

Effekte nationaler Evaluationssysteme für die universitäre Forschung Können wir von angelsächsischen Beispielen lernen?

Stefan Lange
Speyer

Während der letzten drei Jahrzehnte ist die Leistungsbewertung der öffentlich finanzierten Forschung in den meisten Ländern des OECD-Raums zu einem politischen Problem geworden. Der implizite ‚*social contract*‘ zwischen Gesellschaft und Wissenschaft, wonach der Staat Forschung vorbehaltlos zu alimentieren hätte, erodiert zunehmend: Die Regierungen erwarten heute ‚*value for money*‘ von der Wissenschaft und reagieren mit der Schaffung entsprechender Rahmenbedingungen, etwa der in immer mehr Ländern erfolgten Einführung von systematischen Evaluationen der Forschungsperformanz ihrer Universitäten.

Seit Ende der 1990er Jahren wird auch in Deutschland mit solchen Instrumenten experimentiert. Viele Bundesländer haben formelbasierte Mittelverteilungssysteme für ihre Hochschulen eingeführt, die auch Indikatoren zur Bewertung der Forschungsleistung beinhalten. Auf der gesamtstaatlichen Ebene wird mit den Pilotstudien des Wissenschaftsrates zum Forschungsrating und mit der Exzellenzinitiative versucht, Leistung sichtbar zu machen und eine darauf bezogene Differenzierung in der Hochschullandschaft zu initiieren. Damit entsteht die Frage, ob sich Erfahrungen, die in anderen Ländern mit den dort schon länger etablierten

Evaluationssystemen gesammelt wurden, auf die deutsche Situation übertragen lassen. Das Ziel dieses Beitrags besteht darin, die Ursachen für intendierte und nichtintendierte Effekte in den Forschungsevaluationssystemen der angelsächsischen Vorreiterländer Großbritannien und Australien herauszuarbeiten und auf dieser Grundlage die Wahrscheinlichkeit für das Auftreten solcher Effekte in Deutschland abzuschätzen. Können wir von den angelsächsischen Systemen lernen?

1. Die Entstehung universitärer Forschungsevaluationssysteme in Deutschland

Nachdem der Wissenschaftsrat seit den 1970er Jahren anlassbezogen einzelne Forschungseinrichtungen ad hoc evaluiert hatte, kam es im Zuge der Deutschen Wiedervereinigung zu Beginn der 1990er Jahre zu einer ersten systematischen Forschungsevaluation bei den außeruniversitären Forschungseinrichtungen. Dies wirkte sich aber kaum auf die durch die Länderhoheit geprägte deutsche Universitätslandschaft aus. Mit einer Ausnahme (Niedersachsen) haben die Landesregierungen die an ‚ihren‘ Hochschulen durchgeführte Forschung nicht zum Gegenstand systematischer Evaluationen durch Peer review gemacht.

Diese Situation begann sich Ende der 1990er Jahre zu ändern, als viele Hochschulen mit der Novellierung des Hochschulrahmengesetzes und nachfolgenden neuen Landeshochschulgesetzen in eine begrenzte Autonomie entlassen wurden und mit unterschiedlicher Emphase auch ‚schwache‘ Evaluationssysteme für Forschungsleistungen einführten. Dazu lassen sich vor allem die Instrumente ‚Zielvereinbarungen‘ und ‚Leistungsorientierte Mittelvergabe‘ rechnen. Die Zielvereinbarungen zwischen einer Landesregierung und ihren Universitäten spezifizieren Innovationen in Forschung und Lehre und bieten den Hochschulen dafür eine gewisse Erwartungssicherheit hinsichtlich der Wahrung oder nur moderaten Verschlechterung des finanziellen Status quo (König 2007). Häufig zählt zu den in Aussicht gestellten ‚Innovationen‘ die Schaffung profilbildender Bereiche in Form von Forschungszentren und Graduiertenschulen. Die Zielvereinbarungen beinhalten nur in Ausnahmefällen die Bereitstellung zusätzlicher Mittel durch die Länder, unterliegen keiner strengen Kontrolle, und ihre Nichterfüllung wird kaum sanktioniert.

Die Finanzierungsformeln der Leistungsorientierten Mittelverteilung binden Teile der jährlichen Mittelzuweisungen an die Erfüllung von Leistungskriterien in Lehre und Forschung, die mit quantitativen Indikatoren

gemessen werden (Jaeger 2005). So begann etwa Nordrhein-Westfalen schon 1993 damit, den Hochschulen des Landes 10% der laufenden Mittel für Lehre und Forschung nach Leistungsindikatoren zuzuweisen. Heute werden in Nordrhein-Westfalen 20 % und in Rheinland-Pfalz sogar 80 % der laufenden Mittel nach Indikatoren vergeben. Insgesamt haben in Deutschland bislang 13 von 16 Bundesländern die Leistungsorientierte Mittelverteilung eingeführt (Lanzendorf/Pasternack 2008: 53-56). Die durch Indikatoren gesteuerten laufenden Mittel machen aber nur einen kleinen Teil der Hochschulfinanzen aus. Das Gros bilden nach wie vor die nicht in die indikatorbasierte Finanzierung einbezogenen Personalmittel. Wie die Hochschulleitungen das Geld aus dem Leistungsbudget des Landes intern weiterreichen, bleibt bislang weitgehend ihre Angelegenheit.

Bemühungen um bundesweite Evaluationssysteme zur Sichtbarmachung und Steigerung der Forschungsleistungen haben bislang zwei Ergebnisse gezeitigt. Der Wissenschaftsrat hat zwei Pilotstudien zum Forschungsrating durchgeführt, in denen die gesamte Forschung in den Fächern Chemie und Soziologie durch Expertenkommissionen evaluiert und in mehreren Dimensionen vergleichend bewertet wurde. Zwei weitere Pilotstudien dieser Art werden in einer Geisteswissenschaft sowie der Elektro- und Informationstechnik durchgeführt. Allerdings hat der Historikerverband mittlerweile eine Beteiligung abgelehnt. Die Forschungsratings haben keine Konsequenzen für die Finanzierung der Universitäten.

Unter Allokationsgesichtspunkten folgenreicher ist dagegen die gemeinsame Exzellenzinitiative des Bundes und der Länder, in der erhebliche zusätzliche Forschungsmittel für ‚Graduiertenschulen‘, ‚Exzellenzcluster‘ und ‚Zukunftskonzepte‘ vergeben werden. Die Mittel für die Exzellenzinitiative wurden entsprechend auf der Grundlage von Ex ante-Evaluationen durch rein wissenschaftlich besetzte internationale Gutachter-Panels als zeitlich befristete Projektmittel vergeben. Die geförderten Universitäten beziehungsweise Fachbereiche und Forschergruppen können mit den erheblichen staatlichen Sondermitteln eigenständig Strukturentwicklung betreiben (Sondermann et al. 2008). Exzellenzinitiativen unterscheiden sich von Evaluationssystemen und müssen als eigenständiges Governance-Instrument für die Forschung behandelt werden. Sie lassen die Grundfinanzierung der Universitäten unberührt und stellen – meist zeitlich befristet – zusätzliche Mittel für die selektive Steigerung von Forschungsleistung bereit.

2. Internationale Erfahrungen mit Forschungsevaluationssystemen

Großbritannien und Australien sind Vorreiter bei der weltweiten Etablierung von Systemen der Evaluation von Forschungsleistungen. Gleichwohl haben die Regierungen und Wissenschaftsadministrationen der beiden Länder deutlich voneinander abweichende Evaluationssysteme institutionalisiert, von denen sich einzelne Elemente auch in den deutschen Ansätzen zur Forschungsevaluation finden. Deshalb werden im Folgenden in Fallstudien für jedes der beiden Länder die Initialisierungsphase, Etappen der Weiterentwicklung und heutige Merkmale der Evaluationssysteme sowie die intendierten und nichtintendierten Effekte auf die dortigen Universitäten und Forscher aufgezeigt.

2.1. Großbritannien: *Research Assessment Exercise*

Entstehung und Entwicklung

In Großbritannien war die Einführung der nationalen *Research Selectivity Exercise* die Antwort auf wachsende Sorgen über eine allgemeine Degeneration der Forschungsinfrastruktur seit den späten 1970er Jahren. Der *Merrison Report* von 1979 dokumentierte das Ausmaß, in dem die Infrastruktur der Forschung an den Hochschulen erodiert war und argumentierte, dass weiterer Schaden nur dann abgewendet werden könne, wenn die zur Verfügung stehenden Ressourcen vom Staat künftig selektiv verteilt würden. Die erste *Research Selectivity Exercise* zur leistungsbezogenen Verteilung der Forschungsgrundausrüstungen wurde dann in den Jahren 1985/86 durchgeführt. Die Universitäten wurden aufgefordert, Informationen über die Einwerbungen von Forschungsdrittmitteln und die Zahl ihrer Forschungsstudenten (Master/ PhD) einzureichen. Zusätzlich erbat den nationalen Gutachter-Panels von den Universitäten pro Department fünf Publikationen, die die Universitäten als repräsentativ für die dort betriebene disziplinäre Forschung ansahen.

Die der ersten Evaluationsrunde folgenden Transformationen lassen sich als schrittweise Verfeinerung und Professionalisierung des Ausgangssystems charakterisieren (Barker 2007). Die seit der ersten Evaluation konstanten Hauptmerkmale sind dabei das zentralisierte, vergleichende Verfahren in der Verantwortung der öffentlichen Zuwendungsgeber (der *Higher Education Funding Councils* für England, Schottland und Wales), die Verwendung des Peer Review als Evaluationsmethode und die Nut-

zung der Evaluationsergebnisse für die selektive Vergabe von Forschungsgrundausrüstungen an die Universitäten. Die Hauptveränderungen seit der ersten Durchführung sind:

1. Die Einführung einer gemeinsamen Evaluationsskala für alle Disziplinen und die Bewertung der Publikationen von potentiell allen Wissenschaftlern statt jeweils einer von den Departments bestimmten Auswahl.¹ Beide Veränderungen wurden bereits mit dem zweiten Durchgang im Jahre 1989 vollzogen, in dem auch die *Research Selectivity Exercise* in *Research Assessment Exercise* umgetauft wurde.
2. Wurden im ersten und zweiten Durchgang noch alle Wissenschaftler eines Departments in die Evaluation einbezogen, so waren es ab 1992 nur noch diejenigen, die den Panels von den Universitäten als ‚forschungsaktiv‘ gemeldet wurden.
3. Mit der 1996er *Research Assessment Exercise* wurden die wissenschaftlichen Gemeinschaften über ihre Fachgesellschaften an der Auswahl der Gutachter beteiligt. Gleichzeitig wurden erstmals die Evaluationskriterien der Panels vor dem Beginn der *Exercise* veröffentlicht, so dass die Universitäten sich bei der Nominierung der forschungsaktiven Wissenschaftler und diese bei der Auswahl ihrer einzureichenden Publikation bzw. der Planung ihrer Publikationsstrategien vorab darauf einstellen konnten.
4. In der aktuellsten Variante von 2008 wurden die diskreten Rankings in sechs Kategorien (von 5* = weltweit führend bis 1 = auch im nationalen Vergleich unbedeutend) durch ‚Qualitätsprofile‘ ersetzt, die nun die Verteilung der Forschungsqualität einer Einheit (in der Regel eines Departments) über die Qualitätsskala widerspiegeln, ohne – wie früher – für die Evaluationseinheit zu einer Art Gesamtnote zu kommen.

Darüber hinaus wurden die Evaluationszyklen von ursprünglich drei Jahren im Laufe der Zeit auf zuletzt über sieben Jahre ausgedehnt. Die

¹ Die Evaluationsskala setzte sich bis zu ihrer Ersetzung durch Qualitätsprofile in 2008 aus sieben Globalnoten zusammen, wobei die Noten 5* und 5 für international führende Forschung vergeben wurden und die höchsten monetären Ausschüttung bedeuteten, während für die Noten 3a, 2 und 1 in der Regel gar keine Ressourcen verteilt wurden. Im Mittelfeld der zumindest auf nationalem Niveau guten Forschung (4 und 3b) gab es nur geringe Beträge. Die Qualitätsbewertung von 250.000 wissenschaftlichen Erzeugnisse und Kontextdaten (zum Beispiel zu Drittmitteln, PhD-Studenten, Stipendiaten) von 56.000 durch die Universitäten nominierten Wissenschaftlern erfolgte im *Research Assessment Exercise* von 2001 durch 68 disziplinäre Gutachterpanels, die insgesamt 685 Peers zusammenfassten (Roberts 2003).

Research Assessment Exercise war von Beginn an ein in ständiger Entwicklung begriffenes System, dessen kritische Diskussion immer wieder zu politischen Entscheidungen über Veränderungen führte. Alle Reformen lassen sich maßgeblich auf die Vorschläge von Stakeholdern der Research Assessment Exercise zurückführen. Die Gesamtausrichtung des Evaluationsystems auf selektive Forschungsförderung wurde indes nie grundsätzlich in Zweifel gezogen.

Effekte

Die britischen Universitäten verteilen das Gros des aus der Research Assessment Exercise eingeworbenen Forschungshaushalts nach Kriterien des Verdienstes: Die Departments, die durch ihr gutes Abschneiden zum Einkommen ihrer Universität beitragen, bekommen einen großen Teil des erwirtschafteten Geldes, während die anderen in der Regel leer ausgehen. Die zentralen Universitätsleitungen behalten jedoch einen Teil des durch die Research Assessment Exercise vereinnahmten Geldes ein, um sich selbst einen Ressourcenpool für strategische Entscheidungen (Berufungen, Profil- und Zentrenbildung etc.) zu schaffen. Innerhalb der Departments vieler britischer Universitäten werden die Ressourcen – inklusive des Zeitbudgets für die Forschung – von denen, die nicht als ‚forschungsaktiv‘ gemeldet werden können, umverteilt zu denen, die man auch für die nächste Research Assessment Exercise zu nominieren gedenkt (Lucas 2006: 95-132).

Strategische Investitionen sind ein Hauptmerkmal des Forschungsmanagements britischer Universitäten. Internationale Bekanntheit haben sich solche strategischen Managemententscheidungen in Form von Schließungen oder Restrukturierungen von Departments erworben, die darauf abzielten, fachliche Einheiten zu kreieren oder zu stärken, die in kommenden Evaluationsrunden mutmaßlich (noch) bessere Bewertungschancen hätten. Auch das *head-hunting* auf exzellente Einzelforscher, mit denen eine Universität sowohl die eigene Forschungskultur (zusätzlich) stärken als auch gleichzeitig ein gutes Abschneiden in der nächsten Research Assessment Exercise sichern möchte, ist inzwischen zur allgemeinen Praxis geworden.

Das stärkste Argument für positiv-intendierte Effekte des britischen Evaluationssystems lässt sich auf die Mehrheitsmeinung der Universitätsmanager wie auch der Wissenschaftler selbst zurückführen, der zufolge die Forschung qualitativ tatsächlich besser geworden ist. Dies wird auf die durch die Research Assessment Exercise ausgelöste erhöhte Aufmerk-

samkeit und Unterstützung für Fragen der Forschung, die verbesserte materielle Ausstattung an den erfolgreichen Standorten sowie die nach Forschungsgesichtspunkten erfolgten Strukturreformen an manchen Universitäten zurückgeführt.

Negative Effekte der Research Assessment Exercise können einerseits dem verwendeten Evaluationsverfahren und andererseits den Reaktionen der Universitäten zugeschrieben werden. Der der Research Assessment Exercise zugrunde liegende Peer review bevorzugt den Mainstream gegenüber unkonventioneller und riskanter, in ihren Ergebnissen weniger vorhersehbarer Forschung, disziplinäre gegenüber interdisziplinärer Forschung und an die Fachgemeinschaften gerichtete Grundlagenforschung gegenüber der auf außerwissenschaftliche Adressatenkreise zugeschnittenen anwendungsorientierten Forschung (Henkel 2000). Die Wissenschaftler antizipieren dies in ihrer Wahl von Forschungsthemen, wodurch epistemische Veränderungen in der Forschung ausgelöst werden können (Morris 2004). Auch die Universitäten antizipieren solche Entscheidungstendenzen und berücksichtigen sie in ihren Rekrutierungsstrategien. Sie konzentrieren ihre Aufmerksamkeit und ihre Ressourcen auf den evaluierten und mit Anreizen gesteuerten Prozess, was zu einer relativen Vernachlässigung der weniger intensiv und folgenreich evaluierten Aktivitäten führt. Neben der anwendungsorientierten Forschung, deren Ergebnisse insbesondere für das regionale Innovationsumfeld der Universitäten und die Ausbildung von Studenten für den nationalen oder regionalen Arbeitsmarkt wichtig sind, betrifft dies insbesondere die Lehre.

2.2. *Australien: Evaluationsbasierte Indikatorensteuerung*

Entstehung und Entwicklung

Die Australische Regierung hatte Ende der 1980er Jahre in einer umfassenden nationalen Reform ihr zweigliedriges Hochschulsystem abgeschafft und alle Hochschulen im Zuge von Höherstufungen oder Fusionen in Universitätsrang erhoben (Marginson/Considine 2000). In ihrem *White Paper on Higher Education* erklärte die australische Bundesregierung ihre Absicht, ein Finanzierungssystem entwickeln zu wollen, dass die Performanz der Universitäten berücksichtige. Im Jahre 1990 wurde das *Relative Funding Model* eingeführt, dass in der Grundfinanzierung der Universitäten eine lehr- und eine forschungsbezogene Komponente unterschied. Letztere umfasste die Infrastrukturkosten für die intern finanzierte Forschung und für die extern, aber nicht vollständig kostendeckend finan-

zierte Forschung. Die Forschungskomponente machte zunächst 6% der Grundfinanzierung aus und wurde auf der Basis der kompetitiv eingeworbenen Drittmittel vergeben.

In den Folgejahren wurde an der Nutzung quantitativer Indikatoren festgehalten. Seit 1996 wurden zusätzliche Indikatoren zur Leistungsmessung herangezogen, und zwar die Zahl der Publikationen, die Abschlüsse (Master und PhD) und die Zahl der immatrikulierten Forschungsstudenten. Bei der Erhebung der Publikationszahlen wurde zunächst mit insgesamt 22 Kategorien operiert, was sich als nicht praktikabel erwies. Deshalb sind seit 1996 nur noch vier Kategorien von Publikationen für die Finanzierung relevant, und zwar begutachtete Zeitschriftenaufsätze, begutachtete Konferenzpapiere, Bücher sowie Buchkapitel.

Die Gewichte der einzelnen Indikatoren wurden mehrmals verändert. Der Anteil der Forschungskomponente an der Grundfinanzierung wuchs in 2005 auf mehr als 20%. Da aber mit der Einführung von Studiengebühren der Anteil der Grundfinanzierung an den Einnahmen der Universitäten auf 40% zurückging, liegt der Anteil der evaluationsbasierten Forschungsfinanzierung an den Einnahmen der Universitäten heute bei lediglich 7,9% (Gläser et al. 2008: 148).

Effekte

Die australischen Hochschulen reagieren auf ein Evaluationssystem, das ihre Grundfinanzierung vor allem an die eingeworbenen Drittmittel bindet, mit der Strategie, ihre interne Forschungsfinanzierung, strukturelle Maßnahmen und das Management individueller Forschungsleistungen auf die Steigerung der Einnahmen aus Drittmitteln auszurichten. Die beiden anderen Indikatoren – Publikationen und die Betreuung von Doktoranden – spielen demgegenüber eine untergeordnete Rolle, werden aber bei Beförderungen als Kriterium herangezogen (Gläser et al. 2008). Konzentration der Anstrengungen auf die Steigerung der Drittmiteleinwerbung heißt vor allem, die knappen Mittel für die Grundfinanzierung der Forschung in ‚profit centres‘ zu investieren, das heißt in Forschungszentren, die gute Aussichten haben, Drittmittel einzuwerben und sich idealiter nach einer Anschubfinanzierung selbst aus Drittmitteln tragen. Dasselbe Prinzip wird bei der Ausstattung einzelner Wissenschaftler angewendet: Wissenschaftler, die bei der Drittmiteleinwerbung knapp gescheitert sind, neu eingestellte Wissenschaftler und Nachwuchswissenschaftler können sich bei den Universitäten um Mittel bewerben, mit denen sie Drittmittelanträge vorbereiten. Die Fokussierung des Universitätsmanagements auf

die Drittmittelinwerbung kann als „goal displacement“ (Blau/Scott 1962: 229) – im Folgenden frei übersetzt als „Zielaustausch“ – beschrieben werden: Nicht die Erhöhung der Forschungsqualität, sondern die Steigerung der Drittmittelleinnahmen steht im Zentrum der Aufmerksamkeit. Jenseits intern kompetitiv vergebener Anschubfinanzierungen gibt es keine Grundausstattung für die Forschung oberhalb einer basalen Infrastruktur (Gläser/Laudel 2007).

In Australien kann – wie in Großbritannien auch – eine gewachsene Aufmerksamkeit und Unterstützung der Forschung durch die Universitäten als positiver und intendierter Effekt des Evaluationssystems angenommen werden. Auch die Umverteilung von Ressourcen auf die (in den verwendeten Indikatoren) leistungsstärksten Forscher trägt vermutlich zur Qualitätssteigerung bei.

Negative Effekte entstehen aus der exklusiven Orientierung der Universitäten auf die wenigen Indikatoren der Forschungsleistung und aus der generellen Mittelknappheit, die die gesamte universitäre Forschung zwingen, sich an den der Drittmittellandschaft eingeschriebenen Prioritäten zu orientieren. Diese Drittmittellandschaft lässt sich im Vergleich zur deutschen als *monopolistisch*, *karg* und *biased* beschreiben (Laudel 2006). Australische Wissenschaftler sind auf eine einzige bedeutende Drittmittelquelle angewiesen: den direkt dem Wissenschaftsministerium unterstehenden *Australian Research Council* beziehungsweise – für medizinische Forschung – den *National Health and Medical Research Council*. Andere Quellen wie die australische Industrie oder Förderprogramme der Regierung spielen praktisch keine Rolle.

Alle durch die Forschungsräte geförderten Projekte müssen ihren nationalen Nutzen ausweisen, der mit einem Gewicht von 10% in die Bewertung der Projektanträge eingeht. Unter den Förderprogrammen des Australian Research Council hat die Förderung von Kooperationsprojekten mit Industriepartnern deutlich höhere Bewilligungsraten (40-50%) als die Förderung im ‚Normalverfahren‘ (20-30%). Dieser Bias in der Drittmittellandschaft erzeugt einen starken Druck auf die Durchführung anwendungsorientierter Forschung.

Die Anpassung der Universitäten und der Wissenschaftler an diese Situation verändert die Forschung. Erstens kommt es zu einer Umverteilung von Forschungsmöglichkeiten durch entsprechende Maßnahmen der Universitäten und einen Matthäus-Effekt in der Drittmittelinwerbung. Zweitens hat die Orientierung australischer Forscher an den Prioritäten der Drittmittellandschaft eine stärkere Anwendungsorientierung der For-

schung und einen Trend zum Mainstream zur Folge. Letztlich verengen sich die individuellen Forschungsportfolios – die Wissenschaftler bearbeiten tendenziell weniger und einander ähnelndere Themen (Gläser/Lange/Laudel 2009).

3. Abschätzung von Effekten der Evaluationssysteme in Deutschland

Während es noch zu früh ist, um Effekte der deutschen Evaluationssysteme zu beurteilen, lassen sich aus dem Vergleich mit den angelsächsischen Referenzsystemen doch einige Schlussfolgerungen zu ihrem Innovationspotenzial und zu erwartbaren Nebenfolgen der Instrumente ziehen. Alle drei in Deutschland praktizierten Systeme – die beiden bundesweiten Instrumente Forschungsrating des Wissenschaftsrates und Exzellenzinitiative sowie die im Rahmen der Hochschulfinanzierung der Länder praktizierte Leistungsorientierte Mittelverteilung – liefern den forschungspolitischen Akteuren, dem Universitätsmanagement und auch den Wissenschaftlern selbst Informationen über die Leistungsfähigkeit der Forschung an den Universitäten. Auch die Konstitution einer breiteren Öffentlichkeit für Belange der Forschung und die Sichtbarmachung vermeintlich ‚exzellenter‘ Forschungsstandorte können – ähnlich wie in Großbritannien – zumindest als Nebeneffekte der Ratings des Wissenschaftsrates, besonders aber der Exzellenzinitiative, gelten. Die Einführung der Leistungsorientierten Mittelverteilung hat – ähnlich wie in Australien – vor allem hochschulinterne Effekte.

Vergleicht man nun konkret die deutschen Systeme und Instrumente zur Forschungsevaluation mit den beiden Vorreitern, dann zeigen sich Ähnlichkeiten, die Prognosen gestatten. Die Leistungsorientierte Mittelverteilung ist die schwache Version eines indikatorbasierten kompetitiven Systems, die sich von der australischen Variante nur durch die verwendeten Indikatoren und die geringeren finanziellen Konsequenzen unterscheidet.

Der australische Fall demonstriert einen generellen und einen spezifischen Effekt der Indikatorensteuerung. Der generelle Effekt ist ein Zieldaustausch, das heißt die Fokussierung einer Steigerung der Forschungsleistung auf die Indikatoren selbst statt auf die Qualität, die die Indikatoren eigentlich messen sollten. Da Publikationen, Drittmittelwerbung und die Zahl der Doktoranden als Indikatoren der Forschungsleistung von fragwürdiger Validität sind, entsteht eine erhebliche Differenz zwischen

politischer Absicht und dem realen Verhalten der Universitäten. Die wegen der Bedeutung des Indikators für die staatlichen Zuwendungen erfolgende Fokussierung auf die Drittmittelinwerbung hat vor allem dann negative Auswirkungen auf die Forschungsinhalte, wenn die Drittmittellandschaft durch ein Monopol oder auch ein Oligopol weniger, staatlich kontrollierter Förderagenturen mit identischen Präferenzen dominiert wird.

Anders als in Australien ist die deutsche Drittmittellandschaft allerdings *pluralistisch, reich und neutral*, das heißt es gibt eine Vielzahl von öffentlich finanzierten und privaten Förderagenturen und Stiftungen, die in ihren Förderprogrammen bislang keinen inhaltlichen Einfluss auf die Forschungsagenden der individuellen Wissenschaftler nehmen. Dem in Australien von der Regierung erzwungenen Anwendungsbias könnte zukünftig in Deutschland jedoch ein Bias zugunsten der Grundlagenforschung entsprechen. Interviews im Rahmen von Fallstudien an drei deutschen Universitäten zeigen einen Trend zur Bevorzugung von DFG-Drittmitteln als ‚höherwertig‘. In den drei Universitäten wurden DFG-Drittmittel von der Hochschulleitung signifikant höher gewertet als Einwerbungen aus anderen Quellen. Wer nicht als ‚DFG-fähig‘ gilt, büßt an Reputation und Rückhalt in seinem Fachbereich und gegenüber der Hochschulleitung ein – was zum Beispiel in den Ingenieurwissenschaften bereits Probleme erzeugt. Der Bias zugunsten der Grundlagenforschung zeigt sich auch in der Berufungspolitik. Über solche indirekten Effekte indikatorengestützter Evaluationssysteme (und hier vor allem des Drittmittelfaktors) hinaus, werden die Folgen der Indikatorensteuerung für das Forschungshandeln und die Forschungsqualität so lange gering bleiben, wie die Pluralität der Drittmittellandschaft und thematisch neutrale Instrumente wie das DFG-Normalverfahren erhalten bleiben.

Die gegenwärtig durch den Wissenschaftsrat erprobten Forschungsratings ähneln vom Verfahren her dem Forschungsevaluationssystem in Großbritannien. Freilich kann man mit Blick auf die bisherigen Pilotstudien in Deutschland kaum von einem bereits institutionalisierten System sprechen. Die Research Assessment Exercise neigt zu einer Prä-Determinierung des Umgangs mit Informationen über Forschungsqualität, da es die hier vollzogenen Ratings unmittelbar mit einem staatlich vorgegebenen Verteilungsmechanismus verknüpft, den die Universitätsleitungen in der Regel unverändert an ihre Departments weiterreichen. Würden die Forschungsratings des Wissenschaftsrates in ihrer gegenwärtigen Form institutionalisiert, könnten sie voraussichtlich die Handlungsspielräume des Hochschulmanagements erweitern, aber – da sie nicht mit Mittelver-

teilungen verkoppelt sind – wenig zu einer Differenzierung der deutschen Hochschullandschaft beitragen.

Stärker noch als die Forschungsratings des Wissenschaftsrates weist die Exzellenzinitiative Ähnlichkeiten mit der britischen Research Assessment Exercise auf. Beide Verfahren beruhen auf Peer review und haben erhebliche finanzielle Konsequenzen für die Universitäten. Die Unterschiede zwischen beiden Instrumenten sind aber dennoch erheblich: Die Research Assessment Exercise steuert dauerhaft die Grundfinanzierung der Universitäten, während die Exzellenzinitiative zeitlich befristet zusätzliche Grundausstattung vergibt.

Wegen der finanziellen Konsequenzen der Exzellenzinitiative hat diese – ähnlich, aber schwächer als die Research Assessment Exercise – das Potenzial, die Differenzierung der Hochschullandschaft voranzutreiben. Die aufgewendeten erheblichen Mittel gestatten strukturelle Innovationen, verbessern die Ausstattung der Forschung in einigen Universitäten erheblich und leiten damit eine Ausdifferenzierung der deutschen Hochschullandschaft ein, deren Folgen mit Blick auf eine dauerhafte Segmentierung in forschungsstarke und -schwache Standorte einerseits, den Erhalt von Anschlussmöglichkeiten für zur Zeit nicht exzellente Standorte andererseits, noch nicht absehbar sind. Ein wichtiger Aspekt dieser Ausdifferenzierung könnte die Umverteilung der Lehre zwischen und innerhalb von Universitäten sein.

Ein zweiter Effekt, den die Analyse der Research Assessment Exercise für Deutschland erwarten lässt, ist die relative Vernachlässigung aller Aufgaben, die nicht mit gleicher öffentlicher Aufmerksamkeit und finanzieller Konsequenz evaluiert werden wie die Forschung. Ein auch nach mehr als 20 Jahren noch immer nicht gelöstes Problem liegt in dem Mangel an wirksamen komplementären Anreizstrukturen, die eine Vernachlässigung von Aufgaben in der angewandten Forschung, Lehre und Weiterbildung verhindern könnten.

4. Fazit: Was können wir lernen?

Aus dem hier angestellten Vergleich mit den angelsächsischen Ländern lassen sich nun für die Entwicklung in Deutschland erste Schlussfolgerungen ziehen:

1. Eine erste Beobachtung ist, dass Forschungsevaluationssysteme lernende Systeme sind, die sich weiterentwickeln, an neue Bedingungen anpassen und ihre Methoden ausdifferenzieren. Für diesen Lernpro-

zess bedürfen sie ihrerseits einer ständigen Evaluation. Befristete ‚Initiativen‘, gleich welcher Art, sind für solche Lernprozesse weniger geeignet als dauerhaft institutionalisierte Systeme.

2. Bei den Evaluationsverfahren ist das wesentlich aufwändigere qualitative Peer review, wie es im britischen Research Assessment Exercise angewendet wird, den in Forschungsevaluationssystemen wie dem australischen eingesetzten einfachen quantitativen Indikatoren überlegen. Die australischen Erfahrungen zeigen, dass solche Systeme zu einem Zielaustausch führen, bei dem das gute Abschneiden in den numerischen Werten, das Ziel der Steigerung der Forschungsqualität ersetzt. Solche Effekte deuten sich bei der deutschen leistungsorientierten Mittelvergabe bereits an, wenn die Drittmittelinwerbung zu einem dominierenden Kriterium bei Neuberufungen und der leistungsbezogenen Vergabe der Grundausrüstungen wird.
3. Eine wichtige positive Eigenschaft von Instrumenten zur Sichtbarmachung und Steigerung der Forschungsleistungen besteht darin, dass sie die Aufmerksamkeit von Universitäten und Wissenschaftlern auf die Forschung fokussieren. Diese Eigenschaft ist aber zugleich hochproblematisch, da sie in einem komplexen System von Leistungsprozessen die Prioritäten zugunsten der evaluierten und belohnten Aktivitäten verschieben. In Großbritannien und in Australien zeigt sich deutlich, dass die starken Evaluationssysteme zu einer relativen Vernachlässigung derjenigen Leistungsprozesse führen, die nicht gleichermaßen gut belohnt werden – nicht-drittmittelträchtige Forschung in Australien, anwendungsorientierte Forschung in Großbritannien sowie die Lehre in beiden Ländern. Dieser Nebeneffekt lässt sich nicht durch verbesserte Konstruktionen der Instrumente beseitigen, weil er die Balance zwischen evaluierten und belohnten Leistungsprozessen einerseits und nicht evaluierten oder belohnten Leistungsprozessen andererseits betrifft. Es bedarf analoger performanzorientierter Governance-Instrumente für die anderen Leistungsprozesse, die dann auch proportional angemessene Finanzvolumina steuern müssten. Mit Blick auf die Anreizstrukturen, die ein 10 Mio. € starker „Wettbewerb exzellente Lehre“ von Stifterverband und Kultusministerkonferenz verglichen mit den um ein vielfaches höheren Summen der Bundes- und Länderinitiativen für exzellente Forschung setzt, herrscht eine zu eindeutige Asymmetrie, die rationale Entscheidungsträger in den Universitätsleitungen ihr Heil primär in Verbesserungen der (DFG-fähigen) Forschungsleistung suchen lässt.

4. Instrumente zur Steigerung der Forschungsperformanz sollen idealiter die Forschungsinhalte nicht beeinflussen, sondern lediglich die Qualität der Forschung verbessern. Die Studien zur britischen Research Assessment Exercise und zum australischen Forschungsevaluationssystem zeigen jedoch, dass dieses Ziel verfehlt wird. Anwendungsorientierte Forschung (in Großbritannien), riskante Forschung und non-konformistische Forschung – letztlich all jene Forschungen, die keine Drittmittel attrahieren und deren Ergebnisse keinen oder nur geringen Niederschlag in international führenden Fachzeitschriften finden – gefährden in mit Verteilungsentscheidungen verknüpften Evaluationssystemen die Einnahmen der Universitäten und die Karrieren der Forscher. Im Mittelpunkt des Interesses stehen naturgemäß die ‚Milchkühe‘, die den Universitäten das meiste Geld bringen. Die Bereitschaft, reine Kostgänger loswerden zu wollen, steigt entsprechend. Auf eine solche Entwicklung reagieren auch die Forscher selbst mit opportunistischem Verhalten. Im Aggregateffekt kann das auf eine irreversible Einschränkung des Diversitätspools der an Universitäten betriebenen Forschung hinauslaufen. Ein möglicher Verlust von Diversität im Sinne nicht wahrgenommener Forschungsoptionen ist allerdings schwer messbar und damit als Argument weniger schwerwiegend.

Damit ist abschließend ein Grunddilemma jeder Evaluation von Forschung und Forschungsleistung angesprochen, das von Forschungspolitik und Wissenschaftsmanagement gleichwohl berücksichtigt werden sollte: Evaluationssysteme und Leistungsbeurteilungen führen häufig zu einer starken Homogenisierung der institutionellen Umwelt von Forschung, die auch die Forschung selbst einem starken Homogenisierungsdruck aussetzt. Man sollte aber – frei nach Niklas Luhmann – nicht vergessen, dass Universitäten die gesellschaftliche Funktion haben, Komplexität zu erhöhen und nicht, sie zu reduzieren.

Literatur

- Barker, Katharine, 2007: The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System. In: *Research Evaluation* 16, 3-12.
- Blau, Peter M./ Scott, Richard W., 1962: *Formal Organizations*. San Francisco: Chandler.
- Gläser, Jochen et al., 2008: Evaluationsbasierte Forschungsfinanzierung und ihre Folgen. In: Mayntz, Renate et al. (Hrsg.), *Wissensproduktion und Wissenstransfer. Wissen im Spannungsfeld von Wissenschaft, Politik und Öffentlichkeit*. Bielefeld: Transcript, 145-170.

- Gläser, Jochen/ Lange, Stefan/ Laudel, Grit, 2009: Auswirkungen der evaluati-
onsbasierten Forschungsfinanzierung an Universitäten auf die Inhalte der For-
schung. Australien und Deutschland im Vergleich. In: *Wissenschaftsrecht* 42,
329-352.
- Gläser, Jochen/ Laudel, Grit, 2007: Evaluation without Evaluators: The Impact of
Funding Formulae on Australian University Research. In: Whitley, Richard/
Gläser, Jochen (eds.), *The Changing Governance of the Sciences: The Advent
of Research Evaluation Systems*. Dordrecht: Springer, 127-151.
- Henkel, Mary, 2000: *Academic Identities and Policy Change in Higher Education*.
London: Jessica Kingsley.
- Jaeger, Michael, 2005: Leistungsbezogene Mittelvergabe und Qualitätssiche-
rung als Elemente der hochschulinternen Steuerung, HIS-Kurzinformationen
A/12/2005, Hannover: HIS.
- König, Karsten, 2007: Kooperation wagen. 10 Jahre Hochschulsteuerung durch
vertragsförmige Vereinbarungen, HoF-Arbeitsbericht 1/2007, Wittenberg: HoF.
- Lanzendorf, Ute/ Pasternack, Peer, 2008: Landeshochschulpolitiken. In: Hilde-
brandt, Achim/ Wolf, Frieder (Hrsg.), *Die Politik der Bundesländer. Staatstätig-
keit im Vergleich*, Wiesbaden: VS, 43-66.
- Laudel, Grit, 2006: The Art of Getting Funded: How Scientists Adapt to their
Funding Conditions. In: *Science and Public Policy* 33, 489-504.
- Lucas, Lisa, 2006: *The Research Game in Academic Life*. Maidenhead: SRHE/
Open University Press.
- Marginson, Simon/ Considine, Mark, 2000: *The Enterprise University. Power,
Governance and Reinvention in Australia*, Cambridge: Cambridge University
Press.
- Morris, Norma, 2004: *Scientists Responding to Science Policy. A Multi-Level
Analysis of the Situation of Life Scientists in the UK*, Enschede: CHEPS.
- Roberts, Sir Gareth, 2003. *Review of Research Assessment*. London: Higher Edu-
cation Funding Council for England, Quelle: <http://www.ra-review.ac.uk/reports/roberts.asp> (Zugriff 22.05.2009).
- Sondermann, Michael et al., 2008: *Die Exzellenzinitiative: Beobachtungen aus der
Implementierungsphase*. IFQ-Working Paper 5, Bonn: IFQ.