

A photograph of a historic stone building, likely the Sorbonne in Paris, during a protest. In the upper part, a group of people is gathered on a balcony. In the foreground, several riot police officers wearing helmets and shields are visible, facing a crowd of people. The building's facade features large windows with metal grates and classical architectural elements like columns and arches. A red circular sticker with the text 'JULI ET AOUT' is visible on the wall.

Der richtige Zeitpunkt

Reif für Veränderung? Wann Evaluationen sinnvoll sind

Persönliche Erfahrungen als Begutachtender und Begutachteter sprechen dafür, dass Evaluationen nur dann gelingen, wenn eine Institution reif dafür ist und sich beide Seiten uneigennützig für ihre Zukunft engagieren.

Widerstand der Betroffenen:
Im März 2009 besetzten Studierende monatelang die Sorbonne in Paris und andere Universitäten in Frankreich. Sie protestierten gegen die Bildungsreform, die das französische Wissenschaftssystem wettbewerbsfähiger und Forschungsergebnisse international sichtbarer machen soll.

ABB.: REUTERS/BENOIT TESSIER



VON MARTIN LOHSE

EVALUIEREN IST EN VOGUE – und das zunehmend. Auch die Bayerische Akademie hat sich, halb gedrängt, halb gewollt, einer Evaluation unterzogen und müht sich nun mit den Ergebnissen. Manche hoffen auf Neuerung, während andere Altes bewahren wollen. Evaluiert und begutachtet wird aber landauf, landab. Vom kleinen innerfakultären Projekt bis hin zu Forschungszentren und ganzen Institutionen – der externe Sachverstand soll leisten, was der innere vermeintlich oder tatsächlich nicht kann: die Spreu vom Weizen trennen und Einrichtungen zukunftsfähig machen.

Mit der steigenden Bedeutung von Evaluationen haben sich Spezialisten herausgebildet, Gesellschaften für Evaluation, Zeitschriften und Handbücher sowie ganze Einrichtungen wie das Centrum für Hochschulentwicklung (CHE) oder das Institut für Forschungsinformation und Qualitätssicherung (IFQ). Es gibt natürlich auch kritische Stimmen – zu selbsternannten Besserwissern und zu ungesicherten Methoden ebenso wie zum Umsichgreifen des Evaluierens überhaupt. Schon 1994 veröffentlichte Michael Power von der London School of Economics einen kleinen Band „The Audit Explosion“, dem

1997 die „Audit Society“ folgte. Im deutschsprachigen Raum hat sich der Wirtschaftswissenschaftler und Glücksforscher Bruno Frey aus Zürich an die Spitze der Kritik zu setzen versucht. In seinem Aufsatz „Evaluierungen, Evaluierungen ... Evaluitis“ (2007) schreibt er: „Das Fortschreiten von Evaluationen scheint unaufhaltsam und kann mit einer Modewelle verglichen werden, bei der sich ohne viel Überlegung alle beteiligen, die nicht als rückständig gelten wollen.“ Solche Kritik ist ebenso wie die von verschiedenen Fachverbänden – von Mathematikern bis zu Soziologen und Historikern – von den Medien aufgenommen worden und hat etwa Jürgen Kaube von der Frankfurter Allgemeinen Zeitung zu einer Reihe beißender Artikel angeregt.

Dieser Artikel will aus subjektiver Perspektive, wenn auch von Zahlen durchsetzt, über Begutachtungen und Evaluationen berichten und zum Nachdenken darüber anregen, wann Evaluieren und Begutachten angebracht sein können und wann nicht. Die Beispiele sind gegriffen aus Erlebtem, vom Einzelantrag bis zur Exzellenzinitiative, von der Graduiertenschule bis zum Forschungszentrum, aus Deutschland und aus anderen Ländern. Ich habe versucht, von einzelnen Erlebnissen so weit wie möglich zu abstrahieren. Wer sich trotzdem zu erkennen meint, möge denken, er täusche sich. Oder mir verzeihen.

Begutachten versus evaluieren

Wissenschaftliches Arbeiten kann auf verschiedene Weise begutachtet werden: im Voraus oder im Nachhinein, kontinuierlich oder periodisch und mit unterschiedlicher Intensität. Wenn es im Voraus geschieht – Evaluationsprofis sprechen von *ex ante* –, so bezeichnen wir das gemeinhin als Begutachtung, wenn es im Nachhinein erfolgt (für Profis *ex post*), so gilt es als Evaluation.

Begutachtungen dienen in unserem Wissenschaftssystem zwei Hauptzwecken: der Auswahl von Personen (besonders bei Berufungen) und der Allokation von Ressourcen. Begutachtungen sind Projektionen in die Zukunft, die sich im Wesentlichen auf Anträge derjenigen stützen, die sich um Mittel oder Positionen bewerben. Aber auch wenn es hierbei ganz wesentlich um die Beurteilung des konkreten Antrages geht, kann man als Gutachter von der antragstellen-

Abb. 1: Evaluitis – eine neue Krankheit? Der Schweizer Verhaltensökonom Bruno Frey verglich Evaluierungen 2007 mit einer Modewelle, an der sich „ohne viel Überlegung alle beteiligen, die nicht als rückständig gelten wollen“.



den Person nicht abstrahieren – ja, sie und ihre bisherigen Leistungen stellen im Sinne einer ex post-Evaluation einen wesentlichen Faktor bei der Prognose dar, ob ein Projekt gelingen und ob Konzepte und Interpretation originell sein werden.

Evaluationen dagegen betrachten, ob die Versprechungen gehalten wurden, die für ein Projekt, eine Einrichtung oder Organisation gegeben wurden. Sie versuchen, Ergebnisse systematisch zu erfassen und gleichzeitig Vorschläge für Verbesserungen zu machen. Und sie dienen nicht zuletzt der Rechtfertigung von Kosten, vor allem bei öffentlichen Mitteln.

Der Aufwand

Kann man bestimmen, wie groß der Aufwand ist, den die deutsche Wissenschaft für Begutachtungen und Evaluationen betreibt? Für zwei Aspekte, Antragstellung bei der Deutschen Forschungsgemeinschaft (DFG) und Publikationen, will ich eine grobe Schätzung wagen.

Die DFG gibt ihren eigenen Verwaltungsaufwand mit ca. 3 Prozent an – vor allem für die Geschäftsstelle, also die Organisation von Begutachtungen und die Verteilung und Verwaltung der Mittel. Gutachter und Gremienmitglieder arbeiten im Wesentlichen unbezahlt, ihr Aufwand lässt sich aber ungefähr schätzen (s. Abb. 2). Demnach braucht die gesamte Begutachtung für die DFG (ohne Exzellenzinitiative) jährlich etwa 150 Personenjahre, oder – bei Jahreskosten von rund 100.000 Euro für ein Gutachterjahr – 15 Mio. Euro. Bei einem Haushalt von ca. 2,5 Mrd. Euro (2012) entspricht das einer erstaunlich geringen Quote von 0,6 Prozent, oder, wenn die Gutachter ein wenig gründlicher arbeiten, vielleicht 1 Prozent.

Im Gegensatz dazu ist der Aufwand für das Begutachten von Publikationen erheblich höher. Legt man die im Web of Science angegebenen rund 300.000 Publikationen aus Deutschland (2013) zu Grunde und geht vom in Abbildung 3 geschätzten Verlauf einer Publikation und ihrer Begutachtung aus, dann sind für jede einzelne 16 Begutachtungsjahre erforderlich – oder 3.000 Begutachtungsjahre für alle zusammen, d. h. 300 Mio. Euro im Jahr.

Damit könnte es eigentlich auch gut sein. Doch zu diesen rein forschungsbezogenen Begutachtungen kommt eine immer größer werdende Zahl weiterer Bewertungen: von Lehre, von Studiengängen, von Instituten und Institutionen. Für Akkreditierungsagenturen und Ministerien, für Rankings und Zielvereinbarungen. Und jeder Fußbreit Autonomie, den Hochschulen und Forschungseinrichtungen gewinnen, bringt eine stete Berichts- und Evaluationspflicht mit sich, die in ihrer Kleinteiligkeit die gewonnene Freiheit wieder zunichtemacht. „Nie habe ich öfter ins Ministerium fahren müssen, als seit unserer Autonomie“, stöhnte der Direktor eines Universitätsklinikums einmal.

Abb. 2: Abschätzung des Begutachtungsaufwandes für die DFG. Angenommen ist eine Begutachtungszeit von 4 Stunden pro Einzelantrag bzw. eine Vorbereitungszeit, die der Präsenzzeit bei Vor-Ort-Begehungen bzw. bei Gremiensitzungen entspricht. Ein Personen-Jahr ist mit 200 Arbeitstagen angenommen.

	Personen-Jahre/Jahr
Normalverfahren: ca. 16.000 Anträge, je 2 Gutachten à 4 h	80
Vor-Ort-Begehungen: 250 Begehungen, je 8 Gutachter à 4 T	40
Gremiensitzungen: Fachkollegien, Ausschüsse, Senat	30
Summe	150

	Personen-Jahre/Jahr
ca. 300.000 Publikationen <ul style="list-style-type: none"> • Erste Einreichung, abgelehnt: je 2 Gutachten à 3 h • Neu-Einreichung, zur Revision zugelassen: je 2 Gutachten à 3 h • Revision, angenommen: je 2 Gutachten à 2 h Summe: 16 h pro Publikation	
Insgesamt	3.000

Neben- und Hauptziele

Evaluationen brauchen Material, am besten quantifizierbares. Doch das ist in der Wissenschaft schwierig – denn das eigentlich Wichtige, die großen Theorien und Ergebnisse, zeigt sich meist erst spät. Die Evaluation aber findet im Jetzt statt und sucht die schnelle Objektivität. So steigt die Gefahr, dass man sich den Nebensatt den Hauptzielen widmet: Anzahl von Publikationen und Zitationen, Promotionen und Frauenquoten – statt Forschungsergebnisse und fruchtbringende Hypothesen. Über die Problematik solcher Parameter ist viel gesagt und

Abb. 3: Abschätzung des Begutachtungsaufwandes für Publikationen aus Deutschland. Zahl der Publikationen nach ISI Web of Science für 2013.

Abb. 4: Der Mediziner Brian Kobilka (links) bei der Nobelpreisverleihung 2012. Seine strukturellen Forschungen an Rezeptoren, die in der Zellkommunikation eine wichtige Rolle spielen, waren seinen Gutachtern suspekt.

geschrieben worden. Mathematiker kritisierten die mangelnde Validität von Zitationsanalysen (www.math-union.org/fileadmin/IMU/Report/CitationStatistics.pdf), Drittmitteleinwerbung wurde als Maß für die Kosten, nicht für den Wert von Forschungen benannt, und ganze Sammelbände wurden den Verbiegungen von Evaluatoren und Evaluierten gewidmet (etwa: H. Matthies, D. Simon [Hrsg.], *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*, Wiesbaden 2008; R. Hohls, C. Prinz [Hrsg.], *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaften?* Historisches Forum 12, Berlin 2009).



Gelingende und misslingende Evaluationen

Es gibt viele Bedingungen, unter denen Begutachtungen und Evaluationen nicht funktionieren – etwa bei persönlichen Abhängigkeiten, bei zu geringen Ressourcen oder wenn sie nicht ernstgemeint sind. Vor allem aber können sie in aller Regel nur das Mittelmaß erfassen. Wirklich Gutes wird oft nicht erkannt – weil es unfertig ist und seine Schöpfer noch unsicher sind, weil der Autor gegen den Strom schwimmt und das Neue (noch) nicht in die Köpfe will. Brian Kobilka etwa, der 2012 den Nobelpreis für Chemie für die Lösung der Struktur von Rezeptoren erhielt, wurde 2003 aus den Howard Hughes Medical Institutes geworfen, weil Mediziner solche Dinge nicht machen sollten. Zum Glück hielten sein Department in Stanford und enge Freunde zu ihm und stützten ihn, bis wenige Jahre später das Wagnis gelang. James Black, der Erfinder der Beta-Blocker, pflegte die Anekdote zu erzählen, er habe bei einem Treffen von Nobelpreisträgern die Frage gestellt, wessen entscheidende Entdeckung in einem Forschungsantrag begutachtet worden sei – keiner habe sich gemeldet. Offenbar sehen häufig nur die Betroffenen selbst oder enge Vertraute das Potential bedeutender Entwicklungen.

Wie können Evaluationen trotzdem erfolgreich sein? Mir scheint, dass mehrere Bedingungen zusammenkommen müssen, damit dies wirklich gelingt. Erstens muss das Objekt der Evaluation dafür reif sein. Aus sich heraus, etwa wenn sich die Frage nach künftigen Forschungsschwer-

punkten in einer Universität stellt. Oder von außen, wenn ein Zustand an sein Ende kommt, etwa zu Beginn der Exzellenzinitiative die Fiktion, dass alle Universitäten gleich seien. Oder unter Druck, wie in Frankreich, als die mangelnde Wettbewerbsfähigkeit des Wissenschaftssystems ab 2006 zur Etablierung von Einrichtungen der Forschungsförderung und -evaluation führte, wie wir sie aus anderen Ländern schon lange kennen. Druck allein aber reicht selten: Selbst die längst fälligen Veränderungen führten zum Widerstand der Betroffenen, die im Frühjahr 2009 monatelang die Sorbonne und andere Universitäten besetzten.

Der Evaluator

Nicht überschätzt werden kann die Bedeutung der Person des Evaluators, mehr noch seiner Persönlichkeit. Jeder Wissenschaftler kennt und fürchtet den Besserwisser, der in der Machtposition des Gutachters seine schnell gewonnenen Erkenntnisse für die Wahrheit hält und doch zum Gelingen nicht beiträgt.

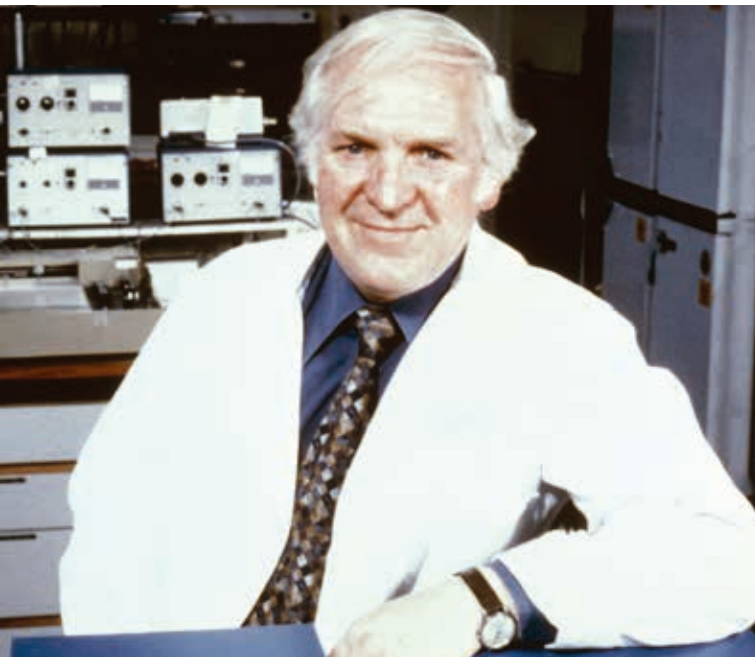
Mir scheint, es gibt zwei Typen von erfolgreichen, nützlichen Evaluatoren: den neutralen Beobachter und den kritischen Freund. Beide sind schwer zu finden. Der neutrale Beobachter, weil Forschung weltweit vernetzt ist und – das haben mir immer wieder die Begutachtungen der Ex-

zelleninitiative gezeigt – auch weil der aus der Ferne Kommende den zu Begutachtenden besser kennt, als Außenstehende denken mögen. Seine Befangenheit ist dann nur halb bewusst.

Auch der kritische Freund ist selten, weil hier Kritikfähigkeit, Erfahrung, beiderseitiges Wohlwollen und Interesse für die Sache zusammentreffen müssen. Kritische Freunde sind mir selbst am wertvollsten gewesen – bei der Diskussion der eigenen Forschung ebenso wie bei der Leitung einer Einrichtung. Und als kritischer Freund habe ich auch selbst am liebsten agiert. Als Begutachtender oder Rat Gebender spürt man dann, dass es in der Institution im Grunde schon das Wissen gibt, wo sie hin will – dass es nur noch nicht an der Oberfläche ist. Damit ist die Rolle des Evaluators nicht weit

lässe waren die Rechtfertigung öffentlicher Mittel, aber auch die veränderte Rolle der Landesakademien und ihre Funktion als Forschungseinrichtungen. Doch im Hintergrund steht ebenso die Bedrohung der Akademie durch andere Einrichtungen, die um Mittel und Bedeutung konkurrieren.

Wie jede Veränderung wird die jetzt anstehende manchmal schmerzhaft sein – wenn auch anders als bei der Sorbonne die Besetzung der Münchner Residenz nicht unmittelbar bevorzustehen scheint. Aber auf das Miteinander von Gelehrtengesellschaft und Forschungseinrichtung, auf die Struktur der vielen Kommissionen, auf die Trennung von Projektleitung und -aufsicht und auf die Auswahl thematischer Schwerpunkte wird mancher Blick zu richten sein.



Eine große Chance scheint mir die Empfehlung zu enthalten, die Akademie möge neben den Langzeitvorhaben ein neues, zweites Forschungsinstrument installieren: Forschungsgruppen und Institute auf Zeit. Sie könnten diese Vorhaben bei sich oder in den Einrichtungen ihrer Mitglieder realisieren, bräuchte sich also nicht durch Forschungsgebäude oder -instrumente zu belasten und zu binden und könnte doch gerade in den Gebieten, auf denen sich Neues tut, schnell und an vorderster Front aktiv werden. Solche Vorhaben könnten aus dem interdisziplinären Gespräch der Akademie entstehen und würden nicht nur zu einer engen Verzahnung von Gelehrtengesellschaft und ihren Forschungen führen, sondern auch die Akademie als einen innovativen Akteur der bayerischen Wissenschaftslandschaft stärken.

Abb. 5: Erfinder der Beta-Blocker: Der britische Pharmakologe und Nobelpreisträger James Whyte Black kritisierte, dass Begutachtungen wirklich Neues nicht erfassen können.

von der eines Therapeuten, der beim Gegenüber nur das zu wecken braucht, was schon darauf wartet. Wozu auch wieder gehört, dass die zu evaluierende Einrichtung dafür reif sein muss, wenn Veränderungen erfolgreich sein sollen.

Die Evaluation der BAdW – Bedrohung oder Chance?

Auch die Bayerische Akademie der Wissenschaften hat nun ihre Evaluation hinter sich – niedergelegt auf fast 300 Seiten im „Evaluierungsbericht Strukturkommission Bayern 2013“. An-

Danksagung

Ich möchte mich bedanken: bei denen, die ich begutachtet habe, für das, was ich von ihnen lernen durfte, bei denen, die mich begutachtet haben, für ihre Nachsicht, bei den Mitgliedern der Reformkommission der Akademie für anregende Gedanken und Gespräche. Beate Konze-Thomas danke ich für die Zahlen zur DFG.

DER AUTOR

Prof. Dr. Martin Lohse, Mitglied der Bayerischen Akademie der Wissenschaften seit 1998, ist Pharmakologe an der Universität Würzburg und leitet dort das Rudolf-Virchow-Zentrum / DFG-Forschungszentrum für Experimentelle Biomedizin. Seit 2009 ist er Vizepräsident für Forschung an seiner Universität und ebenfalls seit 2009 Vizepräsident der Nationalen Akademie der Wissenschaften Leopoldina.