

## Leistungsmessung in der Forschung

Hornbostel, Stefan

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

### Empfohlene Zitierung / Suggested Citation:

Hornbostel, S. (2006). Leistungsmessung in der Forschung. In *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung* (S. 219-228). Bonn: Hochschulrektorenkonferenz (HRK).  
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-25386>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

aus Diskussionen der Peers gehen dann gemeinsame Beurteilungen mit differenzierten Empfehlungen, was zu tun ist, hervor.

Ein derartiges Verfahren ist freilich aufwendig; damit der Aufwand noch erträglich bleibt, muss er durch gute Organisation und Routinen in Grenzen gehalten werden.

6. Bei alledem ist schließlich auch eine Totalisierung von Leistungsdruck zu vermeiden. Wissenschaftler müssen eine erhebliche Fehlerrate ohne Schaden überstehen können.

Insbesondere die Forschung bedeutet immer wieder, hohe Risiken des Scheiterns eingehen zu müssen; Ähnliches kann aber auch für innovative Formen der Lehre gelten. Wenn man nicht will, dass nur noch Mainstream-Forschung und -Lehre betrieben wird, darf man auch wiederholtes Scheitern von Vorhaben nicht deshalb bestrafen, weil bestimmte Leistungsindikatoren nicht erfüllt werden.

Hinzu kommt: Wird der Konkurrenzdruck zu stark, werden übermäßig viele Energien für Akquisitionsstrategien (Anträge, Berichte, Netzwerkpfege, impression management u.ä.) statt für das Kerngeschäft von Forschung und Lehre eingesetzt. Dies schlägt ab einem bestimmten Punkt in Ineffizienz um.

# Forum 1: Leistungsmessung in der Forschung

## Leistungsmessung in der Forschung

**Stefan Hornbostel**

Lange Zeit stand die Lehre im Zentrum von öffentlichen Leistungsvergleichen, internen und externen Evaluationsverfahren und Zielvereinbarungen. Diese Fokussierung auf die Lehre war durchaus begründet, denn – anders als die Forschung – ist die Qualität der Lehre durch ihre lokale Verankerung besonders wenig transparent und entsprechend anfällig für idiosynkratische Entwicklungen. Allerdings ist die Lehre nur ein Teil des universitären Leistungsspektrums. Die Außenwahrnehmung und die Attraktivität für Wissenschaftler hängen weitaus mehr von den Forschungsleistungen einer Hochschule als von den Leistungen in der Lehre ab. In den letzten Jahren entwickelte sich daher die Forschungsevaluation zum wesentlichen Bestandteil eines profitorientierten Qualitätsmanagements in der Hochschule. Die Instrumentarien für eine solche Evaluation sind zwar deutlich besser entwickelt als im Bereich der Lehrevaluation, weil Forschung über einen internationalen Kommunikationszusammenhang von der scientific community regelmäßig beobachtet wird, aber dennoch weit entfernt von einem routinemäßig einsetzbaren Verfahren. Forschungsbewertungen greifen grundsätzlich auf Peer-Urteile zurück.

Im Falle von Forschungsindikatoren werden Peer-Urteile indirekt in Gestalt von begutachteten und bewilligten Drittmittelanträgen, veröffentlichten Zeitschriftenartikeln nach Manuskriptbegutachtung, Zitaten etc. berücksichtigt. Das ermöglicht eine von subjektiven Wahrnehmungen unabhängige, allerdings stark verfahrensabhängige Information über Umfang und Resonanz von Forschung. Explizite Peer-Urteile fallen an, wenn Experten unmittelbar den Forschungsoutput einer Institution begutachten und bewerten. Nachteil solcher Verfahren ist neben dem hohen Aufwand auch die Anfälligkeit für subjektive Fehleinschätzungen. In der Praxis werden häufig Mischformen beider Verfahren verwandt (infor-

med Peer-Review), bei denen Experten ihr Urteil ganz oder teilweise auf Indikatoren stützen (vgl. Abb. 1).

Typisch für ausschließlich indikatorenbasierte, vergleichende Berichte über Forschungsleistungen sind die Forschungsrankings, die international, auf europäischer Ebene oder national publiziert werden (z.B. das CHE-Ranking im „Stern“ bzw. in der „Zeit“, das Focus-Ranking, das DFG-Förderranking). Eine expertenbasierte Evaluation wird regelmäßig in Niedersachsen durch die wissenschaftliche Kommission durchgeführt und hat weniger den Vergleich als vielmehr die Bedingungen „vor Ort“ im Kontext der Hochschulplanung des Landes zum Gegenstand. Auch freiwillige Benchmarkings werden gelegentlich für den Vergleich von Forschungsleistungen eingesetzt. Das jüngst vom Wissenschaftsrat empfohlene „Forschungsrating“ basiert – in Anlehnung an das britische Research Assessment Exercise – auf einer Benotung von Forschungsleistungen durch Experten, die ihr Urteil nicht auf „vor Ort-Begehungen“ stützen, sondern auf Indikatoren und von den Hochschulen übermittelten Daten (vgl. Abb. 2).

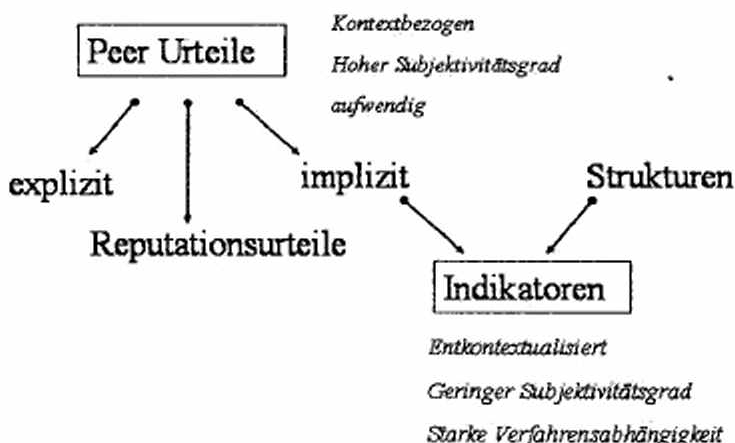


Abb. 1

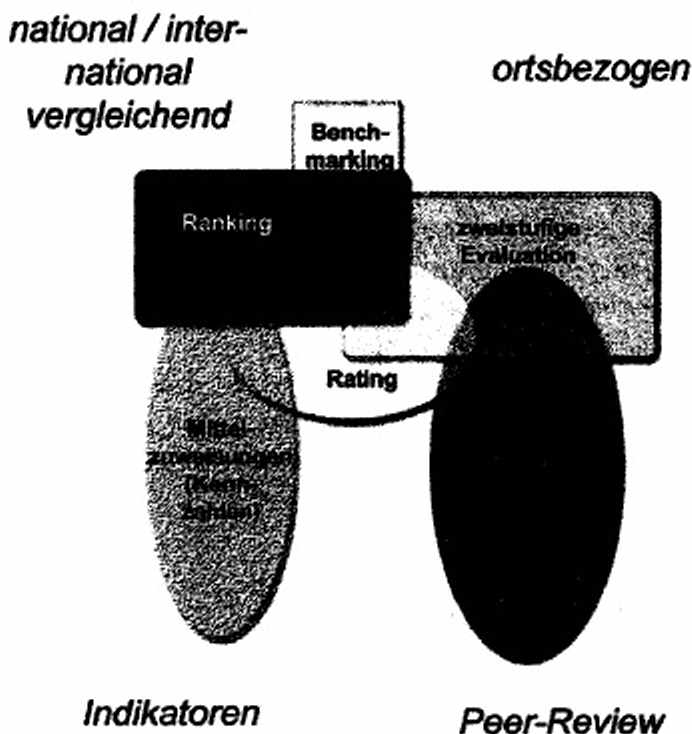


Abb. 2

Auf der Länderebene und hochschulintern werden forschungsbezogene Kennzahlen immer häufiger für Mittelverteilungen eingesetzt und schließlich darf man erwarten, dass – sofern die studiengangsbezogene Akkreditierung fortgesetzt wird – zumindest bei der Akkreditierung von Master-Studiengängen eine Berücksichtigung von Forschungsleistungen notwendig sein wird.

Forschungsindikatoren – insbesondere bibliometrische Indikatoren – sind anspruchsvolle Instrumente, die mindestens in dreierlei Hinsicht einer kompetenten Interpretation bedürfen: Zum einen können Indikatoren fehlerbehaftet sein. Im Unterschied zu Expertenurteilen sind derartige Fehler aber identifizierbar und korrigierbar. Zum anderen kann die Konstruktion eines Indikators den Evaluationszielen mehr oder weniger an-

gemessen sein. Und drittens sind Indikatoren je nach Disziplin unterschiedlich gehaltvoll bzw. überhaupt erhebbar.

Die folgende Tabelle (vgl. Tab. 1) zeigt das Tableau der vom CHE benutzten Indikatoren für das Fach Chemie. Neben Absolutangaben werden dabei jeweils per Capita-Angaben veröffentlicht (Basis sind jeweils die Professoren am Fachbereich/Institut oder alle Wissenschaftler mit Ausnahme des Drittmittelpersonals). „Reputation“ beruht auf den Empfehlungen der Professoren (aus einer Vollerhebung im Fach), Drittmittelanangaben werden von den Hochschulen geliefert, die Angaben zu Publikationen und Zitationen stammen aus dem „Web of Science“, wobei nur die Publikationen der Professoren (deren Namen von den Dekanaten geprüft werden) berücksichtigt werden. Das hat den Vorteil, dass eine Pro-Kopf-Berechnung tatsächlich möglich ist und den Nachteil, dass der akademische Mittelbau unberücksichtigt bleibt. Schließlich wird auch die Anzahl der Promotionen als Indikator ausgewiesen. Wie die Tabelle zeigt, korrelieren alle Indikatoren positiv miteinander. Dabei ist auch zu erkennen, dass das Reputationsurteil deutlich vom absoluten „Output“ eines Fachbereichs beeinflusst wird. Große Fachbereiche sind einfach „sichtbarer“, auch wenn die Effizienz möglicherweise gering ist. Erkennbar ist schließlich auch, dass die Zahl der Promotionen (die meist im Rahmen von Forschungsprojekten angefertigt werden) mit den übrigen Indikatoren positiv korreliert.

Betrachtet man das Indikatorenset, das für das Fach Soziologie benutzt wurde, werden exemplarisch einige Probleme deutlich. Für die Ermittlung der Publikationen ist der internationale „Social Science Citation Index“ nicht tauglich. Nicht nur der Anteil der deutschen Publikationen, die dort nachgewiesen werden, ist verschwindend gering, auch die Begrenzung auf Zeitschriften ist für sozialwissenschaftliche Disziplinen sehr problematisch. Wie die folgende Graphik zeigt, werden auch in den soziologischen Zeitschriften – nicht nur in Deutschland – in hohem Maße Monographien und Sammelbände zitiert (vgl. Abb. 3). D.h. die fachliche Kommunikation läuft nicht wie in den Naturwissenschaften ausschließlich über Zeitschriften. Man erzeugt also schnell ein falsches Bild, wenn man sich ausschließlich auf Zeitschriftenaufsätze kapriziert.

Chemie: Korrelationen zwischen Indikatoren (CHE)									
	Drittmittel pro Jahr (in TE)	DFG-Mittel pro Jahr (in TE)	Publika- tionen pro Jahr	Zitate pro Jahr	Promotionen pro Jahr	Drittmittel je Wissens- schaffler (in TE)	Publika- tionen pro Professor	Zitate pro Papier	Promo- tionen pro Professor
FORSCHTIPP	,664(**)	,532(**)	,443(**)	,440(**)	,635(**)	,211	,253(*)	,213(*)	,447(**)
Drittmittel pro Jahr (in TE)		,761(**)	,572(**)	,586(**)	,676(**)	,648(**)	,805(**)	,378(**)	,600(**)
DFG-Mittel pro Jahr (in TE)			,594(**)	,655(**)	,569(**)	,474(**)	,699(**)	,453(**)	,356(**)
Publikationen pro Jahr				,926(**)	,716(**)	,341(*)	,823(**)	,411(*)	,529(**)
ZITATE/JAHR					,721(**)	,661(**)	,713(**)	,694(**)	,569(**)
Promotionen pro Jahr						,382(*)	,584(*)	,491(*)	,532(*)
Drittmittel je Wissenschaftler (in TE)							,248(*)	,351(*)	,263(*)
Publikationen pro Professor								,471(*)	,453(*)
ZITATE/papier									,385(*)

\* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.  
 \*\* Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2002

Tabelle 1

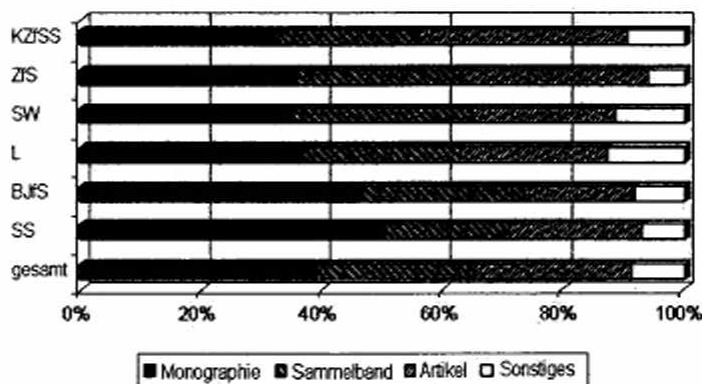


Abb. 3

Zitierte Literatur aus sechs deutschen soziologischen  
 Fachzeitschriften des Jahrganges 2002 nach Art der zitierten Quelle

Die in Tabelle 2 benutzten Daten stammen daher aus der SOLIS Datenbank des IZ Bonn. Wesentlicher Nachteil ist, dass diese Daten keinerlei Information über die Relevanz der Dokumente enthalten (wie sie oben für die Chemie mit Hilfe von Zitationen erfasst wurde). Besonders auffällig ist hier, dass offenbar die Anzahl der Promotionen nicht als echter Forschungsindikator aufgefasst werden kann. In der Soziologie sind Promotionen nur in geringem Maße an Forschungsprojekte des jeweiligen Instituts geknüpft, was sich in den leicht negativen Korrelationskoeffizienten auch ausdrückt. Nachwuchsförderung sollte daher als eigenständige Dimension und nicht als Forschungsleistung interpretiert werden.

Soziologie: Korrelationen zwischen Indikatoren (CHE)							
	Drittmittel insgesamt pro Jahr	DFG- Mittel pro Jahr	Publikationen pro Jahr	Promotionen pro Jahr	Drittmittel je Wissenschaftler	Publikationen je Professor	Promotionen je Professor
Forschungsreputation	,597(**)	,676(**)	,501(**)	,457(**)	,319*	,357(**)	,102
Drittmittel insgesamt pro Jahr		,746(**)	,664(**)	,561(**)	,584(**)	,237	,039
DFG-Mittel pro Jahr			,531(**)	,586(**)	,327(*)	,131	,171
Publikationen pro Jahr				,635(**)	,603	,615(**)	,028
Promotionen pro Jahr					,158	,015	,610(**)
Drittmittel je Wissenschaftler						,146	,181
Publikationen je Professor							,146

\* Die Korrelation ist auf dem Niveau von 0,05 (1-seitig) signifikant.  
 \*\* Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Quelle: CHE Fachbereichserhebung 2001

Tabelle 2

Die Art und Weise der Ermittlung von Publikationen und Zitationen (ganz abgesehen von möglichen Gewichtungen) hat selbst dann erheblichen Einfluss, wenn die gleiche Datengrundlage benutzt wird, wie ein Vergleich von Focus und CHE-Ranking zeigt (beide nutzen das WEB of Science). Im „Focus“ wurden die Publikationen auf Grundlage der institutionellen Adressen in der Datenbank und einer fachlichen Klassifikation, die auf der Zuordnung von Zeitschriften zu Fachgebieten beruht, ermittelt, beim CHE-Ranking erfolgte eine namentliche Überprüfung der Professoren. Abgesehen von Fehlern und höchst unterschiedlichen Schreib-



weisen der institutionellen Adressen wird z.B. ein Psychologieprofessor, der in einer medizinischen Zeitschrift publiziert, im einen Fall als „Mediziner“ behandelt, im anderen als „Psychologe“. Derartige Unstimmigkeiten sind keine Fehler, sondern unterschiedliche Abgrenzungen des Gegenstandsbereiches. Man kann Forschung als einen disziplinären Zusammenhang verstehen oder als einen organisatorischen. Im einen Fall entscheidet die Zugehörigkeit zu einem Institut, im anderen Fall die Nutzung bestimmter Publikationsorgane über die fachliche Zuordnung.

Die Ziele einer Forschungsevaluation sollten daher die Auswahl und Konstruktion von Indikatoren bestimmen – nicht umgekehrt. Welche Probleme entstehen, wenn man die Definition von Forschungsaktivität und -qualität einfach an die Datenbankbetreiber delegiert, wird immer dann sichtbar, wenn Disziplinen einen großen nationalen Publikationsmarkt bedienen. Konzentriert man sich nur auf „internationale Sichtbarkeit“ entsteht möglicherweise ein verzerrtes Bild: Betrachtet man z.B. den Publikationsoutput deutscher VWL-Professoren, fällt auf, dass keineswegs der Zeitschriftenartikel absolute Dominanz hat: ca. 36 Prozent der VWL-Publikationen (von Professoren) sind Zeitschriftenartikel, weitere 32 Prozent sind Beiträge in Sammelbänden und rund 22 Prozent lassen sich als „graue Literatur“ klassifizieren. Diese Publikationen sind sicher nicht alle von gleicher Qualität, allerdings kann auch nicht unterstellt werden, dass alle vom Social Science Citation Index nicht erfassten Publikationen als „wertlos“ für die Forschung eingestuft werden können. In wirtschafts-, sozial- und geisteswissenschaftlichen Forschungsfeldern orientieren sich offenbar viele Autoren weniger an internationalen Spitzenjournalen, sondern am avisierten Auditorium, das häufig gerade nicht über diese Journals erreicht wird. Man mag das beklagen (die Fülle internationaler Rankings demonstriert die marginale deutsche Position in der wirtschaftswissenschaftlichen Forschung), aber man wird kaum umhinkommen, danach zu fragen, ob diese Form der Forschungskommunikation wirklich bedeutungslos ist. Eine Vorstellung vom „Verlust“ erhält man, wenn man einem Vorschlag von Kalaitzidakis (1999) folgt, der sein wirtschaftswissenschaftliches Ranking auf lediglich zehn Spitzenjournals stützt und zusätzlich eine Qualitätsgewichtung benutzt, die sowohl eine Standardisierung des Umfangs der Zeitschriftenartikel wie auch eine auf dem Zeitschriftenimpact beruhende Qualitätskomponente beinhaltet, wobei

das American Economic Review (AER) die Referenzgröße bildet. Wendet man das Verfahren der „AER Standardized pages“ auf das CHE Datenmaterial an, dann bleiben von den 4149 volkswirtschaftlichen Publikationen 32 übrig, also weniger als ein Prozent der deutschen Publikationen. Bereits die absoluten Fallzahlen zeigen, dass ein solides Ranking auf dieser Basis kaum möglich ist.

Universität	AER Adj. Pages	Publikationen
Uni Frankfurt a.M.	35,39	4
Uni Bonn	19,67	2
Uni Mannheim	18,72	3
HU Berlin	18,63	4
TU Dresden	11,75	1
Uni München	11,66	5
Uni Oldenburg	8,69	1
Uni Osnabrück	8,23	2
Uni Freiburg	5,50	2
Uni Essen	2,50	1
Uni Duisburg	1,98	1
Uni Dortmund	1,32	1
Uni Hamburg	1,01	2
Uni Heidelberg	,60	2
Uni Tübingen	,31	1

**Tabelle 3**

CHE Sample 1998 - 2000 VWL (nur Professoren)

Weniger gravierend, aber dennoch erkennbar ist dieses Problem auch in europäischen Rankings (vgl. dazu den Beitrag von Peter in diesem Band) wie dem jüngst erschienenen „Mapping of Excellence“. Auch dort macht sich ein angloamerikanischer bias massiv bemerkbar, erkennbar an den für einen Zehnjahreszeitraum sehr niedrigen Publikationszahlen (vgl. Tab.4, Spalte Papers). Auskunft geben diese Rankings aber über die internationale Resonanz, ausgedrückt als „Zitationserfolg“ im Vergleich zum Wissenschaftsgebiet bzw. zu den übrigen Artikeln der genutzten Zeitschriften (vgl. die letzten Spalten der Tab. 4).

Die wenigen Beispiele sollten deutlich machen, dass Instrumente zur Leistungsmessung in der Forschung Diagnoseinstrumente sind, die sorgfältig auf die Evaluationsziele eingestellt werden müssen. Die Therapie zur gezielten Verbesserung der Effektivität (Zielerreichung) ist damit keineswegs festgelegt. Sicherlich sind die Hochschulen aber gut beraten sich auf einen steigenden Informationsbedarf hinsichtlich der Forschungsleistungen einzurichten und eigenständig eine Informationsbasis zu schaffen, die nicht nur für internes Qualitätsmanagement geeignet ist, sondern auch die kompetente Interpretation (und ggf. kritische Revision) der Forschungsindikatoren und Kennzahlen, mit denen Hochschulen öffentlich gemessen werden.

Germany's performance in sub-fields of economics, according to SSCI 1991-2001						
<i>Macroeconomics and quantitative methods</i>						
U Mannheim	25	108	4.2	54%	1.43	1.51
Humboldt U Berlin	52	200	3.8	42%	1.43	1.51
U Bonn	36	100	2.8	47%	1.79	1.16
U Dortmund	20	44	2.2	50%	0.71	0.78
<i>Labour and development economics</i>						
U Munich	20	85	2.4	50%	1.33	1.46
Humboldt U Berlin	20	90	4.5	30%	1.24	1.29
Max Planck Institutes	20	42	2.8	50%	1.10	1.23
U Bonn	31	71	2.3	52%	0.85	1.16
U Mannheim	31	71	1.4	62%	0.94	0.96
<i>Health economics</i>						
Wissenschaftszentrum Berlin	22	67	3.0	55%	1.41	1.33
U Munich	33	106	3.2	36%	1.26	1.22
U Mannheim	25	36	1.4	49%	0.74	0.99
U Bonn	28	56	2.0	36%	0.94	0.82
Free U Berlin	21	35	1.7	43%	0.59	0.50
<i>Economic development, industrial organization and economic geography</i>						
U Kiel	21	20	1.2	43%	0.65	0.31

Source: DG-Research

Tabelle 4

### Literaturverzeichnis

- Berghoff, S.; Federkeil, G.; Giebisch, P.; Hachmeister, D.; Müller-Böling, D. (2003): Das CHE-Forschungsranking deutscher Universitäten. Arbeitspapier Nr. 50. Gütersloh.
- Berghoff, S./Hornbostel, S. (2003): Das CHE hinter den sieben Bergen. In: Perspektiven der Wirtschaftspolitik. Mai, Volume 4, Issue 2, S.177-294.
- DFG (2004): Förder-Ranking 2003, Bonn.
- European Commission (2004): Mapping of Excellence, Luxembourg, EUR 20889.
- Hornbostel, S. (2004): Leistungsparameter und Ratings in der Forschung. In: Evaluation – ein Bestandteil des Qualitätsmanagements. Beiträge zur Hochschulpolitik 9/2004, S.173-187 (hrsg. von der HRK).
- Hornbostel, S. (2004): Der Drahtseilakt – Evaluationskriterien zwischen fachinterner Relevanz und externer Definition. In: Merkens, Hans. (Hrsg.): Evaluation in der Erziehungswissenschaft, Wiesbaden: VS Verlag, S. 77-88.
- Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (1999), „European Economics: An Analysis Based on Publications in the Core Journals“, European Economic Review, vol. 43, p. 1150-1168.
- Wissenschaftsrat (2004): Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung, Hamburg, Drs. 6285-04.

# Qualitätsbewertung von Forschung in Niedersachsen

## Christof Schiene

Zur Leistungsmessung im Hochschulbereich werden inzwischen eine Reihe unterschiedlicher Instrumente benutzt. Die Variationsbreite hängt dabei sehr stark vom Auftraggeber und der Zielsetzung der Bewertung, sodann von der durchführenden Einrichtung und einer Vielzahl lokaler bzw. regionaler Randbedingungen ab, die sich auf der Ebene der Bundesländer ebenso feststellen lassen wie auf der Ebene der einzelnen Hochschule. Insbesondere bei der Bewertung von Lehre und Forschung haben sich inzwischen Verfahren etabliert, die als „evidence based“- oder „informed peer-review“-Verfahren bezeichnet werden. Sie kombinieren Elemente des (innerwissenschaftlichen) „peer-review“ mit quantitativen Leistungsindikatoren. Da weder Qualitätsbegriffe noch wissenschaftliche Leistungen in Lehre und Forschung als wertfrei begriffen und absolut gesetzt werden können, sind auch die Instrumente und Maßnahmen zur Leistungsmessung nicht als gut oder schlecht, sondern – ähnlich wie in der wissenschaftlichen Theoriebildung – lediglich als mehr oder minder angemessen zu betrachten. So muss sich letztlich erweisen, wie angemessen solche Verfahren dem Gegenstandsbereich sind, den sie bewerten sollen, wie sich das Verhältnis von Aufwand und Ertrag gestaltet und inwieweit die Ergebnisse nachhaltig in die wissenschaftliche und wissenschaftspolitische Entscheidungsfindung eingehen.

### Kontext der Forschungsevaluation

In Niedersachsen werden seit 1999 flächendeckend Evaluationsverfahren in der Forschung an Universitäten und gleichgestellten Hochschulen sowie punktuell an außeruniversitären Forschungseinrichtungen durchgeführt. Die Forschungsevaluation in Niedersachsen ist zu sehen im Kontext institutionalisierter Politikberatung, die von der Wissenschaftlichen Kommission Niedersachsen (WKN) erbracht wird.