

Performanzsteigerung durch Selektivität? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen

Lange, Stefan; Gläser, Jochen

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

Verlag Barbara Budrich

Empfohlene Zitierung / Suggested Citation:

Lange, S., & Gläser, J. (2009). Performanzsteigerung durch Selektivität? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen. *der moderne staat - dms: Zeitschrift für Public Policy, Recht und Management*, 2(2), 411-432. <https://doi.org/10.3224/dms.v2i2.09>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/3.0>

Stefan Lange/Jochen Gläser

Performanzsteigerung durch Selektivität? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen

Zusammenfassung

Während der vergangenen drei Jahrzehnte hat sich die Bewertung der Performanz öffentlich alimentierter Forschung zu einem wichtigen Thema der Wissenschaftspolitik in den OECD-Ländern entwickelt. Viele Länder haben Forschungsevaluationssysteme eingeführt, um die Qualität der Forschungs-Outputs an ihren Universitäten zu messen, zu beurteilen und in Rangskalen zu listen. Die entsprechenden Entwicklungen in Deutschland sind durch einen späten Beginn und ausgeprägte föderale Fragmentierung gekennzeichnet. Der Beitrag beschreibt zunächst die jüngsten deutschen Ansätze zur Einführung mehrerer Forschungsevaluationssysteme auf verschiedenen Politikebenen. Danach werden die Vorreiter-Evaluationssysteme Großbritanniens, der Niederlande und Australiens untersucht und mit Blick auf ihre Effekte auf Universitätsmanagement sowie Forschung und Lehre diskutiert. Dabei interessiert insbesondere die Frage, inwiefern intendierte und nicht-intendierte Effekte dieser Evaluationssysteme für die Einführung und aktuelle Entwicklung der deutschen Ansätze instruktiv sind. Es zeigt sich, dass die deutschen Systeme der Evaluation von Forschungsperformanz zahlreiche Ähnlichkeiten mit denen der Vorreiter-Länder haben und vor allem von den negativen und nicht-intendierten Effekten lernen können, die sich dort in einer über zwanzigjährigen Evaluationspraxis und deren Folgen gezeigt haben.

Schlagerworte: Evaluation, Governance, Forschungsfinanzierung, Wissenschaftspolitik, Hochschulsysteme

Abstract

Performance by Selectivity? On Foreseeable Effects of Research Evaluations at German Universities in the Light of International Experiences

During the past three decades the performance assessment of state-financed research has become a major issue of science policies in OECD countries. Accordingly, many countries started to introduce research evaluation systems (RES) to assess and reward the performance of research in their universities. Germany is a latecomer in this respect. Owing to its federal constitution, a multitude of RES has emerged. This article aims at drawing conclusions about likely intended and unintended effects of the recently introduced German RES from an analysis of first-mover-RES in Great Britain, the Netherlands, and Australia. The article describes recent German attempts to establish several RES at different policy levels, and summarizes effects of the British, Dutch and Australian RES on university management as well as teaching and research. It turns out that German RES bear certain similarities with the first-mover-systems, which enables conclusions concerning the likelihood of negative effects which have shown up in more than twenty years of evaluations abroad.

Key words: Evaluation, Governance, Research Funding, Science Policy, Higher Education Systems

1 Einleitung

Während der letzten drei Jahrzehnte ist die Leistungsbewertung der öffentlich finanzierten Forschung in den meisten Ländern des OECD-Raums zu einem politischen Problem geworden. Bedingt durch die Expansion der nationalen Wissenschaftssysteme entwickelten sich die staatlichen Forschungsbudgets in den 1960er und 1970er Jahren zu wichtigen staatlichen Haushaltspositionen, die vermehrt die Aufmerksamkeit von Politik und Öffentlichkeit auf sich zogen. Spätestens zu Beginn der 1980er Jahre wurde den wissenschaftspolitischen Akteuren dann deutlich, dass sie zu wenig eigenes Wissen über die Leistungsprozesse der Erzeugung und Verwertung von Wissenschaft hatten, dass es keine besonderen Anreize für exzellente Leistungen bzw. für Leistungssteigerungen gab, und dass man die politisch gewollte Expansion der Wissenschaftssysteme der vergangenen Jahrzehnte unter den gegebenen Rahmenbedingungen schrumpfender Verteilungsspielräume der öffentlichen Hand nicht mehr auskömmlich finanzieren konnte. Der implizite Gesellschaftsvertrag mit der Wissenschaft, wonach der Staat Forschung vorbehaltlos zu alimentieren hatte, erodierte. Regierungen erwarten heute ‚value for money‘ von der Wissenschaft und sind geneigt, die Entscheidungen darüber, wie solcher Wert erzeugt werden könne, selbst durch das Setzen veränderter Rahmenbedingungen für die Forschung zu präformieren.

Zu diesem Zweck haben immer mehr Länder systematische Evaluationen der Forschungsperformanz ihrer Universitäten eingeführt. Evaluationssysteme, die Leistungserhebungen, -messungen und -beurteilungen für die Umverteilung der Grundfinanzierung universitärer Forschung nutzen, setzen i.d.R. auf eine Konzentration der knappen Mittel und auf finanzielle Anreize für die Universitäten, ihre Forschungsperformanz zu steigern. Alle Systeme zielen darüber hinaus darauf ab, die Universitäten über ihre Forschungsleistungen zu informieren und ein universitäres Management der Forschungsperformanz zu initiieren.

Damit eröffnet sich für die Forschung zur Governance der Wissenschaft ein neues theoretisches Feld. Die Analyse der neuen Governance-Instrumente sollte zu theoretischen Einsichten über den Zusammenhang von Entstehungs- und Wirkungsbedingungen, Beschaffenheit und Effekten der Instrumente führen. Die einschlägigen empirischen Studien setzen sich aber weder dieses Ziel, noch sind sie in ihrer methodischen Ausrichtung oder ihrem empirischen Zuschnitt ‚theoriefähig‘. Es überwiegen idiosynkratische Beschreibungen einzelner nationaler Instrumente, die sich auf Institutionalisierungsprozesse konzentrieren und die Effekte der Instrumente meist unberücksichtigt lassen.¹

Diese Theorielücke wird für Deutschland auch unmittelbar praxisrelevant, weil hier seit Ende der 1990er Jahren ebenfalls mit Instrumenten zur Steigerung der Performanz universitärer Forschung experimentiert wird. Viele Bundesländer haben indikatorbasierte Finanzierungssysteme für ihre Hochschulen eingeführt, die auch Indikatoren der Forschungsleistung einschließen. Auf Bundesebene wird mit den Pilotstudien des Wissenschaftsrates zum Forschungsrating und mit der Exzellenzinitiative versucht, Performanz sichtbar zu machen und eine darauf bezogene Differenzierung der Hochschullandschaft einzuleiten.

Diese Versuche haben eine teils heftige politische Debatte über die Erfolgsaussichten von Versuchen, die Forschungsperformanz durch evaluationsbasierte Finanzierung zu steigern, über die Validität der Evaluationsverfahren und mögliche negative Nebeneffekte der Verfahren ausgelöst (siehe z.B. Münch 2007; Bloch et al. 2008; Frey 2008; Hornbos-

tel 2009). Die Entscheidung des Historikerverbandes, sich nicht am bundesweiten Forschungsrating des Faches durch den Wissenschaftsrat zu beteiligen, ist nur der sichtbarste Ausweis von in Deutschland verbreiteten Zweifeln an der Sinnfälligkeit ubiquitärer Evaluationen.

Eine in diesen Debatten weitgehend ungenutzte Ressource sind die Erfahrungen, die in anderen Ländern mit den dort schon länger etablierten Evaluationssystemen gesammelt wurden und die Frage nach deren Übertragbarkeit auf die deutsche Situation. Das Ziel dieses Artikels besteht darin, durch einen Vergleich von seit langem etablierten Forschungsevaluationssystemen mit den deutschen ‚Experimenten‘ einen Beitrag zur Schließung der soeben aufgezeigten Theorielücke zu leisten und zugleich die Diskussion über die Steigerung der Forschungsperformanz in Deutschland zu bereichern. Die Ursachen für intendierte und nichtintendierte Effekte² in den Forschungsevaluationssystemen der Vorreiterländer Großbritannien, Niederlande und Australien werden herausgearbeitet und auf dieser Grundlage die Wahrscheinlichkeit für das Auftreten solcher Effekte in Deutschland abgeschätzt.

Zu diesem Zweck rekonstruieren wir zunächst die Entstehung universitärer Forschungsevaluationssysteme in Deutschland und identifizieren charakteristische Elemente, die die Grundlage für einen Vergleich mit seit längerem etablierten Forschungsevaluationssystemen bilden (2). Danach rekonstruieren wir kurz die Entstehung und Entwicklung der Forschungsevaluationssysteme in Großbritannien, den Niederlanden sowie Australien und bilanzieren Aussagen zu den Effekten dieser Systeme (3). Auf dieser Grundlage lassen sich dann zwei der deutschen Experimente – die leistungsorientierte Mittelvergabe und das Forschungsrating des Wissenschaftsrates – direkt mit den Forschungsevaluationssystemen anderer Länder vergleichen (4).³ Abschließend diskutieren wir die Wahrscheinlichkeit des Auftretens der in den anderen Ländern beobachteten Effekte als mögliche Folgen der deutschen Experimente zur Evaluation von Forschungsperformanz (5).

2 Die Entstehung universitärer Forschungsevaluationssysteme in Deutschland

Nachdem der Wissenschaftsrat seit den 1970er Jahren gelegentlich 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 (*Schimank/Lange* 2006). Mit Ausnahme des Bundeslandes Niedersachsen, in dem eine eigens gegründete ‚Wissenschaftliche Kommission Niedersachsen‘ das Evaluationsverfahren des Wissenschaftsrates auf die gesamte universitäre Forschung anwandte (*Schimank/Schiene* 2007), hat keine Landesregierung die an ‚ihren‘ Hochschulen durchgeführte Forschung zum Gegenstand systematischer Evaluationen durch Peer review gemacht. Die teilweise noch bis heute kameralistisch geführten Universitätshaushalte differenzieren kaum zwischen Ressourcen für Forschung und Lehre. Hinzu kommt in den meisten Bundesländern die Bindung der Haushaltsaufstellung an Stellenpläne und die bis vor kurzem ‚ewigen‘ Berufungszusagen an die Professoren.

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, die durch Global-

haushalte, befristete Berufungszusagen, seitens der Universität festgelegte leistungsabhängige Gehaltszulagen für neu berufene Professoren und erweiterte Handlungsbefugnisse für Dekane charakterisiert ist (Lange 2009; Lanzendorf/Pasternack 2008). Parallel zu dieser Neugestaltung des Verhältnisses von Staat, Professorenschaft und Universität wurden – mit unterschiedlicher Emphase – auch ‚schwache‘ Evaluationssysteme für Forschungsperformanz eingeführt.⁴ Dazu lassen sich vor allem die Instrumente ‚Zielvereinbarungen‘ und ‚Leistungsorientierte Mittelvergabe‘ (LOM) 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 (Classen 2009; 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; sie können deshalb als ‚schwach intrusive‘ Evaluationssysteme gelten (Gläser 2007).

Die Finanzierungsformeln der LOM binden Teile der jährlichen Mittelzuweisungen an die Erfüllung von Leistungskriterien in Lehre und Forschung, die mit quantitativen Indikatoren gemessen werden (Jaeger 2005; Leszczensky/Orr 2004). 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 LOM 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. Da die Hochschulen um diese Mittel konkurrieren (ihre Vergabe ist i.d.R. ein Nullsummenspiel), der Anteil der so gesteuerten Mittel gering ist und die internen Leistungsprozesse der Universitäten durch die Instrumente nicht berührt werden, handelt es sich bei diesen Versionen der LOM um ‚schwach kompetitive‘ Evaluationsinstrumente (Gläser 2007).

Die wenig überraschende Bilanz dieser Entwicklungen ist, dass im föderalen System der Bundesrepublik auf 16 verschiedenen Institutionalisierungspfaden landesspezifische Evaluationssysteme entstehen, und zwar das auf Peer review beruhende ‚stark intrusive‘ Evaluationssystem Niedersachsens und die auf Steuerung durch Indikatoren beruhenden schwachen Evaluationssysteme des Gros der Bundesländer. Bemühungen um bundesweite Evaluationssysteme zur Steigerung der Forschungsperformanz haben bislang zwei Ergebnisse gezeitigt. Der Wissenschaftsrat hat zwei Pilotstudien zum ‚Forschungsrating‘ (im Folgenden WR-Rating) 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 sollten in der Geschichte sowie der Elektro- und Informationstechnik folgen, allerdings hat der Historikerverband mittlerweile eine Beteiligung abgelehnt. Die Forschungsratings haben keine Konsequenzen für die Finanzierung der Universitäten.

Wesentlich folgenreicher ist dagegen die gemeinsame Exzellenzinitiative (im Folgenden: ExIn) des Bundes und der Länder, in der erhebliche zusätzliche Forschungsmittel

für ‚Graduiertenschulen‘, ‚Exzellenzcluster‘ und ‚Zukunftskonzepte‘ vergeben werden. Die ExIn wurde von der Bundesregierung initiiert und stellt in ihrer heutigen Ausprägung als Drei-Linien-Konzept letztlich einen Kompromiss zwischen drei Akteursgruppen dar (Pasternack 2008):

- der Bundesregierung, mit ihrem politischen Bedürfnis nach selektiver Förderung weniger, in den internationalen Rankings sichtbarer Eliteuniversitäten,
- der Länder, mit ihrem Interesse an Gleichbehandlung, wenn es um die Verteilung zusätzlicher Bundesmittel geht, und drittens
- der wissenschaftlichen Gemeinschaft, die erfolgreich darauf verwiesen hat, dass sich Exzellenz in der Forschung a) nicht auf der Aggregatebene ganzer Universitäten zeigt und b) diese Exzellenz nur von der Wissenschaft selbst festgestellt werden kann.

Die Mittel für die ExIn wurden entsprechend auf der Grundlage von Ex ante-Evaluationen durch rein wissenschaftlich besetzte internationale Gutachter-Panels als zeitlich befristete Projektmittel vergeben. Die geförderten Universitäten bzw. Fachbereiche und Forschergruppen können mit den erheblichen staatlichen Sondermitteln eigenständig Strukturentwicklung betreiben (Sondermann et al. 2008). Einzelne Bundesländer sind auf das Wettbewerbskonzept der selektiven Sonderförderung qua Exzellenzinitiative aufgestiegen und haben eigene Initiativen aus Landesmitteln aufgelegt (siehe nur LOEWE in Hessen). Exzellenzinitiativen unterscheiden sich von Evaluationssystemen und müssen als eigenständiges Governance-Instrument behandelt werden. Sie lassen die Grundfinanzierung der Universitäten unberührt und stellen – meist zeitlich befristet – zusätzliche Mittel für die selektive Steigerung von Forschungsperformanz bereit (Gläser 2008).

3 Internationale Erfahrungen mit Forschungsevaluationssystemen

Großbritannien, die Niederlande und Australien sind Vorreiter bei der weltweiten Etablierung von Systemen der Evaluation von Forschungsperformanz i.S. eines neuen „policy paradigms“ (Hall 1993; Musselin 2000) der Wissenschaftspolitik. Gleichwohl haben die Regierungen und Wissenschaftsadministrationen dieser Länder deutlich voneinander abweichende Evaluationssysteme institutionalisiert, deren Elemente auch in den deutschen Ansätzen zur Evaluation von Forschungsperformanz auftauchen. Deshalb werden im Folgenden in Fallstudien für jedes Land die Initialisierungsphase, Stufen der Weiterentwicklung und heutige Merkmale der Evaluationssysteme sowie die intendierten und nichtintendierten Effekte auf die dortigen Universitäten und Forscher aufgezeigt. Die Diskussion über Effekte von Systemen zur Evaluation von Forschungsperformanz wird allerdings durch prinzipielle methodologische Probleme bei der Zurechnung möglicher Effekte auf die kausalen Folgen der Evaluationen erschwert (Gläser et al. 2002). Darüber hinaus steckt die empirische Forschung zu intendierten und nichtintendierten Effekten von Systemen zur Evaluation von Forschungsperformanz noch in den Kinderschuhen. Über positive oder negative Effekte des niederländischen Evaluationssystems auf die Forschung ist bspw. fast nichts bekannt. Während die Literatur- und Datenlage über Großbritannien bereits besser – wenn auch nicht hinreichend – ist, gibt es zu Australien erste systematische Untersuchungen (Gläser/Laudel 2007; Gläser et al. 2008; Gläser/Lang/Laudel 2009).

3.1 Großbritannien: Research Assessment Exercise (RAE)

3.1.1 Entstehung und Entwicklung

In Großbritannien war die Einführung der nationalen *Research Selectivity Exercise*⁵ die Antwort auf wachsende Sorgen über eine allgemeine Degeneration der Forschungsinfrastruktur seit den späten 1970er Jahren. Der ‚Merrison Report‘ von 1979 dokumentierte „the extent to which the infrastructure of university research had been eroded and argued that if further damage was to be prevented then resources for scientific research had to be allocated selectively“ (Tapper/Salter 2003: 8). Die desolate Ausgangslage wurde zudem noch von einem weiteren Zurückfahren der staatlich gewährten Globalhaushalte für die britischen Universitäten durch die Thatcher-Regierung zu Beginn der 1980er Jahre flankiert (Taylor 2006: 2). Die erste *Research Selectivity Exercise* 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 erbaten die 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 des britischen Systems zur Evaluation der Forschungsperformanz lassen sich als schrittweise Verfeinerung und Professionalisierung des Ausgangssystems charakterisieren. Drei Hauptmerkmale, die seit der ersten Evaluation beibehalten wurden, sind 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 Nutzung der Evaluationsergebnisse für die selektive Vergabe von Forschungsgrundausrüstungen an die Universitäten. Die Hauptveränderungen seit der ersten *Research Selectivity Exercise* sind⁶:

- 1) Die Einführung einer gemeinsamen Evaluationsskala für alle Disziplinen und die Bewertung von Publikationen aller Wissenschaftler statt jeweils weniger, die Departments repräsentierender, Publikationen. Beide Veränderungen wurden bereits mit dem zweiten Durchgang im Jahre 1989 vollzogen.
- 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* (RAE) 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 (wann und in welchem Journal) 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 sehr schlecht) durch ‚Qualitätsprofile‘ ersetzt, die nun die Verteilung der Forschungsqualität einer Einheit (i.d.R. 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 zunächst auf vier, dann auf fünf und zuletzt auf über sieben Jahre ausgedehnt. Die RAE war von Beginn an ein in ständiger Entwicklung begriffenes System, dessen kritische Diskussion immer wieder zu politischen Entscheidungen über Veränderungen führte. Alle Veränderungen lassen sich maßgeblich auf die Vorschläge von Stakeholdern der RAE zurückführen – Vorschläge, die häufig in umfangreichen Konsultationsprozessen zwischen allen Beteiligten kommuniziert und in verschiedenen, von Politikern geleiteten Kommissionen gesammelt sowie in deren Abschlussberichten veröffentlicht wurden. Nichtsdestotrotz wurde die Gesamtausrichtung des Evaluationssystems auf selektive Forschungsförderung nie grundsätzlich in Zweifel gezogen.

3.1.2 Effekte

Die britischen Universitäten verteilen das Gros des aus der RAE 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 leer ausgehen. Die zentralen Universitätsleitungen behalten jedoch einen Teil des durch die RAE vereinnahmten Geldes ein, um sich selbst einen Ressourcenpool für strategische Entscheidungen (Berufungen, Profil- und Zentrenbildung etc.) zu schaffen (evidence Ltd. 2005). 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 den nächsten RAE 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 Schließungen⁷ oder Restrukturierungen von Departments erworben, die darauf abzielten, fachliche Einheiten zu kreieren oder zu stärken, die in kommenden Evaluationsrunden mutmaßlich (noch) bessere Bewertungschancen hätten (Dainty/Williams/Brown 1999). 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 RAE sichern möchte, ist inzwischen zur allgemeinen Praxis geworden (Johnston/Farrar 2003).

Grundsätzlich wird der RAE eine generelle qualitative Verbesserung der Forschung an britischen Universitäten zugeschrieben; daneben werden auch nicht-intendierte negative Effekte berichtet (z.B. bei Henkel 2000). Hier ist jedoch eine empirische Evidenz, die entsprechende Argumente für und wider stützen könnte, dünn. Das stärkste Argument für positiv-intendierte Effekte des britischen Evaluationssystems lässt sich auf die Mehrheitsmeinung der Universitätsmanager wie auch der Wissenschaftler zurückführen, der zufolge die Forschung qualitativ tatsächlich besser geworden ist. Dies wird auf die durch die RAE ausgelöste

- erhöhte Aufmerksamkeit und Unterstützung für Fragen der Forschung,
- die verbesserte materielle Ausstattung an den erfolgreichen Standorten,
- die nach Forschungsgesichtspunkten erfolgten Strukturreformen an manchen Universitäten sowie
- eine neue Kultur an den britischen Universitäten, die mit Blick auf die Rekrutierung und Beförderung von Wissenschaftlern einen größeren Wert auf Forschungsaspekte legt (Parker 2008),

zurückgeführt.

Allerdings haben die Ergebnisse des jüngsten RAE-Zyklus das Konzept, die Forschungsperformanz durch selektive Finanzierung zu steigern, auch grundsätzlich in Frage gestellt. Die im Ergebnis der letzten Evaluation veröffentlichten Qualitätsprofile demonstrieren, dass es nach mehr als 20 Jahren selektiver Finanzierung in den Departments, denen immer wieder internationale Exzellenz bescheinigt wurde, mittelmäßige und schlechte Forschung gibt, während bislang als mittelmäßig oder schlecht bewertete Departments auch exzellente Forschung enthalten. Tabelle 1 zeigt einen nur losen Zusammenhang zwischen selektiver Finanzierung und Performanz: Forschungseinheiten, denen im Jahre 2008 35% bis 60% „international führender“ und „international exzellenter“ Forschung attestiert wurden, erhielten wegen zu schlechter Globalnoten im Zeitraum 2001 bis 2008 keine Finanzierung aus dem RAE-gesteuerten Budget. Das bedeutet umgekehrt auch, dass die Finanzierung nur bedingt als Ursache für hohe oder niedrige Performanz angesehen werden kann.

Tabelle 1: Vergleich der Grade aus der RAE 2001 mit den Qualitätsprofilen der RAE 2008 für ausgewählte Universitäten im Gebiet Computer Science (Panel 25 in der RAE 2001) bzw. Computer Science and Informatics (Panel 23 in der RAE 2008)⁸

Universität	RAE 2001 Grad ⁹	Gewicht der Grade von 2001 in der Finanzierungsformel für England 2003	RAE 2008 Qualitätsprofil (prozentuale Verteilung der Forschung auf fünf Qualitätsstufen) ¹⁰				
			4*	3*	2*	1*	unclassified
Univ. of York	5*	3.357	25%	50%	20%	5%	0%
Univ. of Birmingham	5	2.793	30%	45%	20%	5%	0%
Univ. of Essex	4	1,00	15%	50%	30%	5%	0%
Loughborough Univ.	3a	0	15%	40%	35%	10%	0%
Goldsmiths College	3b	0	20%	40%	30%	10%	0%
Bournemouth Univ.	2	0	5%	30%	50%	15%	0%

Die negativen Effekte der RAE können einerseits dem verwendeten Evaluationsverfahren und andererseits den Reaktionen der Universitäten zugeschrieben werden. Der der RAE zugrunde liegende Peer review hat zwar gegenüber den in anderen Verfahren verwendeten quantitativen Indikatoren den Vorzug, dass er es gestattet, die Inhalte der Forschung in die Qualitätsbewertung einzubeziehen. Allerdings hat er auch charakteristische Schwächen: er bevorzugt den mainstream gegenüber unkonventioneller und riskanter, in ihren Ergebnissen weniger vorhersehbarer Forschung, disziplinäre gegenüber interdisziplinärer Forschung und an die Fachgemeinschaft gerichtete Grundlagenforschung gegenüber der auf außerwissenschaftliche Adressatenkreise zugeschnittenen Anwendungsforschung. Die Wissenschaftler (insbesondere diejenigen, die sich der positiven Bewertung ihrer Forschungen nicht sicher sind) 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.

Die Universitäten reagieren auf die RAE, indem sie ihre Aufmerksamkeit und ihre Ressourcen auf den evaluierten und mit incentives gesteuerten Prozess konzentrieren, 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 ins-

besondere die Lehre. Die Fokussierung des Universitätsmanagements auf Forschungsleistungen, die sich auf die Bewertung individueller Leistungen erstreckt und die Umverteilung der Lehrbelastung einschließt, verwandelt die Lehre faktisch in eine Strafe für schlechte Forschung. Die daraus entstehende Tendenz zur Trennung von Lehre und Forschung ist in ihrer Dynamik und in ihren Folgen bislang kaum untersucht worden.

3.2 Niederlande: Standard Evaluation Protocol

3.2.1 Entstehung und Entwicklung

Ähnlich wie das britische wurde auch das niederländische System zur Evaluierung von Forschungsperformanz Mitte der 1980er Jahre mit der Absicht eingeführt, der Regierung qualitative Informationen für eine stärker selektive Finanzierung der Forschung zu liefern. Das Hauptmotiv für die Einführung eines Evaluationssystems war in den Niederlanden die Intransparenz und das vollständige Fehlen einer Rechenschaftspflicht für die Forschung an den Hochschulen. Die niederländische „university research was supported to an unusually large extent through the general block grant of the university“ (Blume/ Spaapen 1988: 8), und die Performanz der solchermaßen global vorfinanzierten Forschung wurde nicht kontrolliert. Ein politisch veranlassetes Gutachten aus dem Jahre 1980 (die *Beileidsnota Universitair Onderzoek*) schlug vor, die Hälfte des Globalhaushaltes einer niederländischen Universität der Forschung zuzuwidmen und dann diesen Teil des Haushaltes auf der Basis von Ex ante-Evaluationen sogenannter ‚Forschungsprogramme‘ – das sind kohärente Forschungspläne für zunächst fünf Jahre und für eine Gruppe von mindestens fünf Wissenschaftlern – zu finanzieren. Damit sollte aber zu dieser Zeit keine Umverteilung der Forschungsanteile an den Globalhaushalten zwischen forschungsstarken und -schwachen Universitäten erreicht werden. Es wurde erwartet, dass im Rahmen dieses „Conditional Funding“-Systems alle Universitäten auf den gesetzten Anreiz mit hinreichend vielen guten Forschungsprogrammen reagieren, deren positive Evaluation den Hochschulen die tradierte Höhe ihres Globalhaushaltes sichern würde. Dieses Verfahren wurde 1983 auf nationaler Ebene eingeführt (Jeliazkova/Westerheijden 2004: 328-329). Seitdem wurde es in vier grundlegenden Aspekten verändert:

- 1) Mit der zweiten Evaluationsrunde von 1987 wurde von der Ex ante-Evaluation von Forschungsprogrammen zur Ex post-Evaluation der Ergebnisse dieser Programme übergegangen. Dieser Schritt erschien logisch, da die zweite Evaluationsrunde für die Zeit nach der Beendigung der ersten Fünfjahresprogramme angesetzt war und man natürlich wissen wollte, ob die ‚richtigen‘ Programme mit einem Vertrauensvorschuss bedacht worden waren.
- 2) Mit dem Wechsel zu Ex post-Evaluationen wurde die evaluationsbasierte Finanzierung der Universitätsforschung aufgegeben. Das niederländische Forschungsministerium hatte zwar ursprünglich vor, die Evaluationsergebnisse unmittelbar mit den Entscheidungen über die Verteilung der Forschungshaushalte an die Universitäten zu verknüpfen; es wurde aber kein Konsens über das Verfahren erzielt (Van der Meulen 2007: 196). Statt dessen wurden ab 1987 die Finanzmittel für die Forschungsgrundausrüstung der Universitäten auf der Grundlage von Formeln verteilt, die ein starkes historisches Moment enthielten: Ausschlaggebend für die Gesamthöhe der staatlichen Zuwendung war die Höhe des Vorjahreshaushaltes plus die üblichen inkrementellen Steigerungen.

Die Forschungsindikatoren in der Formel waren sehr simpel konzipiert, mit der Anzahl an Doktoranden als einflussreichstem Indikator (ca. 8 % der Grundfinanzierung).

- 3) Ein dritter interessanter Trend ist die schrittweise Übertragung der Verantwortlichkeit für die Forschungsevaluationen vom Staat auf die Hochschulen, die letztlich der – auch im internationalen Vergleich – starken politischen Position der niederländischen Hochschulen geschuldet ist. Die Verantwortung für die Durchführung der Evaluationen wurde ab dem dritten Durchgang im Jahre 1993 vom Ministerium an die Vereinigung der Niederländischen Universitäten (*Vereniging van Universiteiten/ VSNU*) delegiert. Seit Einführung des zweiten Standardprotokolls im Jahre 2003 sind die einzelnen Universitäten selbst für die Organisation der Evaluation ihrer Forschung verantwortlich: sie müssen im Drei-Jahres-Rhythmus eine interne Forschungsevaluation durchführen und im Sechs-Jahres-Rhythmus eine externe Evaluation organisieren (*VSNU/NWO/KNAW* 2003). Da damit in den meisten Disziplinen die landesweite Evaluation durch separate Evaluationen der Forschung in einer Universität abgelöst wurde, sind auch Vergleiche der Forschungsleistungen zwischen Disziplinen erschwert worden.
- 4) Die sukzessive Übertragung der Verantwortlichkeiten auf die Universitäten selbst war begleitet von einer steigenden Formalisierung der Inputs (der den Panels zu übermittelnden Informationen), der Verfahren und der Outputs (der Qualitätsurteile) der Forschungsevaluationen. Mit der Übertragung der Verantwortlichkeit für die Evaluationen auf die VSNU wurde ein „Standard Evaluations-Protokoll“ (*Standard Evaluation Protocol for Public Research Organizations*) eingeführt. Dieser für die Universitäten und disziplinären Gutachter-Panels verbindliche Leitfaden enthält Standards für die Leistungserhebung, -auswertung und abschließende Qualitätsbewertung. Er schreibt die Evaluation der Forschung in den vier Dimensionen Qualität, Produktivität, wissenschaftliche Relevanz und Entwicklungsfähigkeit vor. In diesen Dimensionen werden Bewertungen auf einer Fünf-Punkte-Skala von „exzellent“ bis „schwach“ vorgenommen. Das Standardprotokoll hat insbesondere in seiner aktuellen Ausgestaltung seit 2003 sehr zu einer Vergleichbarkeit – Kritiker würden sagen: zur Suggestion einer Vergleichbarkeit – der Qualitätsstandards über Fächergrenzen hinweg beigetragen (*Van der Meulen* 2007: 198).

3.2.2 Effekte

In den Niederlanden werden die evaluationsbasierten Informationen zur Forschungsqualität hauptsächlich als Argumente in Budgetverhandlungen zwischen den verschiedenen Ebenen der Universität (Zentrales Universitätsmanagement – Department – Institut – Forschergruppe/ Forschungsschule) verwandt. Die Ergebnisse dieser internen Verhandlungen sind aber keinesfalls durch die Evaluation determiniert: sie transformieren nicht sklavisch die Qualitätsurteile aus der Evaluation in Geldzu- oder -abflüsse, und sie reproduzieren keinen Automatismus, mit dem schlecht evaluierte Einheiten zwingend bestraft oder exzellente zusätzlich belohnt würden. Im Gegenteil: schlecht evaluierte Einheiten können durchaus mehr Geld erhalten, weil sie für die Fakultät oder das Department aus anderen Gründen (z.B. hohe Attraktivität für Studenten, gute Lehre etc.) wichtig sind und deshalb gestärkt werden müssen. Umgekehrt gibt es Fälle, in denen exzellent evaluierte Forschungseinheiten Geld abgeben müssen, wenn z.B. die Studentenzahlen stark absinken (*Jongbloed/Van der Meulen* 2006).

Die Auswirkungen der Evaluationen auf der Mikroebene der Forschungsorganisation an den niederländischen Universitäten und damit auf das Handeln der individuellen Wis-

senschaftler sind bislang nicht systematisch erforscht. In den Augen der wenigen befragten Wissenschaftler haben die Evaluationen keine großen Folgen. Ein gutes Abschneiden bietet einen gewissen Schutz vor Interventionen des Managements wie z.B. Umstrukturierungen und eine stärkere Position in Budgetverhandlungen. Schlechte Ergebnisse verringern diesen Schutz, haben aber nicht automatisch weitergehende Folgen. Die Evaluationen sind also lediglich eine von mehreren Informationsquellen für die Entscheidungen eines sehr aktiven und durchsetzungsfähigen Managements (*Westerheijden 1997; Jongbloed/Van der Meulen 2006*).

3.3 Australien: Evaluationsbasierte Indikatorensteuerung

3.3.1 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 *Upgradings* oder Fusionen in Universitätsrang erhoben (*Marginson/Considine 2000*). Bald darauf setzten Bemühungen ein, die Finanzierung der Forschung selektiv zu gestalten. In ihrem *White Paper on Higher Education* erklärte die australische Bundesregierung ihre Absicht, ein Finanzierungssystem entwickeln zu wollen, dass die Performanz der Universitäten berücksichtige (*Anderson et al. 1996: 2*). 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 finanzierte 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 Messung der Performanz herangezogen, und zwar die Zahl der Publikationen, die Abschlüsse (Masters und PhD) und die Zahl der immatrikulierten Forschungsstudenten. Bei der Erhebung der Publikationszahlen wurde zunächst mit insgesamt 22 Kategorien operiert, was sich als wenig praktikabel erwies. Deshalb sind seit 1996 nur noch vier Kategorien von Publikationen für die Finanzierung relevant sind, und zwar begutachtete Zeitschriftenaufsätze, begutachtete Konferenzpapiere, Bücher und – heute fünffach gewichtet – Buchkapitel (*Butler 2003*).

Die Gewichte der einzelnen Indikatoren wurden mehrmals verändert. Die Einbeziehung von Immatrikulationen und Abschlüssen der Forschungsstudenten reflektiert die Umbewertung von deren Ausbildung. Sie war im Jahre 1990 noch Bestandteil der Lehrkomponente, wurde aber seit 1996 der Forschungskomponente zugerechnet. 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*).

3.3.2 Effekte

Die australischen Hochschulen reagieren auf ein Evaluationssystem, das ihre Grundfinanzierung vor allem an die eingeworbenen Drittmittel bindet, mit der wenig überraschenden

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 in Entscheidungen über Beförderungen als Kriterium herangezogen (Gläser *et al.* 2008). Konzentration der Anstrengungen auf die Steigerung der Drittmittelwerbung heißt vor allem, die knappen Mittel für die Grundfinanzierung der Forschung in ‚profit centres‘ zu investieren, d.h. in Organisationseinheiten (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 Drittmittelwerbung knapp gescheitert sind, neu eingestellte Wissenschaftler sowie Nachwuchswissenschaftler können sich bei den Universitäten um Mittel bewerben, mit denen sie Drittmittelanträge vorbereiten.¹¹ Jenseits dieser intern kompetitiv vergebenen Anschubfinanzierung gibt es keine Grundausstattung der Forschung oberhalb einer basalen Infrastruktur (Gläser/Laudel 2007).

In Australien kann – wie in den beiden anderen Ländern 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 zu deutschen als *monopolistisch*, *karg* und *biased* beschreiben (Laudel 2006). Australische Wissenschaftler sind auf eine einzige bedeutende Drittmittelquelle angewiesen, den *Australien Research Council* (ARC) bzw. – für medizinische Forschung – den *National Health and Medical Research Council* (NHMRC). Andere Quellen wie die australische Industrie oder Förderprogramme der Regierung spielen praktisch keine Rolle. Die Forschungsräte unterstehen dem Wissenschafts- bzw. dem Gesundheitsministerium. Die Minister entscheiden letztlich alle Förderungen und haben die Möglichkeit, ihnen nicht genehme Projekte zu streichen.

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 ARC 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 Drittmittelwerbung. Zweitens hat die Orientierung australischer Forscher an den Prioritäten der Drittmittellandschaft eine stärkere Anwendungsorientierung der Forschung und einen Trend zum Mainstream zur Folge. Drittens verengen sich die individuellen Forschungsportfolios – die Wissenschaftler bearbeiten tendenziell weniger und einander ähnlichere Themen. Viertens schließlich reagieren die Wissenschaftler auf die Mittelknappheit und auf den Verlust von Forschungsmöglichkeiten, indem sie kostenneutrale Forschung betreiben. Sie wählen kostengünstigere Untersuchungsobjekte oder führen we-

niger Experimente, Testreihen oder Beobachtungen durch, was die Validität und Reliabilität der Forschung einschränkt (Gläser et al. 2008; Gläser/Lange/Laudel 2009).

4 Abschätzung von Effekten der Evaluationssysteme in Deutschland

Während es noch zu früh ist, um Effekte der deutschen Evaluationssysteme zu beurteilen, lassen sich aus dem internationalen Vergleich doch einige Schlussfolgerungen zu ihrem Innovationspotenzial und zu erwartbaren Nebenfolgen der Instrumente ziehen. Alle drei in Deutschland praktizierten Systeme – die beiden bundesweiten Instrumente WR-Rating und ExIn sowie die im Rahmen der Hochschulfinanzierung der Länder praktizierte LOM – liefern den forschungspolitischen Akteuren, dem Universitätsmanagement und auch den Wissenschaftlern selbst Informationen über die Forschungsperformanz an den Universitäten. Dies war das wichtigste Motiv für die Etablierung eines Evaluationssystems in den Niederlanden. Auch die Konstitution einer breiteren Öffentlichkeit für Belange der Forschung und die Sichtbarmachung vermeintlich guter und schlechter Forschungsstandorte können – ähnlich wie in Großbritannien – zumindest als Nebeneffekte der Ratings, besonders aber der ExIn, gelten. Die Einführung der LOM hat – ähnlich wie im australischen Hochschulsystem – vor allem hochschulinterne Effekte.

Vergleicht man nun konkret die deutschen Systeme und Instrumente zur Evaluation von Forschungsperformanz mit den zuvor vorgestellten internationalen Beispielen, dann zeigen sich Ähnlichkeiten, die Prognosen gestatten. Die LOM 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, wie Tabelle 2 zeigt.

Tabelle 2: Vergleich der indikatorbasierten Systeme von Nordrhein-Westfalen und Australien (2005)

	Nordrhein-Westfalen	Australien
Indikatoren (relatives Gewicht in %)		
– Drittmittelerwerbungen	81	54,8
– Abschlüsse von Forschungsstudenten		
– Zahl der immatrikulierten Forschungsstudenten	Promotionen: 19	Master/PhD: 29,1
– Publikationen		7,7
		8,4
Beeinflusste Mittel		
– Anteil an der staatlichen Grundfinanzierung (%)		
– Anteil an den Einnahmen der Universitäten (%)	2,1	20,8
	2,1	7,9
Reaktionen der Universitäten		
	Variation der Sachkostenausstattung der Professoren	Internes Kopieren der Finanzierungsformel

Der australische Fall demonstriert einen generellen und einen spezifischen Effekt der Indikatorensteuerung. Der generelle Effekt ist das *goal displacement*, d.h. die Fokussierung der Performanzsteigerung auf die Indikatoren selbst statt auf die Qualität, die die Indikatoren eigentlich messen sollen. Da Publikationen, Drittmittelerwerbungen und die Zahl der Doktoranden als Indikatoren der Forschungsleistung von fragwürdiger Validität sind, ent-

steht eine erhebliche Differenz zwischen politischer Absicht und Verhalten der Universitäten. Die wegen der Bedeutung des Indikators für die staatlichen Zuwendungen erfolgende Fokussierung auf die Drittmittelwerbung 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. Dies führt zu den in Abschnitt 3.3.2 angeführten qualitäts- und diversitätsvermindernden Effekten in der Forschung jener absoluten Majorität unter den Wissenschaftlern, die nicht der internationalen Elite angehören (und dadurch problemlos Zugang zu Drittmitteln erhalten).

Anders als in Australien ist die deutsche Drittmittellandschaft *pluralistisch, reich und neutral*, d.h. es gibt eine Vielzahl von öffentlich finanzierten und privaten Förderagenturen und Stiftungen, die in ihren Förderprogrammen (noch) wenig inhaltlichen Einfluss auf die Forschungsagenden der individuellen Wissenschaftler nehmen. Dem in Australien von der Regierung erzwungenen Anwendungsbias könnte zukünftig in Deutschland ein Bias zugunsten der Grundlagenforschung entsprechen. Unsere Interviews in drei Universitäten zeigen einen Trend zur Bevorzugung von DFG-Drittmitteln als ‚höherwertig‘. In den drei Universitäten werden 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 z.B. 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 den Forschungsevaluationssystemen in den Niederlanden und Großbritannien. Freilich kann man mit Blick auf die bisherigen Pilotstudien in Deutschland kaum von einem bereits institutionalisierten System sprechen. Da mit den fächerbezogenen Ratings in Deutschland bewusst keine Ressourcenallokationsmechanismen verbunden werden – wie dies z.B. im britischen RAE geschieht –, scheint ein Blick auf die Effekte des Standard Evaluation Protocol in den Niederlanden instruktiv zu sein. Hier ließ sich in Abschnitt 3.2.2 resümieren, dass die Evaluationsergebnisse weniger Gegenstand der nationalen Hochschulpolitik als vielmehr der Organisationspolitik der Hochschulen selbst sind und der strategischen Ausrichtung der Universitäten dienen. Dabei werden die Entscheidungen zu Ressourcenverteilungen oder Strukturveränderungen nicht durch Evaluationsergebnisse dominiert. Auch als exzellent evaluierte Einheiten können geschlossen oder mit anderen fusioniert werden, wenn sie z.B. zu wenige Studenten attrahieren oder für das Profil der Hochschule als weniger relevant eingeschätzt werden, während schlecht evaluierte Einheiten zur Stärkung ihrer Forschungsqualität zusätzliches Geld erhalten können. Die Entkopplung von Evaluation und Finanzierung verringert demzufolge den Druck zur Maximierung der Einnahmen und erhöht den Handlungsspielraum des Universitätsmanagements.

Dagegen neigen Systeme wie die britische RAE zu einer Prä-Determinierung des Umgangs mit Informationen über Forschungsqualität, da sie die Ratings unmittelbar mit einem staatlich vorgebenden Verteilungsmechanismus verknüpfen, den die Universitäts-

leitungen i.d.R. unverändert an ihre Departments weiterreichen. Würde das WR-Rating in seiner gegenwärtigen Form institutionalisiert, würde es voraussichtlich die Handlungsspielräume des Hochschulmanagements erweitern, aber wenig zu einer Differenzierung der deutschen Hochschullandschaft beitragen.¹²

Tabelle 3: Vergleich des WR-Ratings in Deutschland mit den Forschungsevaluationssystemen der Niederlande und Großbritanniens

	WR-Rating	Standardprotokoll	RAE
	Deutschland	Niederlande	Großbritannien
Bewertungseinheit	Forschung in einer Disziplin in einer Universität		
Teilnahme	freiwillig (aber starker informeller Druck)	obligatorisch	freiwillig
Eingereichte Informationen	Publikationen und Kontextinformationen z.B. über Drittmittelanträge und Anzahl von Forschungsstudenten (Doktoranden)		
Rekrutierung der Gutachter	überwiegend nationale Peers		
Verfahren	Analyse der Publikationen und Kontextinformationen		
Ergebnis	Rating von 5 („exzellent“) bis 1 („nicht befriedigend“) in drei Dimensionen (Forschung, Nachwuchsförderung, Wissenstransfer)	Rating von 5 („exzellent“) bis 1 („schwach“) in vier Dimensionen (Qualität, Produktivität, wissenschaftliche Relevanz, Entwicklungsfähigkeit)	Qualitätsprofile: Verteilung der Forschung über fünf Qualitätsniveaus von 4* („international führend“) bis „nicht bewertet“ („unter dem nationalen Standard“)
Verbindung zur Finanzierung	nein	nein	ja, stark (ca. 7,5% der Einnahmen der Universitäten)

Während das WR-Rating Ähnlichkeiten mit dem niederländischen System aufweist, lässt sich die ExIn besser mit der britischen RAE vergleichen: beide beruhen auf Peer review und haben erhebliche finanzielle Konsequenzen für die Universitäten. Wegen der Selbst-Selektionen in Großbritannien ist beiden Systemen auch gemeinsam, dass sich nicht alle Universitäten am Wettbewerb beteiligen. Die Unterschiede zwischen beiden Instrumenten sind aber dennoch erheblich – die RAE steuert dauerhaft die Grundfinanzierung der Universitäten, während die ExIn zeitlich befristet zusätzliche Grundausstattung vergibt.

Wegen der finanziellen Konsequenzen der ExIn hat diese – wie die RAE – 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 auf nationaler Ebene 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 Auseinanderdriften der beiden im Humboldtschen Modell verkoppelten Leistungsprozesse von Forschung und Lehre steht in Deutschland erst am Anfang und fällt deutlich schwächer aus als in Großbritannien. Die ExIn bietet aber vielen erfolgreichen ‚Principal Investigators‘ in Exzellenz-Clustern

und Graduiertenschulen die Möglichkeit, mit ihren Dekanen und Universitätsleitungen über befristete Freistellungen von der Lehre zu verhandeln. In welchem Ausmaß sich dadurch Ausdifferenzierungen der Rollen von Lehrenden und Forschenden ergeben, ist noch nicht absehbar.¹³

Ein zweiter Effekt, den die Analyse der RAE 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 RAE 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. In Deutschland ist z.B. die Frage der Verdrängung des Weiterbildungsauftrags der Hochschulen im Exzellenzdiskurs bereits aufgeworfen worden (*Kloke* 2009).

Die einschlägigen Initiativen privater Stiftungen sind in diesem Kontext zwar ein interessantes Signal; es ist aber deutlich, dass z.B. die von Stifterverband und KMK hälftig bereitgestellten € 10 Mio. für die Initiative zur Förderung exzellenter Lehre nicht annähernd den gleichen Effekt haben können, wie die € 1,9 Mrd. für die ExIn in der Forschung.

5 Fazit

Vergleiche der Evaluationssysteme von Vorreiterländern mit den Entwicklungen in Deutschland werden dadurch eingeschränkt, dass erstere in unitarischen politischen Systemen (Niederlande und Großbritannien) oder in föderalen Systemen, in denen aber die Finanzierung der Hochschulen in der Verantwortung der Bundesregierung liegt (Australien) entwickelt wurden.¹⁴ Diese politischen Systeme bieten dem Aufbau kohärenter nationaler Systeme der Evaluation von Forschungsperformanz bessere Chancen als der föderative Aufbau Deutschlands mit Universitäten unter Länderhoheit und einem in Sachen Hochschulpolitik weitgehend entmachteten Bundesministerium. Deshalb beobachten wir in Deutschland eine diversifizierte Institutionalisierung von Forschungsevaluationssystemen in den Bundesländern und erst mit dem WR-Rating und der ExIn Versuche eines Einstiegs in ein nationales Evaluationssystem. Wohin allerdings die ersten Runden der ExIn als auch die Pilotstudien des Wissenschaftsrates zum Forschungsrating führen, ob es also zu einer echten Institutionalisierung der Forschungsevaluation von Universitäten kommt, ist noch nicht abzusehen. Es überrascht jedenfalls nicht, dass die deutsche Wissenschaftspolitik den Weg einer Exzellenzinitiative beschreitet und nicht gleich Kurs auf ein dauerhaftes Evaluationssystem für das ganze Land nimmt: Den Ländern fehlt es vielfach an Geld, eine Performanzsteigerung der universitären Forschung über zusätzliche, auf Dauer gestellte monetäre Anreize zu verfolgen, und die Bundesregierung darf nur eine ebenso zeitlich befristete wie projektförmige Zusatzfinanzierung verteilen. Ob sich die Initiativen auf Bundesebene zu einem nationalen System verstetigen, lässt sich bislang kaum prognostizieren.

Dennoch können aus dem internationalen Vergleich einige Schlussfolgerungen gezogen werden. Eine erste Beobachtung ist, dass Systeme zur Evaluation von Forschungsperformanz lernende Systeme sind, die sich weiterentwickeln, sich an neue Bedingungen anpassen und ihre Methoden ausdifferenzieren. Für diesen Lernprozess bedürfen sie ihrerseits einer ständigen Evaluation. Befristete ‚Initiativen‘ gleich welcher Art sind für solche Lernprozesse weniger geeignet.

Bei den Evaluationsverfahren ist der wesentlich aufwändigere qualitative Peer review den in Forschungsevaluationssystemen eingesetzten einfachen quantitativen Indikatoren deutlich überlegen. Die australischen Erfahrungen zeigen, dass starke indikatorbasierte Systeme ein *goal displacement* auslösen, bei dem das gute Abschneiden in den numerischen Werten, die die Indikatoren zum Ausdruck bringen, das Ziel der Steigerung der Forschungsqualität ersetzt. Solche Effekte deuten sich bei der deutschen LOM bereits an, wenn die Drittmiteleinwerbung zu einem dominierenden Kriterium bei Neuberufungen und der leistungsbezogenen Vergabe der Grundaustattungen wird. Mit der durchaus möglichen Verwandlung der LOM von einem ‚schwachen‘ in ein ‚starkes‘ Evaluationssystem und der erwartbaren weiteren Ausweitung des Handlungsspielraums der Universitäten wird sich das *goal displacement* auch an deutschen Hochschulen voraussichtlich weiter verstärken.

Zwei wichtige internationale Erfahrungen, die gleichermaßen auf Evaluationssysteme und Exzellenzinitiativen anwendbar sind, betreffen die Grenzen der Selektivität und die Balance der Leistungsprozesse der Universitäten. Die britische RAE, die mit dem primären Ziel einer Leistungssteigerung durch selektive Finanzierung eingeführt worden war, hat durch den Übergang von ‚Globalnoten‘ für fachbezogene Einheiten der Universität auf Qualitätsprofile unfreiwillig demonstriert, dass die Forschungsperformanz im britischen System breiter verteilt ist, als die Globalnoten nahe legten. Offensichtlich gibt es überall im Wissenschaftssystem leistungsbezogene Binnendifferenzierungen, die der Konzentration und Steigerung der Performanz durch selektive Finanzierung Grenzen auferlegen. Selektive Forschungsfinanzierung läuft deshalb immer Gefahr, ‚Exzellenznischen‘ nicht angemessen zu belohnen und mittelmäßige Forschung im ‚Huckepackverfahren‘ mitzufinanzieren.

Eine wichtige positive Eigenschaft von Instrumenten für die Steigerung der Forschungsperformanz besteht darin, dass sie die Aufmerksamkeit von Universitäten und Wissenschaftlern auf die Forschungsleistungen fokussieren. Diese Eigenschaft ist aber zugleich hochproblematisch, da sie in einem komplexen System von Leistungsprozessen die Prioritäten zugunsten der evaluierten und belohnten Leistungsprozesse verschieben. In Großbritannien und in Australien zeigt sich deutlich, dass die starken Evaluationssysteme zu einer relativen Vernachlässigung derjenigen Leistungsprozesse führen, die nicht gleichermaßen stark 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 KMK 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) Forschungsperformanz suchen lässt.

Ein weiterer Gegenstand von Lernprozessen, die aus der Analyse von Effekten der Evaluationssysteme der Vorreiterländer resultieren können, wird hier abschließend nur kurz angerissen: Instrumente zur Steigerung der Forschungsperformanz sollen idealiter die Forschungsinhalte nicht beeinflussen, sondern lediglich die Qualität der Forschung

verbessern. Die Studien zur britischen RAE und zum australischen Forschungsevaluationssystem zeigen jedoch, dass dieses Ziel verfehlt wird. Anwendungsorientierte Forschung, riskante Forschung und nonkonformistische Forschung – letztlich all jene Forschung, die keine Drittmittel attrahiert und deren Ergebnisse keinen oder nur geringen Niederschlag in international führenden Fachzeitschriften findet – gefährden in mit Allokationsentscheidungen 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 i.S. nicht wahrgenommener Forschungsoptionen ist schwer messbar und wiegt deshalb als Argument – obwohl nicht selten vorgebracht – weniger schwer. Damit enden wir bei einem Grunddilemma jeder Evaluation von Forschungsperformanz, das von Forschungspolitik und Wissenschaftsmanagement gleichwohl berücksichtigt werden sollte: Evaluationssysteme und Performanzbeurteilungen führen häufig zu einer starken Homogenisierung der institutionellen Umwelt von Forschung, die auch die Forschung 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.

Anmerkungen

- 1 Siehe pars pro toto für Spanien *Cruz-Castro/Sanz-Menéndez* (2007), für Italien *Reale/Barbara/Costantini* (2007), für Norwegen *Schneider* (2009) und die in den folgenden Länderabschnitten angeführte Literatur.
- 2 Siehe zu einer theoretisch-reflexiven Diskussion über intendierte und nichtintendierte bzw. intentionale und transintentionale Folgen von Evaluationen im Wissenschaftssystem *Braun* (2008).
- 3 Die Rekonstruktionen und Analysen der Effekte beruhen für Großbritannien und die Niederlande auf einer Sekundäranalyse von Publikationen, Forschungsberichten und internen Dokumenten der Universitäten. Das australische Forschungsevaluationssystem war Gegenstand einer umfangreichen empirischen Analyse, die auch deutsche Universitäten als Vergleichsfälle einschloss. In Interviews mit Leitern und Wissenschaftlern an sieben australischen und zwei deutschen Universitäten wurden die internen Reaktionen der Universitäten auf die Forschungsevaluationssysteme, die Veränderungen der Situationen der Wissenschaftler und die Reaktionen der Wissenschaftler in sechs Disziplinen analysiert, was eine kausale Zurechnung von epistemischen Effekten auf Governance-Instrumente ermöglichte (*Gläser et al.* 2008; *Gläser/Lange/Laudel* 2009).
- 4 Siehe zu entsprechenden Typologien *Whitley* (2007) und *Gläser* (2007).
- 5 Die *Research Selectivity Exercise* wurde ab 1989 in *Research Assessment Exercise* (im Folgenden: RAE) umgetauft. Dennoch kommt die erste Namensgebung dem erklärten Zweck bis heute am nächsten: Die Erhöhung der Selektivität bei der Verteilung öffentlicher Forschungsmittel – die zugleich eine Umverteilung zwischen den Universitäten beinhaltet – ist bis heute die ultima ratio dieses Forschungsevaluationssystems.
- 6 Siehe neben der bereits erwähnten Literatur: *Roberts* (2003), *Barker* (2007) und *HEFCE* (2008).
- 7 Zuletzt haben Schließungspläne für mehrere sozialwissenschaftliche Departments an der Universität von Liverpool für überregionales Aufsehen gesorgt. Hier sind nicht einmal dezidiert schlechte RAE-Ergebnisse der Grund für solche Managemententscheidungen, sondern allein die Wahrnehmung des Universitätsmanagements, dass eine Bewertung unterhalb des Höchstniveaus „international führend“ in der RAE trotz hoher Auslastung in der Lehre keine Bestandssicherheit rechtfertigt (*Newman* 2009).
- 8 Quelle: Eigene Zusammenstellung nach *HEFCE* 2001, *HEFCE* 2008, *Barker* 2007 und *Science and Technology Committee* 2004.

- 9 Die verbalen Beschreibungen der Grade lauten: 5*: >50% international exzellent, sonst national exzellent; 5: <50% international exzellent, sonst national exzellent; 4: durchweg national exzellent, einige Hinweise auf internationale Exzellenz; 3a: >2/3 national exzellent mit einigen Hinweisen auf internationale Exzellenz; 3b: >50% national exzellent; 2: <50% national exzellent; 1: nicht national exzellent (*Barker 2007: 5*, eigene Übersetzung).
- 10 Die verbalen Beschreibungen der Qualitätsstufen lauten: 4*: international führend; 3*: international exzellent; 2*: international anerkannt; 1*: national anerkannt; unclassified: unter nationalem Niveau (*Barker 2007: 5*, eigene Übersetzung).
- 11 Die Fokussierung des Universitätsmanagements auf die Drittmittelinwerbung kann als „goal displacement“ (*Blau/Scott 1962: 229*) beschrieben werden – nicht die Erhöhung der Forschungsqualität, sondern die Steigerung der Drittmittelleinnahmen steht im Zentrum der Aufmerksamkeit.
- 12 Kritiker befürchten allerdings beim WR-Rating disziplinäre Effekte, die perspektivisch zu einem Diversitätsverlust in den Fächern führen können. Das Rating hat demzufolge einen faktischen bias zugunsten ganz bestimmter Indikatoren für exzellente Forschungsperformanz – vor allem des Fachbeitrags in einem peer reviewed Journal – der hierauf spezialisierte disziplinäre Ausrichtungen und ihre Produktionsstandorte bevorteilt. Vgl. am Beispiel der Pilotstudie Soziologie *Münch/Baier (2009)*.
- 13 Siehe hierzu auch *Meier/Schimank (2009)*.
- 14 Siehe zum Einfluss des Regierungssystems auf die Pfadabhängigkeit von Hochschul- und Wissenschaftspolitikern auch *Lange/Schimank (2007)*.

Literaturverzeichnis

- Anderson, Don/Johnson, Richard/Milligan, Bruce, 1996: Performance-Based Funding of Universities. Commissioned Report No. 51. Canberra, Australian National Board of Employment, Education, and Training. Quelle: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/66/8d.pdf (Zugriff 22.05.2009).*
- Barker, Katharine, 2007: The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System, in: Research Evaluation 16, S. 3-12.*
- Blau, Peter M./Scott, Richard W., 1962: Formal Organizations. San Francisco: Chandler.*
- Bloch, Roland et al. (Hrsg.), 2008: Making Excellence. Grundlagen, Praxis und Konsequenzen der Exzellenzinitiative, Bielefeld: Bertelsmann.*
- Blume, S.S./Spaapen, J.B., 1988: External Assessment and “Conditional Financing” of Research in Dutch Universities, in: Minerva 26, S. 1-30.*
- Braun, Dietmar, 2008: Evaluation und unintendierte Effekte – eine theoretische Reflexion, in: Matthies, Hildegard/Simon, Dagmar (Hrsg.), Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. Leviathan-Sonderheft 24, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 103-124.*
- Butler, Linda, 2003. Modifying Publication Practices in Response to Funding Formulas. In: Research Evaluation 12, S. 39-46.*
- Classen, Claus Dieter, 2009: Zielvereinbarungen – Ein neues Steuerungsinstrument der Hochschulpolitik aus juristischer Sicht. In: Hochschulmanagement 4, S. 31-35.*
- Cruz-Castro, Laura/ Sanz-Menéndez, Luis, 2007: Research Evaluation in Transition: Individual versus Organisational Assessment in Spain, in: Whitley, Richard/Gläser, Jochen (Eds.), The Changing Governance of the Sciences: The Advent of Research Evaluation Systems. Dordrecht: Springer, S. 205-223.*
- Dainty, Roger/Williams, Syd/Brown, Paul, 1999: Current Trends and Changes to Research Strategy in UK Universities and Medical Schools. Paper presented at the 50th Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Pittsburgh.*
- evidence Ltd., 2005: Impact of Selective Funding of Research in England, and the Specific Outcomes of HEFCE Research Funding. Leeds: evidence Ltd.*
- Frey, Bruno S., 2008: Evaluitis – eine neue Krankheit, in: Matthies, Hildegard/Simon, Dagmar (Hrsg.), Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 125-140.*

- Gläser, Jochen, 2007: The Social Orders of Research Evaluation Systems, in: *Whitley, Richard/Gläser, Jochen* (Eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*. Dordrecht: Springer, S. 245-266.
- Gläser, Jochen, 2008: Trends im evaluationsbasierten Management universitärer Forschung. Bericht an die interdisziplinäre Arbeitsgruppe ‚Exzellenzinitiative‘ der Berlin-Brandenburgischen Akademie der Wissenschaften (unveröffentlicht).
- Gläser, Jochen/Lange, Stefan/Laudel, Grit/Schimank, Uwe, 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, S. 145-170.
- Gläser, Jochen/Lange, Stefan/Laudel, Grit, 2009: Auswirkungen der evaluationsbasierten Forschungsfinanzierung an Universitäten auf die Inhalte der Forschung. Australien und Deutschland im Vergleich, in: *Wissenschaftsrecht* 42, im Erscheinen.
- Gläser, Jochen/Laudel, Grit/Hinze, Sybille/Butler, Linda, 2002: Impact of Evaluation-Based Funding on the Production of Scientific Knowledge: What to Worry About, and How to Find Out (Expertise für das BMBF), Quelle: <http://www.sciencepolicystudies.de/dok/expertise-glase-lau-hin-but.pdf>.
- 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, S. 127-151.
- Hall, Peter, 1993: Policy Paradigms, Social Learning, and the State. The Case of Economic Policy Making in Britain, in: *Comparative Politics* 25, S. 275-296.
- HEFCE (Higher Education Funding Council for England), 2001: RAE 2001: Results. Excel-Spreadsheet. Quelle: <http://www.hero.ac.uk/rae/Results> (Zugriff 22.05.2009).
- HEFCE (Higher Education Funding Council for England), 2008: RAE 2008: Results. Quelle: <http://submissions.rae.ac.uk/results/outstore/RAEOutcomeFull.pdf> (Zugriff 22.05.2009).
- HEFCE (Higher Education Funding Council for England), 2008: Research Assessment Exercise 2008. Quelle: <http://www.rae.ac.uk> (Zugriff 22.05.2009).
- Henkel, Mary, 2000: *Academic Identities and Policy Change in Higher Education*. London: Jessica Kingsley.
- Hornbostel, Stefan, 2009: Alles eine Frage der Perspektive? Münchs Kritik am deutschen Wissenschaftssystem, in: *Soziologische Revue* 39, S. 15-20.
- Jaeger, Michael, 2005: Leistungsbezogene Mittelvergabe und Qualitätssicherung als Elemente der hochschulinternen Steuerung, HIS-Kurzinformationen A/12/2005, Hannover: HIS.
- Jeliazkova, Margarita/Westerheijden, Don F., 2004: The Netherlands: A Leader in Quality Assurance Follows the Accreditation Trend, in: *Schwarz, Stefanie/Westerheijden, Don F.* (eds.), *Accreditation and Evaluation in the European Higher Education Area*, Dordrecht et al.: Kluwer, S. 323-346.
- Johnston, Chris/Farrar, Steve, 2003: 100 New Chairs Created in Bid to Lift RAE Scores, in: *London Times Higher Education Supplement*. 12..12.2003.
- Jongbloed, Ben/Van der Meulen, Barend, 2006: *Investeren in dynamiek. Eindrapport Commissie Dynamisering*. Deel 2, Enschede: CHEPS.
- Kloke, Katharina, 2009: ‚Nischen-Dasein, Aschenputtel-Existenz und Nebenschauplatz‘? Die Rolle der wissenschaftlichen Weiterbildung an deutschen Hochschulen. In: *Hochschulmanagement* 4, S. 43-49.
- König, Karsten, 2007: *Die Reform der staatlichen Hochschulsteuerung durch Kontraktmanagement und ihre Bedeutung für Mitbestimmung und Chancengleichheit an den Hochschulen*. Wittenberg: Institut für Hochschulforschung.
- Lange, Stefan, 2009: Neuer gesellschaftlicher Legitimierungsbedarf der Wissenschaft in Deutschland – Veränderungen in der Wissenschafts-Governance am Beispiel der Universitäten, in: *Böhlke, Effi et al.* (Hrsg.), *Exzellenz durch Steuerung? Neue Herausforderungen für das deutsche und französische Wissenschaftssystem*, WZB-Discussion Paper SP III 2009-602, Berlin: WZB, S. 70-96.
- Lange, Stefan/Schimank, Uwe, 2007: Zwischen Konvergenz und Pfadabhängigkeit: New Public Management in den Hochschulsystemen fünf ausgewählter OECD-Länder, in: *Holzinger, Katharina/Joergens, Helge/Knill, Christoph* (Hrsg.), *Transfer, Diffusion und Konvergenz von Politiken*, Sonderheft 38 der Politischen Vierteljahresschrift, Wiesbaden: VS, S. 522-548.

- Lanzendorf, Ute/Pasternack, Peer, 2008: Landeshochschulpolitiken, in: *Hildebrandt, Achim/Wolf, Frieder* (Hrsg.), *Die Politik der Bundesländer. Staatstätigkeit im Vergleich*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 43-66.
- Laudel, Grit, 2006: *The Art of Getting Funded: How Scientists Adapt to their Funding Conditions*, in: *Science and Public Policy* 33, S. 489-504.
- Leszczensky, Michael/Orr, Dominic, 2004: *Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung. Dokumentation und Analyse der Verfahren in 11 Bundesländern*, HIS-Kurzinformationen A/2/2004, Hannover: HIS.
- 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.
- Meier, Frank/Schimank, Uwe, 2009: *Matthäus schlägt Humboldt? „New Public Management“ und die Einheit von Forschung und Lehre*, in: *Beiträge zur Hochschulforschung* 1/09, S. 42-61.
- Morris, Norma, 2004: *Scientists Responding to Science Policy. A Multi-Level Analysis of the Situation of Life Scientists in the UK*, Enschede: CHEPS.
- Musselin, Christine, 2000: *The Role of Ideas in the Emergence of Convergent Higher Education Policies in Europe: The Case of France*. Center for European Studies Working Paper Series No. 73, Boston: Harvard University.
- Münch, Richard, 2007: *Die akademische Elite. Zur sozialen Konstruktion wissenschaftlicher Exzellenz*, Frankfurt a.M.: Suhrkamp.
- Münch, Richard/Baier, Christian, 2009: *Die Konstruktion der soziologischen Realität durch Forschungs-rating*, in: *Berliner Journal für Soziologie* 19, S. 295-319.
- Newman, Melanie, 2009: *Future of Eight Liverpool Departments in Doubt*, in: *Times Higher Education* v. 9.3.2009. Quelle: <http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=405704&c=2> Zugriff: 22.05.2009).
- Parker, Jonathan, 2008: *Comparing Research and Teaching in University Promotion Criteria*, in: *Higher Education Quarterly* 62, S. 237-251.
- Pasternack, Peer, 2008: *Die Exzellenzinitiative als politisches Programm – Fortsetzung der normalen Forschungsförderung oder Paradigmenwechsel?*, in: *Bloch, Roland et al.* (Hrsg.), 2008: *Making Excellence. Grundlagen, Praxis und Konsequenzen der Exzellenzinitiative*, Bielefeld: Bertelsmann, S. 13-36.
- Reale, Emanuela/Barbara, Anna/Costantini, Antonio, 2007: *Peer Review for the Evaluation of Academic Research: Lessons from the Italian Experience*, in: *Research Evaluation* 16, S. 216-228.
- Roberts, Sir Gareth, 2003: *Review of Research Assessment*. London: Higher Education Funding Council for England, Quelle: <http://www.ra-review.ac.uk/reports/roberts.asp> (Zugriff 22.05.2009).
- Schimank, Uwe/Lange, Stefan, 2006: *Hochschulpolitik in der Bund/Länder-Konkurrenz*, in: *Weingart, Peter/Taubert, Niels C.* (Hrsg.), *Das Wissenschaftsministerium. Ein halbes Jahrhundert Forschungs- und Bildungspolitik in Deutschland*, Weilerswist: Velbrück Wissenschaft, S. 311-346.
- Schimank, Uwe/Schiene, Christoph, 2007: *Research Evaluation as Organizational Development: The Work of the Academic Advisory Council in Lower Saxony (FRG)*, in: *Whitley, Richard/Gläser, Jochen* (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht: Springer, S. 171-190.
- Schneider, J.W., 2009: *An Outline of the Bibliometric Indicator used for Performance-based Funding of Research Institutions in Norway*, in: *European Political Science* 8, S. 364-378.
- Science and Technology Committee*, 2004: *Research Assessment Exercise: a re-assessment*. Eleventh Report of the Session 2003-2004. HC586. Published on 23 September 2004 on the authority of the House of Commons. London: The Stationery Office Limited. Quelle: <http://www.parliament.the-stationery-office.co.uk/pa/cm200304/cmselect/cmsctech/586/586.pdf> (Zugriff 22.05.2009)
- Sondermann, Michael et al., 2008: *Die Exzellenzinitiative: Beobachtungen aus der Implementierungsphase*. IFQ-Working Paper 5, Bonn: IFQ.
- Tapper, Ted/Salter, Brian, 2003: *Interpreting the Process of Change in Higher Education: The Case of the Research Assessment Exercise*, in: *Higher Education Quarterly* 57, S. 4-23.
- Taylor, John, 2006: *The Evaluation of Research: Motives, Methods, and Misunderstandings*. Paper presented to the Research Group for Science, Technology and Innovation Studies (TaSTI), University of

- Tampere. Quelle: http://www.uta.fi/tasti/papereita/taylor_evaluation_of_research.pdf (Zugriff 22.05.2009).
- Van der Meulen, Barend*, 2007: Interfering Governance and Emerging Centres of Control: University Research Evaluation in the Netherlands, in: *Whitley, Richard/Gläser, Jochen* (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht: Springer, S. 191-202.
- VSNU/ NWO/ KNAW*, 2003: Standard Evaluation Protocol for Public Research Organizations, Utrecht/ Den Haag/ Amsterdam. Quelle: <http://www.quanu.nl./comasy/uploadedfiles/sep2003-2009.pdf> (Zugriff 22.05.2009).
- Westerheijden, Don F.*, 1997: A Solid Base for Decisions: Use of VSNU Research Evaluations in Dutch Universities, in: *Higher Education* 33, S. 397-413.
- Whitley, Richard*, 2007: Changing Governance of the Public Sciences: The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Fields, in: *Whitley, Richard/Gläser, Jochen* (eds.), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht: Springer, S. 3-27.

Anschriften der Autoren:

Dr. Stefan Lange, Deutsche Hochschule für Verwaltungswissenschaften Speyer, Stiftungslehrstuhl für Wissenschaftsorganisation, Hochschul- und Wissenschaftsmanagement, Freiherr-vom-Stein-Str. 2, 67346 Speyer.
E-Mail: lange@dhw-speyer.de

PD Dr. Jochen Gläser, Technische Universität Berlin, Zentrum Technik und Gesellschaft, Hardenbergstr. 36A, 10623 Berlin.
E-Mail: jochen.glaeser@ztg.tu-berlin.de