

Sonderdruck aus

BERLIN-BRANDENBURGISCHE
AKADEMIE DER WISSENSCHAFTEN

(vormals Preußische Akademie der Wissenschaften)

Berichte und Abhandlungen

Band 9



Akademie Verlag
Berlin 2002

Randolf Menzel, Friedhelm Neidhardt, Peter Weingart

Einleitung

Das Thema „Evaluation“ der Wissenschaft und ihrer Einrichtungen ist in der Bundesrepublik in den neunziger Jahren mit Macht auf die Agenda der Wissenschaftspolitik geraten, und alles spricht dafür, daß dieses Thema nicht mehr verschwinden wird. Inzwischen werden in großer Zahl Universitäten, Forschungsinstitute und sogar die Max-Planck-Gesellschaft und die DFG evaluiert. Jürgen Mittelstraß geht so weit, von einem „neuen Furor“ zu sprechen: Wir seien „in der Wissenschaft auf dem besten Weg, ein Volk von Evaluierern zu werden“¹. Es gibt keinen Zweifel: Die Evaluation der Wissenschaft „von außen“, das heißt durch andere Instanzen als sie selbst, stellt einen grundlegenden Wandel im Verhältnis der Wissenschaft zur Gesellschaft dar. Sie ist ein sicheres Anzeichen der Veränderung des „Gesellschaftsvertrags“ für die Wissenschaft, und die Bundesrepublik wird von dieser Entwicklung nicht einmal allein und schon gar nicht zuerst erfaßt.

Auch wir „Herausgeber“ dieser Beiträge sind nolens volens damit befaßt, weil man auch uns evaluiert und weil wir selber als Gutachter in Evaluationen eingespannt sind. Hinzu kommt, daß die Diskussionen, die sich in Wissenschaft und Politik hierzulande entwickelt haben, uns zu interessierten, manchmal auch etwas irritierten Beobachtern des allgemeinen Umgangs mit dem Evaluationsthema gemacht haben.

Aus dieser mehrschichtigen Berührung mit dem Thema entwickelte sich bei uns als Mitgliedern der Berlin-Brandenburgischen Akademie der Wissenschaften die Vorstellung, die Akademie, die selber nicht unter aktuellem Evaluationsdruck steht, sei ein guter Ort, um einigermaßen unbefangen über den Stand der Dinge und über Sinn und Unsinn weiterer Entwicklungen zu diskutieren. Dabei interessierte uns weniger die ausweglose Fundamentalisierung der Frage, ob die Qualität von Forschung überhaupt zuverlässig bestimmbar und wenn bestimmbar, sogar meßbar und damit auch vergleichbar sei. Wir halten Evaluationen der Forschung vor allem dann, wenn diese jenseits der Märkte mit Steuergeldern alimentiert wird, für un-

¹ Gegenworte. Zeitschrift für den Disput über Wissen, 5 (2000), S. 24.

umgänglich, und wir halten sie mit tolerierbaren Annäherungswerten an das, was sich als Forschungsqualität definieren läßt, grundsätzlich auch für möglich. Die Frage ist nur, wie, also mit welchen Instrumenten und Verfahren man Evaluationen betreibt, und wie man dann mit ihren Ergebnissen umgeht.

Im Hinblick darauf interessierten uns bei dem Vorsatz, zum Evaluationsthema in der Berlin-Brandenburgischen Akademie der Wissenschaften eine Konferenz zu veranstalten, eher die Erfahrungen der Evaluationspraktiker als die Meinungen (wissenschafts)theoretisch argumentierender Bedenkensträger. In diesem Sinne haben wir – unterstützt von Dieter Simon, dem Akademiepräsidenten, und assistiert von Wolf-Hagen Krauth – für den 27. November 1999 insgesamt sechs Referenten und mehr als 30 Gäste zu einer eintägigen Konferenz in die Akademie eingeladen, und wir freuen uns besonders darüber, daß auch die ausländischen Experten, auf die wir neugierig waren, unserer Einladung gefolgt sind. Was ist herausgekommen?

Im folgenden werden die Texte der meisten Vorträge abgedruckt, die auf der Konferenz gehalten wurden. Selbstverständlich handelt es sich dabei nur um einen kleinen Ausschnitt aus der Diskussion. Für die Auswahl war bestimmend, daß auch die Erfahrungen aus zwei Ländern präsentiert werden sollten, in denen die Evaluation der Wissenschaft bereits seit einiger Zeit praktiziert wird und dementsprechendes Wissen vorliegt. Ben R. Martin berichtet deshalb über die Forschungsevaluierung im Vereinigten Königreich, Barend van der Meulen über die in den Niederlanden. Sodann sollten Erfahrungen und Reaktionen mit der Evaluierung in der Bundesrepublik vor allem aus der Sicht der Wissenschaftspolitik dargestellt werden. Helga Schuchardt und Hans Joachim Meyer haben sich aus der politischen Perspektive geäußert. Der Beitrag von Stefan Hornbostel fehlt hier. Eine veränderte Fassung ist in den *Gegenworten*, Heft 5, 2000, abgedruckt, zusammen mit einer Sammlung weiterer Beiträge zum Thema, die jedoch nicht Gegenstand der Konferenz waren. In unserer kurzen Einleitung wollen wir Veranstalter einige summarische Eindrücke aus der Wahrnehmung der Vorträge und über die Diskussion dieser Vorträge formulieren:

(1) Nachdem in vergleichbaren Ländern – in Großbritannien und den Niederlanden, um nur europäische Beispiele zu nennen – eine nahezu alle Wissenschaftsdisziplinen erfassende Evaluationspraxis schon seit längerem geübt und eingewöhnt ist, haben auch hierzulande zahlreiche Feldexperimente eingesetzt, an denen eine Fülle von Akteuren noch ziemlich unkoordiniert beteiligt ist. Auch wenn dieser Prozeß gegenwärtig einige Dissonanzen erzeugt, sollte man ihn noch offenhalten für den Wettbewerb um akzeptable Lösungen. Man sollte sich auch gar nicht wünschen, daß am Ende geschlossene Standardformeln entstehen, die sich ritualisieren lassen und dann nur noch ministerialbürokratisch exekutiert werden müssen. Bei allem Bedarf an Standardisierungen müssen Evaluationsprogramme offenbleiben sowohl für die Besonderheiten der Institutionen und Akteure, auf die sie sich beziehen, als auch für Revisionen, die die Arbeit mit ihnen nahelegt.

(2) Bei allen Meinungsverschiedenheiten, die gegenwärtig vorhanden sind, lassen sich doch die Übereinstimmungen nicht übersehen, die sich zu Evaluationsverfahren zu bilden beginnen. Es gibt wohl kaum noch Zweifel daran, daß hinreichend valide Evaluationen komplexe und mehrschichtig angelegte Prüfprozesse voraussetzen.

Dies gilt zuerst einmal in sozialer Hinsicht. Hier ist entscheidend, daß gute Evaluationsergebnisse Interaktionsprodukte sind, deren Qualität die Mitwirkung der Betroffenen voraussetzt. Fremdevaluation muß als grundlegendes Prinzip berücksichtigen, daß erfolgreiche Evaluation das Vertrauen der Betroffenen voraussetzt. Daß diese befangen sind, wenn es um sie selber geht, darf nicht vergessen lassen, daß sie in vielerlei Hinsicht über ihre eigenen Angelegenheiten mehr wissen können als externe Gutachter, die bei ihnen Stippvisiten machen. Man muß in den Evaluationsprozeß also ihre Vorstellungen über Ziele und Verpflichtungen, ihre Interpretation der eigenen Stärken und Schwächen sowie ihre Definition der Kriterien, mit denen ihre Arbeit gemessen werden kann, hineinnehmen. Dies nicht, um es unbesehen zu übernehmen, wohl aber, um es ernst zu nehmen.

Daß es ernst genommen werden kann, setzt allerdings voraus, daß die Institutionen der Wissenschaft, um deren Evaluation es geht, sich mit sich selbst dauerhaft kritisch auseinandersetzen, also Selbstevaluation routinisieren. Fremdevaluation sollte vor allem eine Art Metaevaluation sein können, nämlich Evaluation von Selbstevaluation. Für die Außenwahrnehmung ist letztlich nicht entscheidend, daß eine Wissenschaftseinrichtung mit ihrer Leistungsproduktion bestimmte Schwächen und Probleme hat; solche sind immer vorhanden. Entscheidend für die Qualität einer Wissenschaftseinrichtung ist, ob sie die eigenen Schwächen und Probleme selbst wahrnimmt, angemessen deutet und nachdrücklich bearbeitet. Wissenschaftspolitisch erscheint es in diesem Sinne als vorrangig, Universitäten und Forschungseinrichtungen auf Selbstevaluation zu verpflichten. Dennoch, Fremdevaluation wird und muß zu Konsequenzen führen, im unterstützenden und im begrenzenden Sinne. Um so mehr ist es wichtig, daß sie sich um Konsens mit den Betroffenen ernsthaft bemüht.

(3) Auch für die methodische Anlage von Evaluationen gilt, daß sie nicht nur so einfach (und ökonomisch) wie möglich, sondern auch so vielschichtig und komplex wie nötig sein muß. Um sie einfach (und vergleichbar) zu halten, bedarf es sicher der Standardisierung von möglichen Indikatoren der Leistungsmessung. Sofort ist hinzuzusetzen, daß jeder einzelne Indikator nur ein „Partialindikator“ sein kann. Die verschiedenen Seiten dessen, was „Qualität“ von Forschung (oder auch Lehre) darstellt, bedürfen zu ihrer angemessenen Erfassung eines mehr oder weniger umfangreichen „Indikatorensatzes“. Hinzu kommt nun, daß die Interpretation von Indikatoren und Indikatorensätzen vernünftiger Erwägungen darüber bedarf, was und wie viel diese von einer Wissenschaftseinrichtung angesichts ihrer speziellen Zwecksetzungen und ihrer besonderen Umstände tatsächlich valide messen. Zahlen sprechen nicht für sich selbst. Insofern ist auch die Entgegensetzung quantitativer

und qualitativer Bewertungen und Bewertungsmaßstäbe unhaltbar, wenn sie als konkurrierend angesehen wird. Beide kontrollieren sich und bedürfen einander, um ihre eigene Geltung abzusichern.

Von daher ist auch der Sinn von „peer review“ genauer zu bestimmen. Nicht nur bedürfen vorhandene Meßgrößen der kundigen Deutung durch „peers“. Man schützt diese vor ihren eigenen Befangenheiten, wenn man sie – wo immer es möglich ist – zwingt, sich mit „harten Daten“ auseinanderzusetzen.

(4) Uns ist auf der Akademiekonferenz von unseren englischen und niederländischen Kollegen berichtet worden, daß die in beiden Ländern schon seit längerem eingeführten Evaluationspraktiken nach mehreren Korrekturen erkennbar positive Effekte in der Wissenschaft ausgelöst haben und daß es darüber auch bei den Betroffenen ein beachtliches Maß an Übereinstimmungen gibt. Uns selbst hat allerdings eine Information über Evaluationsfolgen beunruhigt, die offenbar sowohl in den Niederlanden als auch in England beobachtbar sind. Berichtet wird, daß durch Evaluationen eine dezidiert monodisziplinäre Forschung deutlich privilegiert erscheint. Das wird damit zusammenhängen, daß sich interdisziplinäre Forschungsprozesse schlechter messen lassen und daß die Qualität ihrer Ergebnisse überdies von Gutachtern unterschätzt werden, die selber disziplinär eingebunden sind – zum Beispiel in den Universitäten, deren Binnenorganisation sehr stark nach Disziplinen segmentiert ist.

Im Hinblick auf ursprüngliche Ängste, Evaluationen würden Einfallstore für außerwissenschaftliche Interessen sein, mag man es nun einerseits beruhigend finden, daß sich in den Evaluationen tatsächlich eine Art „akademischer Bias“ durchsetzt. Andererseits muß man zur Kenntnis nehmen, daß dieser Bias leicht zur autistischen Selbstisolierung der Wissenschaft im allgemeinen und der wissenschaftlichen Disziplinen im besonderen führen kann. Nicht ohne Grund ist in Deutschland genau mit diesem Vorbehalt die Tendenz der Wissenschaftsratsevaluationen im Umgang mit den außeruniversitären Forschungseinrichtungen kritisiert worden, die ausdrücklich auf ein gewisses Maß an Anwendungsorientierung und – damit zusammenhängend – Interdisziplinarität verpflichtet sind. Nimmt man seine neuesten „Thesen zur künftigen Entwicklung des Wissenschaftssystems in Deutschland“ mit den Appellen zur „Stärkung von Anwendungsorientierung und Praxisbezug“ wahr, dann wird deutlich, daß dem Wissenschaftsrat an den ungleichgewichtigen Effekten auch der von ihm selbst durchgeführten Evaluationen nicht gelegen sein kann. Praktisch vordringlich sind Konventionen zur Bestimmung und Messung auch der außerwissenschaftlich wirksamen und dann in der Regel interdisziplinären Leistungen wissenschaftlicher Einrichtungen. Denn das gehört zu den Folgen einer auf Standardisierungen drängenden Evaluationspraxis: Wofür man keine Kriterien hat, das verliert Beachtung und verkümmert.

(5) Positive wie negative Folgen von Evaluationen sind langfristig allerdings nur zu erwarten, wenn aus dem Wissen, das über Evaluationen geschaffen wird, auch praktische Folgerungen gezogen werden. Dies betrifft zuerst (und vielleicht auch

vor allen Dingen) den Vorgang der Selbstevaluationen in den Wissenschaftseinrichtungen. Selbstevaluation wird zur Selbstkontrolle erst mit der Durchsetzung von Konsequenzen. Diese setzt aber robuste Leitungsstrukturen voraus, die notfalls auch gegen den Widerstand der Betroffenen entscheidungsfähig sind. In Wissenschaftseinrichtungen, in denen mancherlei für eine stark kollegial verfaßte Leitungsstruktur spricht, ist dies eine heikle Angelegenheit; aber man muß auch in dieser Hinsicht für Balancen Sorge tragen, welche die Evaluationen nicht zu einer folgenlosen Veranstaltung machen.

Das gilt ebenso für Fremdevaluationen, in denen letztlich auch staatliche Instanzen in die Rolle geraten, Konsequenzen ziehen zu müssen – und es gilt um so mehr, je weniger die Wissenschaft fähig ist, sich selbst zu steuern. Es ist jedenfalls bedauerlich, daß im föderalen System der Bundesrepublik besonders bei Bund/Ländergeförderten Forschungseinrichtungen mehrfach Sanktionsempfehlungen des Wissenschaftsrats, die sich auf aufwendig ermittelte Evaluationsergebnisse stützen konnten, politisch völlig ins Leere gelaufen sind. Spricht sich dergleichen herum, dann wird es nicht verwunderlich sein, daß in den Wissenschaftseinrichtungen das ganze Thema der Evaluationen nicht ernst genommen wird. Die Evaluierungen der Wissenschaft sind durch die Erwartung der Politik initiiert worden, daß die mit öffentlichen Mitteln geförderten Einrichtungen wie Universitäten und Forschungsinstitute rechenschaftspflichtig gegenüber der Öffentlichkeit seien. Die Legitimität dieser Erwartung ist unabweisbar. Mit ihr wird jedoch die Politik zugleich in die Verantwortung eingebunden, sich selbst an die Bedingungen zu halten, unter denen sie die Rechenschaftspflicht allererst einfordert. Letztlich geht es um die Etablierung einer Evaluationskultur zwischen Politik und Wissenschaft, in der die Zuweisung öffentlicher Ressourcen selbstverständlich mit Leistungserwartungen verknüpft wird und ebenso selbstverständlich die Ergebnisse von Leistungsbilanzen Grundlage von Verteilungsentscheidungen sind. Die Erwartungen müssen den Möglichkeiten der Wissenschaft entsprechen und die Leistungsbilanzen deren Erfüllung angemessen widerspiegeln. Eine solche Kultur ist die Grundlage eines neuen „contrat social“ für die Wissenschaft.