

Geeignete Ansätze zur Messung wissenschaftlicher Leistung¹

Ulrich Schmoch

Die Hochschulfinanzierung wird zunehmend von der Input- auf die Outputsteuerung umgestellt. Dies und die Stärkung der Hochschulleitungen bei der Mittelverteilung haben einen großen Bedarf ausgelöst, wissenschaftliche Performanz zu bewerten. Neben qualitativen Verfahren werden dabei auch in großem Umfang Performanzindikatoren genutzt, deren Eignung jedoch umstritten ist. Der Beitrag zeigt einen Ansatz auf, der über spezifische Indikatoren unterschiedliche Dimensionen von wissenschaftlicher Performanz erfassen kann. Darüber hinaus werden geeignete Verfahren der Auswertung erläutert, die diese Mehrdimensionalität erhalten. Abschließend werden Möglichkeiten einer breiteren Einführung dieser Ansätze in die Hochschulpraxis erörtert.

1 Einführung in die Fragestellung

In den letzten Jahrzehnten ist in führenden Industrieländern ein deutlicher Strukturwandel hin zu wissensbasierten Produkten und Dienstleistungen zu beobachten; gleichzeitig stehen nur begrenzte öffentliche Mittel zur Forschungsförderung zur Verfügung. Der Aufbau einer wissensbasierten Wirtschaft benötigt jedoch ein starkes wissenschaftliches Fundament. Von daher muss die Politik nach geeigneten Wegen suchen, trotz begrenzter Mittel eine gute wissenschaftliche Basis zu gewährleisten. Die naheliegendste Lösung dieses Dilemmas ist eine Steigerung der Effizienz der Forschung an den Hochschulen, doch sind die Wege zur Erreichung dieses Ziels umstritten.

Vor diesem Hintergrund haben viele Regierungen in Europa und auch in Übersee verschiedene Maßnahmen in die Wege geleitet, die mit dem Begriff des „New Public Management (NPM)“ assoziiert werden (*Schimank 2005*). Diese Strukturreform beinhaltet ein Bündel institutioneller Veränderungen und Maßnahmen, insbesondere eine Zurücknahme des direkten Staatseinflusses, verbunden mit einer Stärkung der Leitungen der Universitäten sowie eine Umstellung bei der Finanzierung von der Input- zur Output-Steuerung. Bei der Input-Steuerung werden die Aktivitäten durch die Höhe

¹ Der vorliegende Beitrag beruht auf Ergebnissen des Projekts „Performanz-Indikatoren für Forschungseinrichtungen, insbesondere Forschungsgruppen“, welches von der deutschen Forschungsgemeinschaft gefördert wird (SCHM 1719/1–2). Das Projekt ist Teil der Arbeit einer interdisziplinären Forschergruppe zum Thema „Internationale Wettbewerbs- und Innovationsfähigkeit von Universitäten und Forschungsorganisationen – Neue Governanceformen (FOR 517)“. Ich danke Dr. Torben Schubert für wertvolle Diskussionen und Anregungen.

der bereitgestellten Grundmittel gesteuert, deren ordnungsgemäße Verwendung vom Staat – in Deutschland von den Ländern – kontrolliert wird. Bei der Output-Steuerung ist demgegenüber die Vergabe zukünftiger Mittel an die aktuelle wissenschaftliche Performanz von Forschungseinheiten gekoppelt, die jetzt jedoch von den Hochschulleitungen überprüft wird.²

Diese neuen Strukturen haben sich auch in Deutschland je nach Bundesland mehr oder weniger stark durchgesetzt (*Lesczensky/Orr 2004; Schubert 2008*); der Trend zu einer breiten, nachhaltigen Einführung ist in jedem Fall unübersehbar. Die Stärkung der Hochschulleitungen, verbunden mit einer Output-Steuerung, hat zur Folge, dass diese in zunehmendem Maße Informationen über die wissenschaftliche Performanz ihrer Forschungseinheiten benötigen, wobei es sich bei großen Universitäten auf der Ebene von Instituten und Lehrstühlen um mehrere hundert Einheiten handeln kann. Eine unmittelbare Folge dieser stärkeren Erfassung wissenschaftlicher Performanz ist eine steigende Zahl von Evaluationen, die traditionell rein qualitativ von Fachexperten im Wege des sogenannten Peer Review durchgeführt werden. Da diese qualitativen Evaluationen sehr aufwendig sind, setzt sich in der Praxis immer mehr die Erhebung von Leistungsindikatoren durch und damit die quantitative Messung von wissenschaftlicher Performanz.

Schon die qualitativen Evaluationen sind umstritten, da es bei den Evaluatoren mehr oder weniger versteckte Interessenkonflikte geben kann und die Subjektivität der Bewertung nicht unerheblich ist (*Bornmann/Daniel 2003; Kuhlmann/Heinze 2004a/b*). Außerdem fehlt es bei der steigenden Zahl von Evaluationen an kompetenten Evaluatoren; deren Knappheit ist mittlerweile zu einem Problem geworden. Die Eignung von quantitativen Indikatoren ist aber noch umstrittener. Bemängelt wird hier, dass viele Indikatoren nicht einheitlich für alle Disziplinen anwendbar sind. Ein gutes Beispiel sind hier die Drittmittel, die sich in technischen Fächern einfacher als in geisteswissenschaftlichen akquirieren lassen. Eine weitere, noch grundlegendere Kritik besteht darin, dass durch die Auswahl spezifischer Indikatoren nur ein enger Ausschnitt wissenschaftlicher Performanz erfasst werde und damit andere wesentliche Aspekte nicht berücksichtigt würden.

Der vorliegende Beitrag befasst sich mit der Problematik, wie wissenschaftliche Performanz auf quantitativem Wege adäquat erfasst werden kann, etwa durch eine geeignete Auswahl von Indikatoren oder angemessene Verfahren der Auswertung. Damit fokussieren die Ausführungen auf die Performanz in der Forschung und lassen die der Lehre unbeachtet. Dies soll jedoch nicht bedeuten, dass die Leistungen in der Lehre unwichtig wären. Angesichts des schon aktuell spürbaren Mangels an hoch

² Nähere Details hierzu finden sich in *Schubert/Schmoch (2009a)* und insbesondere in *Schubert (2008)*.

qualifizierten Fachkräften, die eine wesentliche Voraussetzung für eine wettbewerbsfähige wissensbasierte Wirtschaft sind, wird die Lehre vielmehr an Bedeutung gewinnen (*Achleitner et al. 2008, S. 36 ff., 98 ff.*). Die adäquate Erfassung der Performanz in der Lehre erfolgt jedoch nach anderen Kriterien, weshalb ihre gleichzeitige Behandlung den Rahmen eines Aufsatzes sprengen würde.

2 Dimensionen wissenschaftlicher Performanz

Das wesentliche Ziel wissenschaftlicher Forschung ist die Genese neuen Wissens (*Weingart 2003*), welche nicht direkt erfassbar ist, wohl aber indirekt über Fachpublikationen, Patente, Prototypen und Ähnliches. Hier hat sich insbesondere die Auswertung der Quantität und des Einflusses von Fachpublikationen in Form der Bibliometrie durchgesetzt, was für die Mehrzahl der Disziplinen zu durchaus angemessenen Ergebnissen führt (*Moed 2005*). Einschränkend ist jedoch festzuhalten, dass in vielen Disziplinen der Geistes- und Sozialwissenschaften neben Zeitschriften-Artikeln auch Monographien und Buchbeiträge wichtige Publikationsformen sind, die aber in den gegenwärtig verfügbaren bibliometrischen Datenbanken nicht berücksichtigt werden (*Hicks 2004*). Letztere haben darüber hinaus eine starke Orientierung auf englischsprachige Fachzeitschriften, was aus deutscher Perspektive für die Geistes- und Sozialwissenschaften, aber auch die Ingenieurwissenschaften zu einer unzureichenden Erfassung des wissenschaftlichen Outputs führt (*Schmoch 2004*). So ist festzuhalten, dass in diesen spezifischen Bereichen bibliometrische Verfahren (noch) zu Fehlinformationen führen.

Eine weitere zentrale Dimension wissenschaftlicher Performanz ist die Ausbildung von Doktoranden und Habilitanden, also des Nachwuchses an hoch qualifizierten Forschern. Diese Leistung ist in einem Verlaufsmodell der wissenschaftlichen Forschung zwar nur ein Input oder eventuell ein intermediärer Output. Dennoch ist diese Leistung entscheidend für die mittel- und langfristige Gewährleistung relevanter Forschungsergebnisse. Die Ausbildung des Forschernachwuchses ist bei der Performanzerfassung in jedem Falle zu berücksichtigen, da die betroffenen Hochschullehrer und Arbeitsgruppen durch diese Aktivitäten weniger Zeit in die unmittelbare Forschung investieren können.

Eine Leistung, die den dauerhaften Bestand der wissenschaftlichen Forschung gewährleistet, ist das Engagement im Bereich der Infrastruktur der Wissenschaft, etwa über leitende Funktionen in wissenschaftlichen Vereinigungen, oder in der wissenschaftlichen Kommunikation, beispielsweise über die Herausgeberschaft von Zeitschriften. Auch diese Tätigkeiten sind zeitintensiv und begrenzen die Möglichkeit einer unmittelbaren Forschungstätigkeit.

Schließlich ist zu berücksichtigen, dass in den letzten Jahren der Technologietransfer neben Forschung und Lehre als wichtige „Dritte Mission“ der Hochschulen betrachtet wird, was im Bayerischen Hochschulgesetz³ in Art. 2 Abs. 5 auch explizit festgeschrieben ist. Da der Technologietransfer eng mit der wissenschaftlichen Forschung verbunden ist, kann er als weitere Dimension wissenschaftlicher Performanz betrachtet werden. In diesem Kontext gibt es eine wissenschaftliche Debatte, ob transferaktive Hochschulgruppen gleichzeitig auch solche mit hoher Leistung im Bereich wissenschaftlicher Publikationen sind, oder ob hier eher ein Widerspruch besteht (*Perkman/Walsh 2007*). In jedem Fall gibt es eine Reihe von Forschungseinheiten mit hoher Transferleistung, die bei einer reinen Betrachtung des Publikationsoutputs nicht adäquat bewertet würden.

Sicherlich lässt sich noch eine Vielzahl von anderen Dimensionen wissenschaftlicher Performanz finden, die aber meist mit den besonderen Eigenheiten einer Disziplin verknüpft sind. Schon bei den hier beschriebenen zentralen Dimensionen wird deutlich, dass es eine zeitliche Konkurrenz gibt und daher die einseitige Förderung einer Dimension zu einem Ungleichgewicht im Gesamtsystem führen würde. So implizieren einseitige Anreize zugunsten des Publikationsoutputs eine Vernachlässigung von Infrastrukturaufgaben, der Nachwuchsausbildung oder der wissenschaftlichen Kommunikation, was für das Gesamtsystem auf längere Sicht problematisch wäre. Die Erfassung spezifischer Indikatoren, an die die Höhe der Mittelvergabe geknüpft ist, kann solche Fehlanreize setzen. Dafür ist die Bindung der Mittelzuweisung an die Zahl der Publikationen im australischen Universitätssystem ein bekanntes Beispiel. Diese hat zwar zu einem Anwachsen der Publikationszahlen, gleichzeitig aber zu einem Rückgang der Forschungsqualität geführt (*Butler 2004*). Ein weiterer illustrativer Fall ist die britische „Research Assessment Exercise (RAE)“, bei der die Qualität des Publikationsoutputs in den Mittelpunkt gestellt wird. Diese einseitige Orientierung auf den unmittelbaren Forschungsoutput führte zu einer Vernachlässigung von Infrastrukturleistungen und von Technologietransfer, was durch die schrittweise Einführung spezieller Förderprogramme in diesen Bereichen kompensiert werden musste (*Brennan 2008*).

Auch auf der Ebene einzelner Hochschullehrer und Forschergruppen ist die Fokussierung auf eine einzelne Performanzdimension problematisch, da sich etwa ein Hochschullehrer mit Stärken in der Nachwuchsausbildung oder beim Transfer durch entsprechende Anreize auf Grundlagenforschung konzentriert und damit letztlich suboptimal arbeitet. Im Sinne einer optimalen Nutzung individueller Leistungen und einer Effizienz des Gesamtsystems ist es deshalb wichtig, Anreize für alle zentralen Dimensionen wissenschaftlicher Performanz zu schaffen, was insbesondere bei Indikatorsystemen im Kontext der Output-Steuerung von Forschung zu berücksichtigen ist.

³ vom 23. Mai 2006 (GVBl. S. 245)

Denn die Auswahl von Indikatoren hat eine starke Signalwirkung und setzt spezifische Anreize für wissenschaftliche Aktivität.

3 Datengrundlage

Die vorliegende Untersuchung beruht auf einer E-Mail-Befragung von deutschen Forschungsgruppen, die im Jahr 2007 in den Feldern Astrophysik, Nanotechnologie, Biotechnologie und Volkswirtschaft durchgeführt wurde. Die Befragung hatte den Zweck, Performanzindikatoren aus den vier vorgenannten Bereichen abzufragen, die für unterschiedliche Eigenschaften und Ausrichtungen der wissenschaftlichen Forschung stehen, insbesondere Grundlagenorientierung (Astrophysik, Volkswirtschaft) und Anwendungsorientierung (Biotechnologie, Nanotechnologie) sowie Natur- und Sozialwissenschaften.

Insgesamt wurden dafür 1.935 Leiter von Forschungsgruppen angeschrieben mit einem Rücklauf von 473 Antworten, also mit einer Rücklaufquote von etwa 25 Prozent. Die Forschungsgruppen wurden mithilfe der bibliometrischen Datenbank „Science Citation Index (SCI)“ im Bereich der Biotechnologie und Nanotechnologie identifiziert. In den Fällen der Volkswirtschaft und Astrophysik wurde die institutionelle Datenbank „Vademecum“ hinzugezogen. Zusätzlich wurden Recherchen im Internet durchgeführt, um mögliche Verzerrungen der benutzten Datenbanken auszuschließen. In den Fällen der Astrophysik und Volkswirtschaft entspricht die Zahl der Befragten nahezu einer Vollerhebung, da die Zahl der diesbezüglichen Forschungseinheiten beschränkt ist. In der Biotechnologie und der Nanotechnologie ist die Zahl der befragten Einheiten zwar sehr hoch; es handelt sich aber letztlich nur um Stichproben innerhalb einer größeren Gesamtpopulation. Tabelle 1 dokumentiert die Einzelheiten zum Umfang der Erhebung in verschiedenen Feldern.

Tabelle 1: Struktur der Erhebungsstichprobe

	Befragte Einheiten	Auswertbare Antworten	Rücklaufquote (%)
Astrophysik	97	34	35
Nanotechnologie	674	201	30
Volkswirtschaft	477	102	21
Biotechnologie	687	136	20
Gesamt	1935	473	24

In der Umfrage wurden verschiedene Strukturvariablen wie die Gruppengröße oder Finanzierungsquellen abgefragt, die den einzelnen Forschungsgruppen bekannt und mit begrenztem Aufwand zugänglich waren, sodass sinnvolle und sachgerechte Ant-

worten zu erwarten waren. Darüber hinaus wurden Angaben zur Governance-Struktur und zur Durchsetzung von Maßnahmen des New Public Management erhoben, die im vorliegenden Kontext weniger relevant sind. In Bezug auch auf mögliche Performanzindikatoren wurden in der Frühphase des Projekts eine breite Literaturrecherche durchgeführt und die dort vorgeschlagenen Indikatoren gesammelt, was zu einer umfangreichen Liste führte. In einer begrenzten Befragung wurden dann Experten aus den vier Feldern gebeten, die Relevanz der vorgeschlagenen Indikatoren für ihre Disziplin zu bewerten. Hieraus ergab sich eine relativ große Übereinstimmung hinsichtlich der wichtigsten Indikatoren. Mit einer entsprechend reduzierten Liste wurde dann zunächst eine kleinere Befragung von 77 Einheiten durchgeführt, wobei die Performanzindikatoren im Rahmen von Interviews ermittelt wurden. Dies führte zu einer teilweise präzisierten Formulierung einzelner Indikatoren, die in die breitere E-Mail-Erhebung eingingen. Vor der Durchführung der breiten E-Mail-Befragung wurde ein Pretest durchgeführt, um sicherzustellen, dass die Fragen richtig verstanden und beantwortet werden. Tabelle 2 zeigt in einer Übersicht die erhobenen Performanzindikatoren und ihre jeweiligen Ausprägungen.

Tabelle 2: Übersichtsstatistik zu den Performanzindikatoren

Variable	Zahl der Antworten	Mittelwert	Standardabweichung	Minimum	Maximum
Abgeschl. Promotionen	471	3.772824	4.454663	0	52
Abgeschl. Habilitationen	471	.4416136	.9633103	0	13
Berufungen	471	.3715499	.8505052	0	10
Monographien	470	1.802128	6.57426	0	90
Herausgeberschaften	471	1.11465	2.026321	0	20
Unternehmensprojekte	471	3.494692	12.75061	0	200
Unternehmensgutachten	471	.8322718	3.301112	0	50
Mitgl. wiss. Gremien	470	3.271277	4.147767	0	30
Publikationen	471	31.47558	40.83947	0	320
Zitate	450	143.4911	258.6801	0	3030
Int. Kopublikationen	450	13.03556	19.64526	0	136

Die Zahl der letztendlich erforderlichen Indikatoren für die Erfassung der wesentlichen Performanzdimensionen ist somit begrenzt, was im Hinblick auf die Präzision der Antworten und die Rücklaufquote einer solchen Befragung von großem Vorteil ist. Die Einheiten wurden auch zur Zahl ihrer Publikationen befragt; es wurden jedoch zusätzlich Datenbankrecherchen im Science Citation Index für die Felder Astrophysik, Biotechnologie und Nanotechnologie sowie in Scopus für die Volkswirtschaft durchgeführt, um auch Zitate und internationale Kopublikationen erheben zu können. Beim Vergleich der Ergebnisse aus der Fragebogenerhebung und der Datenbankrecherche ergaben

sich bei den Publikationen zum Teil deutliche Abweichungen, die zum einen auf eine unsystematische Dokumentation bei den Einheiten, zum anderen auf die Angabe von Publikationen in weniger hochrangigen Zeitschriften zurückzuführen sind. Hier wurde letztlich die Entscheidung getroffen, nur die Datenbankergebnisse zu verwenden, um einheitliche Kriterien der Publikationszählung für alle betrachteten Einheiten zu gewährleisten.

Eine Verzerrung der Antworten im Sinne sozialer Erwünschtheit ist grundsätzlich möglich, aber wir gehen von einer validen Erhebung aus, weil sich die meisten Fragen auf klare Sachverhalte beziehen, die in der Regel nicht mit Performanz in Verbindung gebracht werden (Tabelle 2). Gewisse Möglichkeiten der „Optimierung“ bestehen lediglich bei den Publikationsindikatoren, die aber, wie oben beschrieben, nicht auf Eigenangaben, sondern Datenbankauswertungen beruhen. Im Übrigen wurde bei der Erhebung Vertraulichkeit zugesichert, und aus den aggregierten Ergebnissen kann in der Tat nicht auf einzelne Einheiten zurück geschlossen werden.

4 **Aktivitätsprofile**

Bei der Auswertung des Datensatzes ist zunächst zu beachten, dass sich die Ausprägung der einzelnen Indikatoren zwischen den Feldern erheblich unterscheidet, sodass ein unmittelbarer Vergleich nicht möglich ist. Beispielsweise ist die Zahl der Publikationen pro Wissenschaftler in den Feldern Astrophysik, Nanotechnologie und Biotechnologie relativ ähnlich mit einem Niveau von etwa 2,5; sie liegt dagegen in der Volkswirtschaft mit 0,2 um den Faktor 10 niedriger. Dagegen ist die Zahl der Habilitationen je Wissenschaftler in der Volkswirtschaft mit 0,1 etwa dreimal so hoch wie in der Nanotechnologie mit 0,03 (*Jansen et al. 2007*). Um eine gemeinsame Auswertung der Daten aus den verschiedenen Teilfeldern in sinnvoller Weise durchführen zu können, wurde deshalb eine Z-Transformation mit Referenz auf die Durchschnittswerte der jeweiligen Felder durchgeführt.

Eine weitere wichtige Frage ist, wie in der Auswertung die Mehrdimensionalität des Outputs behandelt wird. Viele Autoren greifen in diesem Zusammenhang auf sogenannte „composite indicators“ zurück, bei denen die verschiedenen Dimensionen standardisiert und dann miteinander verrechnet werden, sodass am Ende eine einzige Maßzahl zur Bewertung der jeweiligen Untersuchungseinheit entsteht. In dieser Maßzahl ist allerdings nicht mehr sichtbar, welcher Faktor zu einer positiven oder negativen Bewertung geführt hat. Außerdem sind solche integrierten Indikatoren anfällig gegenüber Gewichtungen der Einzelkomponenten, deren Effekt häufig nicht genügend beachtet wird (*Grupp/Mogee 2004; Schubert 2008, S. 86 ff.*).

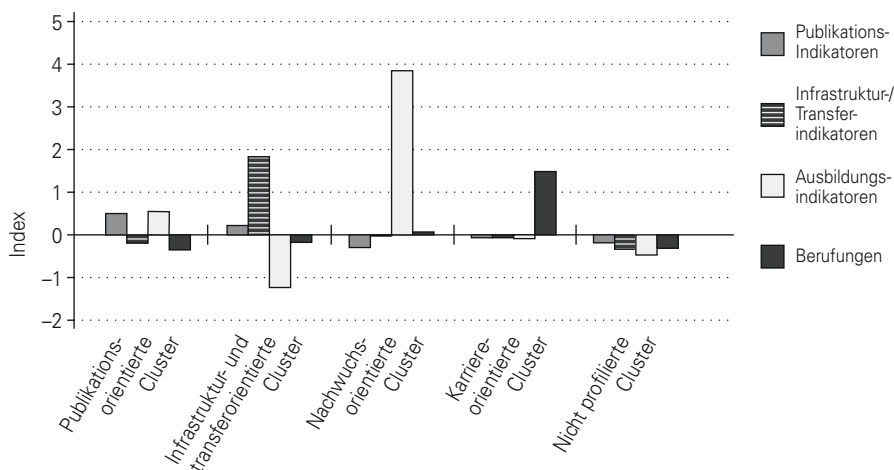
Für die vorliegende Untersuchung wurde in einem ersten Schritt eine Faktorenanalyse über sämtliche erhobenen Output-Maße durchgeführt, um so relevante Indikatoren bzw. Kombinationen von Indikatoren zu ermitteln. Als erster Faktor wurden dabei Publikationsindikatoren ermittelt, in die insbesondere die Zahl der Publikationen (pro Wissenschaftler) und die Zitatquoten eingehen.⁴

Der zweite Faktor führte auf verschiedene Infrastruktur- und Transferindikatoren. Hier zeigte sich, dass Forschungseinheiten, die stark in der Organisation und Kommunikation der wissenschaftlichen Gemeinschaft engagiert sind, häufig auch eine deutliche Transferorientierung auf Unternehmen aufweisen. Dieses letztgenannte Ergebnis sollte allerdings nicht überbewertet werden, da in allen untersuchten Feldern, entgegen den Erwartungen auch in der Biotechnologie und der Nanotechnologie, die Orientierung auf die Grundlagenforschung ausgeprägt ist und der Technologietransfer zu Unternehmen eine untergeordnete Rolle spielt. Von daher ist die hier vorliegende Datenbasis zur statistischen Auswertung der Rolle des Technologietransfers begrenzt.

Die weitere Infrastrukturleistung der Ausbildung des wissenschaftlichen Nachwuchses erwies sich als separater Faktor, ebenso wie die Zahl der Berufungen an externe Universitäten, die in der Literatur häufig als Leistungsindikator angeführt wird, und auch bei der frühen Befragung von Fachvertretern als relevanter Indikator identifiziert wurde.

Aufbauend auf diesen Faktoren wurden in dem Datensatz für die Forschungseinheiten verschiedene Cluster wissenschaftlicher Performanz mit unterschiedlichen Aktivitätsprofilen ermittelt. In der entsprechenden grafischen Darstellung zeigt – aufgrund der oben erwähnten Z-Transformation – ein Index mit dem Wert 0 eine durchschnittliche Aktivität in Relation zur Gesamtstichprobe an (Abbildung 1), positive Werte indizieren eine überdurchschnittliche, negative eine unterdurchschnittliche. Beispielsweise haben im „nachwuchsorientierten“ Cluster die Ausbildungsindikatoren einen positiven Wert von fast 4 und liegen damit deutlich über dem Durchschnitt (Abbildung 1).

⁴ Alle absoluten Werte wie Publikationen wurden jeweils auf die Gruppengröße bezogen, um Verzerrungen durch Größeneffekte auszuschließen. Dies ist der Übersichtlichkeit halber bei den übrigen Indikatoren nicht mehr ausdrücklich erwähnt.

Abbildung 1: Forschungsgruppentypen nach Output-Dimensionen

Innerhalb des „publikationsorientierten Clusters“ sind neben den Publikationsindikatoren selbst auch die Ausbildungsindikatoren überdurchschnittlich ausgeprägt. Dieses entspricht den Erwartungen, da bei Doktorandinnen und Doktoranden mit einem besonders hohen Interesse an der Publikation ihrer Forschungsergebnisse zu rechnen ist. Im „publikationsorientierten“ Cluster sind dagegen die Infrastruktur/Transferindikatoren leicht unterdurchschnittlich (Abbildung 1). Bemerkenswert ist bei diesem Cluster, dass das besondere Merkmal der hohen Publikationsaktivität weniger stark ausgebildet ist als die spezifischen Merkmale der anderen Cluster. Offensichtlich diskriminiert die Publikationsaktivität in der Wissenschaft weniger stark als erwartet, denn auch in den andern Clustern liegen die Publikationsaktivitäten in der Nähe des Durchschnittswerts und sind insbesondere in keinem Cluster deutlich unterdurchschnittlich.

Im „infrastruktur- und transferorientierten“ Cluster ist vor allem die Dimension „Infrastruktur/Transfer“ stark überdurchschnittlich, dagegen ist die der Ausbildung stark unterdurchschnittlich vertreten (Abbildung 1). Die deutliche Transferorientierung lässt sich offensichtlich schlecht mit einer relevanten Ausbildung von Doktorandinnen und Doktoranden vereinbaren.

Im „nachwuchsorientierten“ Cluster ist der Fokus auf den Ausbildungsindikatoren extrem überdurchschnittlich; die Publikationsindikatoren fallen dagegen unterdurchschnittlich aus. Während im „publikationsorientierten“ Cluster die Indikatoren zu „Publikationen“ und „Ausbildungen“ beide leicht überdurchschnittlich sind, also miteinander kompatibel sind, ist im „nachwuchsorientierten“ Cluster offensichtlich eine starke Ausbildungsorientierung nicht mit hohen Publikationsaktivitäten vereinbar

(Abbildung 1). Infrastruktur/Transfer ist in diesem Cluster dagegen durchschnittlich vertreten.

Für die Forschungseinheit mit einer hohen Berufungszahl wurde ein eigener Cluster ermittelt, der etwas plakativ als „karriereorientiert“ bezeichnet wurde. Hier zeigt sich, dass die drei Leistungsdimensionen Publikationen, Infrastruktur/Transfer und Ausbildung lediglich durchschnittlich ausgeprägt sind (Abbildung 1). So ist nicht, wie vermutet werden könnte, bei den Publikationen eine überdurchschnittliche Performanz zu beobachten. Dieses flache Profil dürfte damit zusammenhängen, dass bei Berufungen eine Performanz in allen Dimensionen erwartet wird und nicht eine einseitige Betonung einer Dimension. Der Cluster spiegelt also letztlich die Praxis üblicher Berufungsverfahren und eignet sich somit weniger als wirklicher Performanzindikator. Es wird allenfalls gezeigt, dass die jeweiligen Gruppenleiter eine gute Außendarstellung haben.

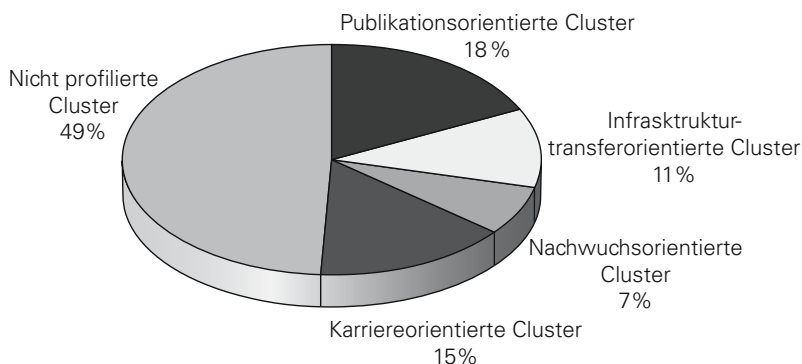
Diese allgemeine Aussage zu Berufungen als Performanzindikator soll nicht bedeuten, dass bei allen Berufungen die Publikationsleistung nur durchschnittlich wäre. Bei der Clusterbildung steht die Profilähnlichkeit und nicht das absolute Niveau der Indikatoren im Vordergrund. Es gibt somit bei den Berufungen eine Reihe von Fällen mit einem hohen Publikationsniveau, bei denen dann in der Regel auch die anderen Indikatoren überdurchschnittlich sind. In Abbildung 1 sind die Durchschnittswerte für den jeweiligen Cluster und nicht die Profile aller Einheiten dargestellt.

In dem letzten Cluster, der als „nicht -profilierter“ etikettiert wurde, sind alle Dimensionen leicht unterdurchschnittlich ausgeprägt. Dieser Cluster ließe sich polemisch mit Einheiten „unprofilierter Faulenzer“ erklären, wie es sicher einigen Kritikern des wissenschaftlichen Systems ins Konzept passen würde (*Kamenz/Wehrle 2007*). Für die „nicht -profilierter“ Einheiten sind zwar keine Spezialisierungsvorteile hinsichtlich einer spezifischen Dimension wissenschaftlicher Performanz festzustellen; aus Sicht der Lehre kann es aber durchaus vorteilhaft sein, wenn Aktivitäten in allen Dimensionen vorhanden sind, die Lehrenden somit über breite Erfahrungen aus verschiedenen Bereichen verfügen. Allerdings sind gerade im Cluster der nicht Profilierten viele Einheiten mit unterdurchschnittlichen Werten in allen Dimensionen zu finden (*Schmoch/Schubert 2008*).

Rein quantitativ entfällt auf den Cluster der nicht Profilierten mit 49 Prozent aller Antworten fast die Hälfte aller analysierten Forschungseinheiten (Abbildung 2). Die meisten Forschungseinheiten bilden also kein dezidiertes Spezialisierungsprofil heraus und stellen den Regelfall dar. Lediglich 18 Prozent entfallen auf den publikationsorientierten Cluster, elf Prozent auf den transfer- und infrastrukturorientierten und nur sieben Prozent auf den nachwuchsorientierten. Da beim karriereorientierten Cluster das Profil bei den anderen Performanzdimensionen flach ist, würden diese

Forschungseinheiten bei Herausnahme dieses Kriteriums überwiegend dem „nicht profilierten“ Cluster zugeordnet.

Abbildung 2: Verteilung der verschiedenen Forschungsgruppentypen in der Erhebungsstichprobe auf die Cluster



Auf den ersten Blick mag es erstaunlich erscheinen, dass bei den Performanzindikatoren nicht die Drittmiteleinahmen vertreten sind, die in der aktuellen Praxis in vielen Bundesländern als zentraler Leistungsindikator erhoben werden (*Leszczensky/Orr 2004*). Diese zentrale Stellung des Drittmittelindikators in der Praxis ist letztlich darauf zurückzuführen, dass in den letzten Jahren die Grundmittel allmählich reduziert wurden, was nur durch einen ständigen Zuwachs der Drittmittel kompensiert werden konnte. So betrug Anfang der 1980er-Jahre der Anteil der Drittmittel an den gesamten Forschungsmitteln der Universitäten 23 Prozent; dieser Wert ist aktuell auf 40 Prozent angewachsen (*Schmoch 2008*). Das Erreichen einer hohen Drittmittelquote ist deshalb aus Sicht der Universitäten ein zentrales Ziel, um die Forschungsfinanzierung zu sichern. Aus Sicht der wissenschaftlichen Performanz stellt sich dieses jedoch anders dar, da letztlich Drittmittel ein Instrumentalziel im Hinblick auf das finale Ziel der Wissensgenese sind. Wenn in diesem Sinne Drittmittel als alternativer Indikator für Publikationen genutzt werden sollen, ist eine notwendige Anforderung, dass für beide Indikatoren über den gesamten Wertebereich ein monotoner positiver Zusammenhang besteht. Dies ist aber nicht der Fall, da ein gleichzeitiger Zuwachs von Drittmitteln und Publikationsoutput nur im unteren Bereich feststellbar ist. Bei Überschreiten eines kritischen Werts, der für jedes Themenfeld spezifisch ist, sinken die Publikationen mit steigenden Drittmittelquoten wieder ab. Dies kann darauf zurückgeführt werden, dass bei sehr hohen Drittmittelquoten der Aufwand zur Akquisition zusätzlicher Drittmittel so hoch ist, dass nicht mehr genügend Zeit für Forschung verbleibt. Ein weiterer Aspekt dürfte sein, dass mit wachsender Drittmittelquote die Transaktionskosten innerhalb der Forschungsgruppe ebenfalls ansteigen.

5 Kontinuierliches Monitoring wissenschaftlicher Performanz

Die vorliegende Untersuchung bezieht sich auf einen beschränkten Datensatz und kann nur einige grundsätzliche Phänomene wissenschaftlicher Performanz illustrieren. Die vier Themenfelder wurden in ihrer Ausrichtung bewusst sehr unterschiedlich gewählt, können aber das gesamte Spektrum der wissenschaftlichen Disziplinen nicht abbilden. Insgesamt wird jedoch deutlich, dass die Erfassung unterschiedlicher Dimensionen wissenschaftlicher Performanz sinnvoll ist, um das gesamte Performanzspektrum zu erfassen und nicht durch falsche Prioritätensetzung einzelne durchaus leistungsfähige Forschergruppen auszugrenzen.

Es stellt sich jedoch die Frage, wie ein solches Indikatorensystem im hochschulpolitischen Alltag implementiert werden kann. Zunächst einmal ist festzustellen, dass schon jetzt verschiedene Indikatoren erhoben werden, wobei die konkrete Auswahl je nach Bundesland unterschiedlich ist und oftmals die berechtigte Frage gestellt werden kann, ob die Auswahl adäquat ist. Bei der Betrachtung der Indikatorenliste, die für die vorliegende Untersuchung verwandt wurde, wird deutlich, dass die Mehrzahl der Variablen ohne großen Aufwand zu erheben ist und von den Fachbereichen oder Universitätsleitungen leicht in jährlichem Rhythmus abgefragt werden könnte. Dazu gehören Strukturvariablen wie die Größe und Zusammensetzung der Forschergruppe, die Anzahl der abgeschlossenen Promotionen oder der Berufungen oder die Herausgeberschaft von Zeitschriften. Aufwendiger, letztlich aber unverzichtbar, ist die Ermittlung der Anzahl der Publikationen und der entsprechenden Zitate in Bezug auf die Forschergruppe. Hier hat es sich – wie oben dargelegt – als problematisch erwiesen, sich auf die Angaben der Forschergruppen selbst zu verlassen, da das Qualitätsniveau der dabei gemeldeten Publikationen oft sehr unterschiedlich ist und im Übrigen eine Ermittlung der Zitate nicht realisiert werden kann. Von daher bleibt nur der Rückgriff auf geeignete Datenbanken wie Web of Science oder Scopus, wobei die Analysen von fachkundiger Seite realisiert werden müssen. Für Fächer, in denen solche Analysen problematisch sind, müssen hier andere Lösungen erarbeitet werden, wie das in den Wirtschaftswissenschaften bereits auf hohem Standard der Fall ist. Dort werden die wichtigsten Fachzeitschriften nach ihrem wissenschaftlichen Niveau verschiedenen Kategorien zugeordnet und die einzelnen Publikationen entsprechend bewertet, was als Ersatz für Zitatanalysen genutzt werden kann (*Schneider/Ursprung 2008*).

Voraussetzung für eine solche regelmäßige Datenerhebung ist, dass sie nicht nur an einer einzelnen Universität, sondern an mehreren parallel nach gleichen Kriterien durchgeführt wird, um so für die einzelnen Disziplinen aussagekräftige Durchschnittswerte für die einzelnen Performanzdimensionen zu erhalten. Diese Einbeziehung einer größeren Zahl von Forschungsgruppen pro Fach war in der vorliegenden Untersuchung

die Voraussetzung für die Durchführung einer Z-Transformation und damit einer Standardisierung mit den Durchschnittswerten.

Liegt ein solcher Datensatz für eine größere Zahl von Universitäten vor, ist es möglich, fachspezifisch die durchschnittliche Ausprägung der verschiedenen Performanzprofile zu ermitteln und damit sinnvolle Benchmarks zu generieren. Innerhalb eines Profils ist es dann relativ einfach, schon durch bloße qualitative Anschauung Gruppen mit hoher oder niedriger Performanz zu identifizieren. Diese sind dann im Sinne der leistungsorientierten Mittelverteilung potentielle Kandidaten für Mittelserhöhungen oder -reduzierungen. Insbesondere bei den Forschungsgruppen, die auf den ersten Blick stark unterdurchschnittliche Werte aufweisen, ist es aber erforderlich, eine genauere qualitative Bewertung im Sinne eines Peer Reviews durchzuführen. Denn durch die vorgeschlagene Analyse werden zwar wichtige Dimensionen der wissenschaftlichen Performanz abgebildet; es ist aber durchaus möglich, dass die Gruppen bei anderen Aspekten stark sind, die bei der beschränkten Indikatorenzahl nicht berücksichtigt wurden. Es wäre jedoch kein geeigneter Weg, diese Fälle durch eine Erhöhung der Zahl der Indikatoren besser zu erfassen, da sich dann die Zahl der Indikatoren leicht auf über 100 erhöhen würde und eine sinnvolle Auswertung nicht mehr möglich wäre. Ein zweiter wichtiger Aspekt der qualitativen Evaluation ist darüber hinaus, dass nicht nur Leistungen in der Vergangenheit berücksichtigt werden, sondern dass die Evaluatoren auch gute Pläne für zukünftige Forschungsvorhaben würdigen können und so den Forschergruppen einen Vertrauensvorschuss geben können, was bei einer reinen Indikatorenauswertung nicht möglich ist.

Die Indikatorenauswertung ist in jedem Fall geeignet, Gruppen mit einem durchschnittlichen Performanzprofil zu ermitteln, bei denen sich in der Mittelverteilung auch nach einer vertieften Evaluation nichts Entscheidendes ändern würde. Damit ist es möglich, qualitative Evaluationen auf die Forschergruppen mit überdurchschnittlichen oder unterdurchschnittlichen Performanzindikatoren zu konzentrieren und damit die Zahl der qualitativen Evaluationen insgesamt deutlich zu reduzieren. Es geht also nicht darum, qualitative Verfahren vollständig durch quantitative zu ersetzen, sondern vielmehr um eine Fokussierung der qualitativen Verfahren auf sinnvolle Fälle.

6 Schlussfolgerungen

Die Untersuchung auf der Basis einer großen Umfrage für verschiedene Fächer zeigt, dass wissenschaftliche Performanz zumindest zu großen Teilen durchaus messbar ist, wobei aber unterschiedliche Dimensionen berücksichtigt werden müssen. Dabei ist darauf zu achten, dass neben Indikatoren für den Output der wissenschaftlichen Forschung auch intermediäre Aspekte berücksichtigt werden, die auf die Infrastruktur der wissenschaftlichen Forschung und damit auf ihren mittel- und langfristigen Bestand

gerichtet sind. Dazu gehören auf der einen Seite die Ausbildung des wissenschaftlichen Nachwuchses, die schon heute in vielen Bundesländern abgefragt wird, auf der anderen Seite aber auch die Organisation der wissenschaftlichen Kommunikation und das Engagement in wissenschaftlichen Vereinigungen, ein Aspekt, der in der Regel nicht beachtet wird.

Bei der Auswertung sollten Verfahren eingesetzt werden, bei denen die Ausprägungen in den verschiedenen Performanzdimensionen erhalten bleiben und nicht zu einfachen Ergebniswerten verrechnet werden. Vielmehr geht es darum, typische Performanzprofile zu ermitteln, um die Spezialisierung von Forschergruppen in einzelnen Bereichen angemessen würdigen zu können.

Eine Erhebung der Mehrzahl der Indikatoren zu den verschiedenen Performanzdimensionen ist in der Praxis mit einem begrenzten Aufwand realisierbar, wobei die Generierung eines größeren Datensatzes wünschbar wäre, um fachspezifische Durchschnittswerte ermitteln zu können. Neben der Beurteilung der Profile auf der Ebene einzelner Hochschulen kann dieser Datensatz eine sinnvolle Information für disziplinäre Fachgemeinschaften sein, um sinnvolle Relationen zwischen den einzelnen Profilen zu diskutieren, etwa den Anteil von Forschergruppen, die auf die Ausbildung des wissenschaftlichen Nachwuchses spezialisiert sind.

Die obigen Überlegungen zu Performanzindikatoren können zweifellos unter dem Aspekt kritisiert werden, dass sie die Tätigkeit einzelner Forschergruppen zu stark erfassen und damit eine freie wissenschaftliche Forschung behindern. Allerdings gibt es schon jetzt eine Vielzahl von indikatorgestützten Erhebungen und Evaluationen, und es geht nicht um eine Ausweitung der aktuellen Analysen. Ziel dieser Betrachtung ist es vielmehr, die Qualität der quantitativen Analysen zu verbessern und Fehlbewertungen sowie Fehlanreize zu vermeiden. Darüber hinaus geht es darum, die Zahl der qualitativen Evaluationen deutlich zu reduzieren und damit auch deren Qualität zu verbessern.

Literatur

Achleitner, Ann-Kristin; Allmendinger, Jutta; Grupp, Hariolf; Harhoff, Dietmar; Luther, Joachim; Soete, Luc; Schmoch, Ulrich; Stenke, Gero (2008): Gutachten zu Forschung, Innovation und Technologischer Leistungsfähigkeit. EFI-Gutachten 2008, Expertenkommission Forschung und Innovation (EFI). Berlin

Bornmann, Lutz; Daniel, Hans-Dieter (2003): Begutachtungen durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. In: Schwarz, Stefanie; Teichler, Ulrich (Hrsg.): Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung. Frankfurt/New York, S. 207–225

Brennan, John (2008): Experience from the United Kingdom: Performance measures unlimited? In: Kehm, Barbara; Mayer, Evelies; Teichler, Ulrich (Hrsg.): Hochschulen in neuer Verantwortung. Strategisch, überlastet, divers? Bonn, S. 268–269

Butler, Linda (2004): What Happens when Funding is Linked to Publication Counts? In: Glänzel, Wolfgang; Moed, Henk; Schmoch, Ulrich (Hrsg.): Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies on R&D Systems. Dordrecht/Norwell/New York/London, S. 389–406

Grupp, Hariolf; Moguee, Mary Ellen (2004): Indicators for National Science and Technology Policy, in: Glänzel, Wolfgang; Moed, Henk; Schmoch, Ulrich (Hrsg.): Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies on R&D Systems. Dordrecht/Norwell/New York/London, S. 75–94

Hicks, Diana (2004): The Four Literatures of Social Sciences. In: Glänzel, Wolfgang; Moed, Henk; Schmoch, Ulrich (Hrsg.): Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies on R&D Systems. Dordrecht/Norwell/New York/London, S. 473–496

Jansen, Dorothea; Wald, Andreas; Franke, Karola; Schmoch, Ulrich; Schubert, Torben (2007): Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 59, 1, S. 125–149

Kamenz, Uwe; Wehrle, Martin (2007): Professor Untat: Was faul ist hinter den Kulissen. Berlin

Kuhlmann, Stefan; Heinze, Thomas (2004a): Evaluation von Forschungsleistungen in Deutschland. Erzeuger und Bedarf. Teil I: Konzeptionelle Grundlage. Wissenschaftsrecht 37, S. 53–69

Kuhlmann, Stefan; Heinze, Thomas (2004b): Evaluation von Forschungsleistungen in Deutschland. Erzeuger und Bedarf. Teil II: Produktion und Verwendung evaluativer Information sowie Möglichkeiten ihrer künftigen Organisation. Wissenschaftsrecht 37, S. 125–149

Leszczensky, Michael; Orr, Dominic (2004): Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung – Dokumentation und Analyse der Verfahren in 11 Bundesländern. HIS-Kurzinformation A2/2004. Hannover

Moed, Henk (2005): Citation Analysis in Research Evaluation. Dordrecht

Perkmann, Marcus; Walsh, Kathryn (2007): Engaging the Scholar: Three Types of Academic Consulting and their Impact on Universities and Industry. AIM Research Working Paper Series. London

Schimank, Uwe (2005): "New Public Management" and the Academic Profession: Reflections on the German Situation. In: Minerva 43, S. 361–376

Schmoch, Ulrich (2004): Leistungsfähigkeit und Strukturen der Wissenschaft im internationalen Vergleich. In: BMBF (Hrsg.): Studien zum deutschen Innovationssystem, Nr. 6-2005. Bonn/Berlin. <http://www.technologische-leistungsfahigkeit.de> (Zugriff am 20. Oktober 2008)

Schmoch, Ulrich (2008): The Role of Universities in Economic Growth – The German Situation. UniDev Discussion Paper Series Paper no. 1. Lund: Lund University, Research Policy Institute. <http://developinguniversities.blogspot.com/category/discussion-papers-country-reviews/> (Zugriff am 20. Oktober 2008)

Schmoch, Ulrich; Schubert, Torben (2008): Appropriateness of Bibliometrics for Assessing Scientific Performance. Vortrag auf der "10th international Conference on Science and Technology Indicators" vom 17. bis 20. September 2008 in Wien

Schubert, Torben; Schmoch, Ulrich (2009a): Finanzierung der Hochschulen, insbesondere in der Forschung. Von der Input- zur Outputsteuerung. In: Simon, Dagmar; Hornbostel, Stefan; Knie, Andreas (Hrsg.): Handbuch „Wissenschaftspolitik.“ Wiesbaden

Schubert, Torben; Schmoch, Ulrich (2009b): Sustainability of incentives for excellent research – The German case. In: Scientometrics, im Erscheinen

Schneider, Friedrich; Ursprung, Heinrich W. (2008): The 2008 GEA Journal-Ranking for the Economics Profession. In: German Economic Review 9, 4, S. 532–540

Schubert, Torben (2008): New Public Management an deutschen Hochschulen. Strukturen, Verbreitung und Effekte. Stuttgart

Weingart, Peter (2003): Wissenschaftssoziologie. Bielefeld

Anschrift des Verfassers:

Dr. Ulrich Schmoch
Fraunhofer-Institut für System- und Innovationsforschung
Breslauer Straße 48
76139 Karlsruhe
E-Mail: ulrich.schmoch@isi.fraunhofer.de

Ulrich Schmoch arbeitet als Abteilungsleiter und aktuell als Leiter der Taskforce „Innovationsindikatoren“ am Fraunhofer-Institut für System- und Innovationsforschung (ISI) in Karlsruhe. Er ist Privatdozent für Wissenschafts- und Techniksoziologie an der Universität Karlsruhe. Seine Arbeitsgebiete sind Innovationsindikatoren, Innovationssysteme, Wissenstransfer und Wissenschaftsforschung.