die negativen Ergebnisse zu verschweigen oder sogar zu beseitigen, obwohl die Falsifikation von Hypothesen zu den Kernaufgaben der Wissenschaft gezählt wird. Einer Untersuchung zufolge haben nicht weniger als ein Drittel der amerikanischen Wissenschaftler sich unehrlich oder gar betrügerisch verhalten, indem Ideen anderer ohne angemessene Zitierung übernommen oder unerwünschte Ergebnisse unterdrückt wurden.

Schlimmer noch ist die Fälschung von Forschungsergebnissen. In Experimenten wurde gezeigt, dass sich kontrolliert fühlende Personen in weit stärkerem Ausmaß bereit sind, zu betrügen. Jüngste Skandale zeigen, dass dies auch im Wissenschaftsbetrieb vorkommt.

Die mit der Evaluation einhergehende Beurteilung der Leistung beeinflusst die Arbeitsmotivation negativ, wenn eine solche Bewertung von den Betroffenen als kontrollierend empfunden wird. Es tritt ein Verdrängungseffekt auf, wonach die intrinsische Arbeitsmotivation abnimmt und die extrinsisch bestimmten Anreize an Gewicht gewinnen. Die Gesamtlei-

tung – gemessen nach den vordefinierten Kriterien – vermindert sich nicht notwendigerweise, sondern steigt mitunter sogar. Es muss jedoch bezweifelt werden, dass die Auswirkungen auf die Qualität und Originalität der Forschung günstig sind. Kreative Forschung zeichnet sich gerade dadurch aus, dass neue Maßstäbe geschaffen werden, die sich erst langsam durchsetzen. Sie ist deshalb in hohem Maß auf Eigenantrieb angewiesen, der nicht verdrängt werden darf. Darüber hinaus hat die permanente Evaluation einen negativen Signalisierungseffekt. Sie demonstriert, dass den Wissenschaftlern das Vertrauen entzogen wird, eigenverantwortlich eine hohe Leistung in Forschung und Lehre zu erbringen. Diese Misstrauensbekundung kann die Loyalität zur beschäftigenden Institution verringern.

Empirische Befunde zum Zusammenhang von Leistungskontrolle und Leistung im Wissenschaftsbetrieb sind bisher kaum vorhanden, weil die Messung der Leistung äußerst schwierig ist. Aus anderswo gewonnenen Befunden lässt sich jedoch schließen, dass variable monetäre

Entlohnung unter bestimmten Bedingungen zu einer geringeren Leistungsbereitschaft führt als eine als fair empfundene fixe Entlohnung, während nichtmonetäre Belohnungen (zum Beispiel Entlastung von der Lehre) die Motivation nicht oder allenfalls schwach verdrängen. Nicht erwartete monetäre Belohnungen, etwa Schenkungen, scheinen die Eigenmotivativität nicht zu verdrängen, symbolische Belohnungen (zum Beispiel Auszeichnungen) stärken sie.

Werden Institutionen oder Personen für eine Evaluation vorgesehen, können sie sich nicht dagegen wehren, selbst wenn sie überzeugt sind, dass diese sich für ihre Verhältnisse nicht eignen. Es würde ihnen vorgeworfen, sie hätten Angst vor dem Ergebnis. Da die Evaluation meist mit einer Mittelvergabe einhergeht, müssen sie sich wider bessere Einsicht an der Evaluation beteiligen. Sie tun sogar gut daran, begeistert mitzumachen. Damit wird ein Einverständnis vorgetuschelt, das in Wirklichkeit nicht vorhanden ist.

Werden sie alsdann in der Evaluation gut eingeschätzt, sind sie erfreut und hoffen auf entsprechend höhere Ressourcenzuweisungen. Die Verlierer werden hingegen große Anstrengungen unternehmen, sich gegen die Auswirkungen zu wehren. Dafür stehen immer Argumente zur Verfügung: Überlastung in Lehre oder Administration, zu wenig Ressourcen oder ganz einfach Pech. Im Nachhinein wird versucht, die Kriterien zum eigenen Vorteil anders zu gewichten.

Evaluationen haben nicht nur versteckte, meist übersehene Kosten. Vielmehr wird ihr Nutzen zu hoch veranschlagt. Die durch die Evaluation gewonnenen Informationen tragen oft wenig dazu bei, die Entscheidungen über die Ressourcenzuweisung für wissenschaftliche Forschung zu verbessern. Gerade in der Scientific Community ist meist wohlbekannt, welche Institutionen und Personen nach den gängigen Kriterien besonders gute oder schlechte Forschung betreiben. Hinzu kommt, dass die angewendeten Verfahren nur im oberen und im unteren Leistungsbereich übereinstimmende und verlässliche Ergebnisse liefern. Im mittleren Leistungsbereich – wo die Informationen am nötigsten wären – differenzieren sie unzuverlässig.

Evaluationen suchen in aller Regel das bestehende Leistungsniveau zu erfassen. Für politische Entscheidungen sind jedoch diese Informationen oft von gering-

Die Autoren



Margit Osterloh

Das gemeinsame wissenschaftliche Zuhause von Margit Osterloh und Bruno S. Frey ist die Universität Zürich. Die 1943 in Brandenburg an der Havel geborene Betriebswirtin und der 1941 in Basel geborene Volkswirt verfügen über



Bruno S. Frey

kann. Zu den Spezialgebieten von Frau Osterloh, die sich an der Freien Universität Berlin promovieren ließ und an der Universität Erlangen-Nürnberg habilitiert hat, gehören die Organisations- und Managementtheorie, aber auch das Management von Innovationen sowie die Regeln guter Unternehmensführung (Corporate Governance). Sie kennt die Wirtschaft nicht nur aus der Forscherperspektive. Sechs Jahre lang hat sie in einem mittelständischen Unternehmen gearbeitet, heute ist sie Mitglied in Aufsichts- und Verwaltungsräten. Mit Frey verbindet sie das Interesse an Fragen der Motivation. Wie müssen Unternehmen und Institutionen organisiert sein, damit die Mitarbeiter aus eigenem Antrieb gute Leistungen hervorbringen? (hig.)