

Wissenschaftsindikatoren: Mittel zur Selbstbeobachtung oder Schiedsrichter im Verteilungskampf?

Hornbostel, Stefan

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Hornbostel, S. (2001). Wissenschaftsindikatoren: Mittel zur Selbstbeobachtung oder Schiedsrichter im Verteilungskampf? In E. Keiner (Hrsg.), *Evaluation (in) der Erziehungswissenschaft* (S. 23-42). Weinheim: Beltz.
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-23744>

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/4.0/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/4.0>

Wissenschaftsindikatoren: Mittel zur Selbstbeobachtung oder Schiedsrichter im Verteilungskampf?

Gegenwärtig kommt man kaum umhin, sich mit Fragen der Leistungsmessung, des Vergleichs von Institutionen, der Verteilung knapper Ressourcen anhand von Kennzahlen oder subjektiver Bewertungen zu befassen. Derartiges ist nicht nur in der öffentlichen, sondern auch in der hochschulpolitischen Diskussion allgegenwärtig. Und zunehmend sind auch in der universitären Lebenswelt vieler Wissenschaftler die ganz konkreten Spuren dieser Debatte zu finden: Indikatorengestützte Mittelzuweisungen, Lehrevaluationen, Fachbereichsevaluationen oder öffentliche Hochschulrankings, bench-marking-Vereinbarungen, leistungsbezogene Besoldungen usw.

Vordergründig scheint allein das Diktat leerer öffentlicher Kassen oder das Aufkommen neoliberaler Steuerungsmodelle für diese Erscheinungen verantwortlich zu sein. Entsprechend abwehrend reagieren die einen, die vor allen Dingen die Autonomie der Wissenschaft im allgemeinen und der Hochschulen im besonderen in Gefahr sehen. Andererseits bestehen kaum Zweifel daran, dass das deutsche Hochschulwesen nicht nur in Lehre und Studium, sondern auch in manchen Teilen der Forschung erhebliche Ineffizienzen und Verkrustungen aufweist, in seiner Reaktionsfähigkeit begrenzt ist, in seinem Management und seiner organisatorischen Gestaltung den Anforderungen kaum gerecht wird. Die Lage der deutschen Hochschulen wird gegenwärtig fast unisono als lamentabel dargestellt. Dementsprechend euphorisch werden neue Steuerungsmodelle, inhaltliche und rechtliche Reformen von den anderen begrüßt (vgl. KIESER 1999; HORNBOSTEL 1999a).

Weder die Debatte noch die diskutierten Lösungsansätze sind sonderlich neu: Seit Mitte der 80er Jahre hat der Wissenschaftsrat immer wieder auf die Dilemmata des deutschen Hochschulwesens aufmerksam gemacht und als einen Kernpunkt der Umgestaltung des Hochschulwesens die stärkere Einführung wettbewerblicher Elemente in das Wissenschaftssystem gefordert. Dazu sei eine Stärkung der Selbststeuerungsfähigkeit des Wissenschaftssystems bei gleichzeitigem Rückzug des Staates aus vielen Regelungsbereichen notwendig (vgl. WISSENSCHAFTSRAT 1985, 1996, 2000; SCHIMANK 1999).

Was wir gegenwärtig erleben ist ein erster - noch zaghafter - Versuch, die Steuerungsprobleme in Lehre und Forschung mit neuen Mitteln anzugehen. Standen vor dem überwiegend quantitatives Wachstum und eine in hohem Maße politisch initiierte Schwerpunktpolitik im Vordergrund, so wird seit mehr als 10 Jahren zunehmend die Stärkung der Selbststeuerungsfähigkeit des Wissenschaftssystems hervorgehoben. Allerdings mit einer erheblichen Verschiebung dessen, was bisher unter Autonomie der Wissenschaft verstanden wurde. In der Nachkriegsgeschichte galt die rechtlich garantierte Autonomie der Wissenschaft in erster Linie als Abwehrrecht

gegen staatliche Steuerungsversuche, als kulturstaatliche Verpflichtung zur Alimenterung der Wissenschaft und als positive Begründung für eine thematisch und strukturell sich frei entwickelnde Wissenschaft.

Womit wir heute in der hochschulpolitischen Debatte konfrontiert sind, ist ein Autonomieverständnis, das man als Gestaltungsauftrag formulieren könnte. Es ist mit Begriffen wie Rechenschaftspflicht, Verantwortlichkeit für das Klientel, Verpflichtung zur strukturellen Reform, zur effizienten Produktion von Wissen, zur Auseinandersetzung mit gesellschaftlichen Leistungserwartungen und Ausweis eines funktionierenden internen Steuerungs- und Allokationsmechanismus geprägt.

In Bewegung geraten ist sowohl die Diskussion um die richtige Balance von thematisch ungebundener Forschungsförderung und forschungspolitischer Schwerpunktbildung in und zwischen den Disziplinen (vgl. INTERNATIONALE KOMMISSION zur Systemevaluation der DFG und der Max-Planck-Gesellschaft 1999), als auch die Debatte um den Leistungsstand der einzelnen Fachbereiche und Hochschulen in Lehre und Forschung, intern (Lehrveranstaltungsevaluation, Fachbereichsevaluation; vgl. z.B. KREMPKOW 1999) wie auch im Vergleich mit anderen Institutionen (RÖBBECKE & SIMON 1999, GEGENWORTE 2000).

Die Deutsche Gesellschaft für Erziehungswissenschaft hat sich dieser Debatte nicht verschlossen: Die Grundzüge des „Evaluationskataloges für die Erziehungswissenschaft“ sehen eine vergleichende, datengestützte Dauerbeobachtung von Lehre, Forschung und disziplinärer Entwicklung vor, die nicht nur strukturelle Beschreibungen und Leistungsvergleiche von Institutionen vorsieht, sondern auch zwischen Fächern und Ländern. Zudem sollen disziplinäre Entwicklungen beobachtet und ggf. gefördert und Profilbildungen vor Ort angestoßen werden (vgl. VORSTAND DER DGfE 1998; OTTO u.a. 2000). Das ist ein anspruchsvolles Programm. Es beinhaltet die Erwartung, dass ein empirisch gesättigtes Wissen über das Fach und seine Institutionen generiert werden kann und dass dieses Wissen in einem Selbststeuerungssystem Wirkungen entfaltet. Zugleich – eher im Subtext – enthält dieser Evaluationskatalog die Befürchtung, dass die Verweigerung eines indikatoren gestützten Leistungsnachweises erhebliche Nachteile in der Ressourcenzuteilung mit sich bringen kann, sowohl für die Disziplin insgesamt wie auch für einzelne Institutionen.

Obwohl Wissenschaft und Forschung auch heute mit Begriffen wie Unwägbarkeit, Zufall, Nicht-Planbarkeit, Nicht-Messbarkeit etc. assoziiert sind, wirkt ein solches Programm heute weniger provokant, denn inzwischen haben wir uns daran gewöhnt, dass fast alles, wenn nicht exakt vermessen, so doch öffentlicher Evaluation zugeführt wird. Dabei spielen Indikatoren, die ursprünglich meist in einem analytischen Kontext entwickelt wurden, eine immer wichtigere Rolle. Zum einen stoßen die klassischen Verfahren des Peer-Review angesichts von Verfahrensmängeln, aber auch weiter steigenden Bedarfs an Evaluation, monitoring, öffentlicher Rechenschaftslegung und öffentlichen Leistungsvergleichs an ihre Grenzen (vgl. FRÖHLICH 1999). Kompetente Peers stehen nur in begrenzter Zahl zur Verfügung und sind auch nur begrenzt belastbar. Zudem sind derartige Begutachtungen sehr aufwändig und für

Routineberichterstattungen kaum geeignet. Schließlich geht es zunehmend um den Vergleich von Institutionen, um „best practice“-Modelle und nicht so sehr um die Erfüllung von Mindeststandards. Dazu aber sind meist Vergleiche notwendig, die eine Quantifikation von Leistungen und eingesetzten Ressourcen erfordern. Last not least spielen Indikatoren eine immer größere Rolle bei der Verteilung von (zunehmend globalisierten) Haushaltsmitteln.

Indikatoren

Indikatoren sollen Repräsentanten für eine meist wenig überschaubare und hochkomplexe Realität sein, und sie sollen diese Realität in wenigen Zahlenwerten repräsentieren. Allerdings ist die zu indizierende Realität – nicht nur im Falle von Wissenschaft – in der Regel nur sehr vage gekennzeichnet: Schwammige Begriffe wie Armut, wirtschaftliche Gesundheit, Fortschritt, gute Forschung etc. stehen bei Wirtschafts-, Sozial-, aber auch Wissenschaftsindikatoren im Hintergrund. Voraussetzung für eine Repräsentation von Realitätsausschnitten ist aber, dass diese Begriffe soweit operationalisiert und modelliert werden können, dass sich ein Indikator konstruieren lässt. Das unterscheidet Indikatoren von Kennzahlen, hinter Indikatoren stehen modellhafte Annahme über die Realität und ein Operationalisierungskonzept.

In ihrem Entstehungskontext, nämlich im Rahmen akademischer Betrachtungen und Analysen der schnell expandierenden und gesellschaftlich immer wichtiger werdenden Wissenschaft in den 50er und 60er Jahren, wurden Wissenschaftsindikatoren eher als das analytische Instrumentarium einer spezialisierten Wissenschaftsforschung angesehen, jedenfalls waren sie wissenschafts- und hochschulpolitisch eine undramatische Angelegenheit. Werden Indikatoren aber über diesen Rahmen hinaus bedeutsam, dann sind Interessen tangiert, und die Angemessenheit von Indikatoren steht zur Debatte. Dass die Auswahl und Konstruktion von Indikatoren bestimmte Akteure bevorzugt bzw. zurücksetzt, ist kein Defekt der Indikatoren, sondern lediglich ein Verweis darauf, dass Indikatoren nicht per se eine Wirklichkeit abbilden. Indikatoren antworten auf Fragen, und sie tun dies meist mit allerlei Unzulänglichkeiten. Ähnliches gilt für die Frage der Steuerungswirkungen von Indikatoren. Ob Indikatoren gewünschte Effekte erzeugen, akzeptable Kollateralschäden produzieren oder kontraproduktiv wirken, wenn Verteilungsentscheidungen an sie geknüpft werden, lässt sich nur anhand einer definierten Zielvorstellung und Wissen über die Responsivität des zu steuernden Gebietes abschätzen.

Strukturindikatoren: Disziplin und Profession

Folgt man dem britischen Wissenschaftsphilosophen Steven TOULMIN (1983), dann lässt sich unter Disziplin ein kognitives System von Erkenntnisidealen, Fragestellun-

gen, Problemen, Begriffen und Methoden verstehen, das sich historisch entwickelt und verändert. Fortschritt, zumindest Entwicklung, lässt sich disziplinar und ex-post als wachsende Erklärungskraft von Theorien, als Lösung von Rätseln rekonstruieren.

Allerdings ruht dieses „geistige Unternehmen“ auf einer sozialen Basis, die man als akademische Profession bezeichnen kann: Ein Ensemble von Institutionen, Rollen, Organisationsregeln und Personen. Die Funktionsweise dieser Profession beeinflusst mit einem eigenständigen Anteil die Selektionen aus dem Pool von Ideen, Theorien und Erkenntnisansprüchen. Disziplin und Profession sind aufeinander bezogen, sie stehen jedoch in keinem linearen Abhängigkeitsverhältnis.

Auf der Ebene der Disziplin lassen sich in aller Regel über den Weg methodologischer Prüfungen keine Entscheide fällen über die Qualität von Theorien und Erkenntnisansprüchen. Will man sich der kognitiven Seite nähern, so ist dies nur über die soziale Seite, über die Praxen wissenschaftlichen Handelns möglich. Die Kommunikation der Wissenschaftler über Erkenntnisansprüche, ihr Anschließen an vorhandenes Wissen, ihre Bewertungen und Selektionen aus den Produkten der Kollegen bilden den empirisch zugänglichen Teil disziplinärer Entwicklung.

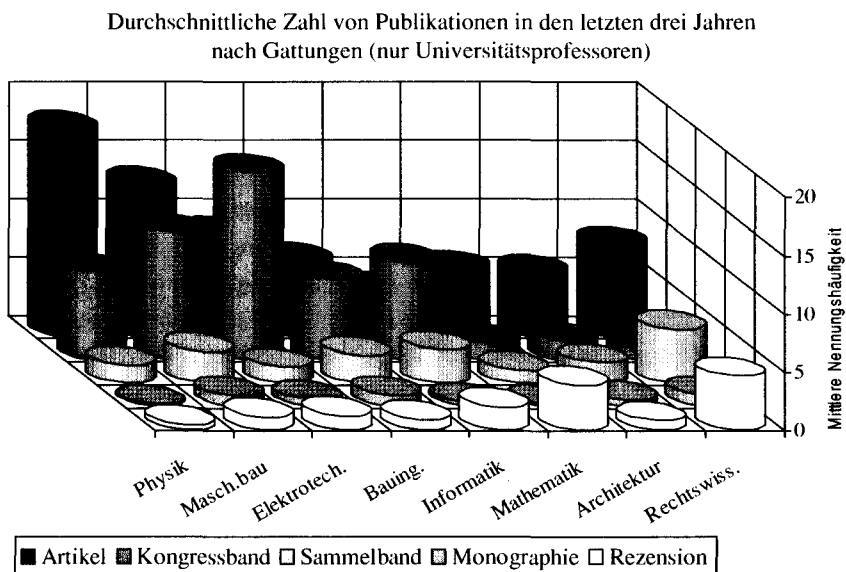
Dabei stand zunächst die Frage im Mittelpunkt, inwieweit die Profession mit ihren Institutionen (Berufungen, Zeitschriften, Förderungen, Tagungen etc.) normativen Ansprüchen wie Universalität, Uneigennützigkeit etc. entspricht. Bis heute sind dies äußerst sensible Fragen, die im Falle von Betrug und Fälschung, der Entdeckung partikularistischer Kriterien bei der Bewilligung von Forschungsprojekten oder der Abweisung von Manuskripten durch Herausgeber zu heftigen Reaktionen führen (vgl. HORNBOSTEL 1997).

In unserem Zusammenhang wichtiger ist aber eine strukturelle Beschreibung des Kommunikationsraumes einer Disziplin. Wer nimmt überhaupt teil an der Forschungskommunikation?, welches sind die Medien, über die kommuniziert wird?, welche thematischen Felder mit welchen gegenseitigen Bezügen gibt es?, gibt es zentrale Methoden, Fragestellungen, Paradigmen, die sich identifizieren lassen?, wie homogen oder inhomogen ist das Theoriegebäude?, mit welcher Geschwindigkeit veralten Forschungsbefunde?, wo sind Forschungsfronten, die Neuland betreten?, wie sehen die Kooperationsstrukturen aus?, wie werden Forschungsvorhaben finanziert und begutachtet? usw. Zur Beantwortung derartiger Fragen steht inzwischen ein beachtliches Instrumentarium bibliometrischer und anderer Analyseverfahren zur Verfügung. Sie lassen sich historisch ausdehnen, als Bestandsaufnahme konzipieren oder auch als Prospektion einsetzen.

Einer der zentralen Befunde quantitativer Analysen ist dabei eine enorm hohe Varianz zwischen den Disziplinen und z.T. auch innerhalb der Disziplinen. Zwei Beispiele mögen dies illustrieren: Fragt man beispielsweise danach, über welche Medien die fachliche Kommunikation abgewickelt wird, zeigen sich erhebliche Differenzen zwischen den verschiedenen Fächern. Wie die Abbildung 1 zeigt, ist in der Physik mit Artikeln in Fachzeitschriften und Kongressbeiträgen der ganz überwiegende Teil der Publikationen erfasst. Zudem ist die Reputation der verschiedenen journals nicht

sonderlich umstritten, so dass z.B. mit dem Science Citation Index oder spezialisierten Fachdatenbanken relativ problemlos „wichtige“ Beiträge identifiziert werden können. In den Ingenieurwissenschaften spielt sich ein ganz erheblicher Teil der wissenschaftlichen Kommunikation in Kongressbänden ab. Ganz anders die Situation bei den Rechtswissenschaften: Dort spielen Festschriften (Sammelbände) und Monographien und als Reflex darauf Rezensionen eine wichtige Rolle. Artikel in Fachzeitschriften streuen über die unterschiedlichsten journals (auch nicht-juristische). Die Unterscheidung zwischen originärer Forschung, Kommentierung und Dokumentation ist schwer zu ziehen. Über die Wertigkeit einzelner Zeitschriften herrscht wenig Einverständnis. Wollte man die juristischen Publikationen zu einem Indikator verarbeiten, wäre es nicht nur notwendig, die unterschiedlichen Publikationstypen auf irgendeine Weise zu verrechnen, sondern auch den diversen Zeitschriften Gewichte zuzuordnen. Wollte man die Erziehungswissenschaft(en) in diesem Spektrum verorten, so wäre sie sicherlich näher bei den Juristen als bei den Physikern anzusiedeln.

Abbildung 1:



Quelle: Hornbostel, CHE Studienführer 1999 & 2000. Erhebungszeitraum: Masch.bau, Elektrotech., Bauing., Architektur 1999, alle anderen 1998

Zu den wichtigsten und gebräuchlichsten Forschungsindikatoren, die nicht auf Publikationen zurückgreifen, gehören die Drittmittelinwerbungen (vgl. DFG 1997, 2000). Sie sind vor allen Dingen deshalb interessant, weil sie - anders als etwa bibliometrische Indikatoren, die nur retrospektiv berichten - sehr zeitnahe oder sogar prospektive In-

formationen liefern. Bei diesem Indikator ist allerdings strittig, ob man ihn als Input- oder Output-Messung interpretieren kann (vgl. GILLET 1989). Für erstere Deutung spricht, dass die Mitteleinwerbungen noch nichts über den Ertrag eines Projektes aussagen. Für letztere spricht, dass erstens der Drittmittelbewilligung i.d.R. ein aufwändiger Begutachtungsprozess zugrunde liegt, zweitens mit dem Drittmittelantrag häufig bereits ein sehr voraussetzungsvolles Zwischenprodukt vorliegt. Ein Indiz dafür, dass die Begutachtung von Drittmittelanträgen durchaus prognostische Validität besitzt, ergibt sich, wenn man einmal den Publikationsoutput daraufhin untersucht, ob die aus drittmittelgeförderten Projekten hervorgegangenen Publikationen eine andere fachliche Resonanz erzeugen als die übrigen Publikationen (vgl. Tabelle 1).

Tab. 1: Publikationen (1983-87) und Zitate in den ersten drei Jahren nach Erscheinen für die physikalischen Fachbereiche der Universitäten TU Berlin, Köln und Konstanz							
mit Drittmittelförderung				Ohne Drittmittelförderung			
Zahl der Artikel *	Mean Zitate **	Nicht zitiert ***	Viel zitiert ****	Zahl der Artikel *	Mean Zitate **	Nicht zitiert ***	Viel zitiert ****
543	7,4	127 (23,4%)	42 (7,7%)	459	4,9	171 (36,5%)	14 (3,0%)

- * : Zahl der in SCISEARCH nachgewiesenen Artikel (nach institutioneller Adresse).
- ** : Mittelwert (mean) der in den ersten 3 Jahren nach Erscheinen erhaltenen Zitate.
- *** : Artikel, die in den ersten 3 Jahren nicht oder einmal zitiert wurden.
- **** : Artikel, die in den ersten 3 Jahren mehr als 20 mal zitiert wurden

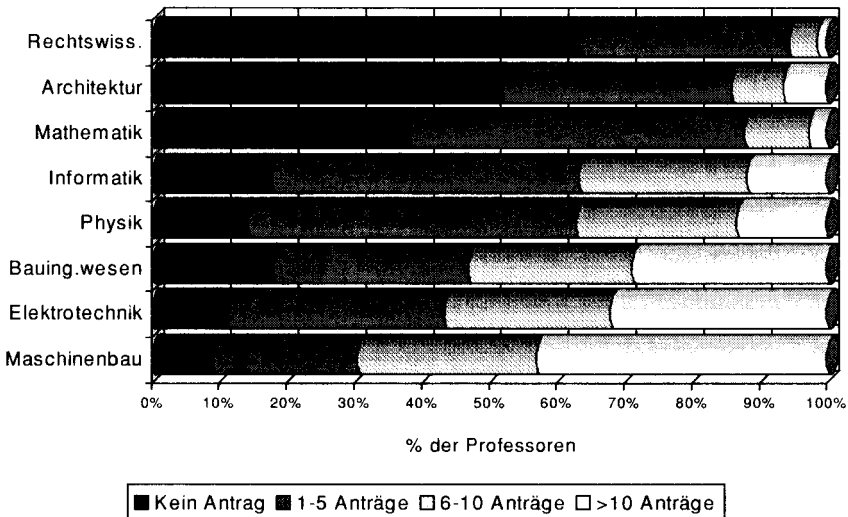
Quelle: HORNBOSTEL 1997

Es spricht also viel dafür, Drittmiteleinwerbungen, insbesondere in Ergänzung zu Publikationsanalysen, als Forschungsindikator zu benutzen. Voraussetzung ist allerdings, dass Drittmittelforschung in den untersuchten Fachgebieten üblich ist und nicht ausschließlich im Rahmen subdisziplinärer Spezialisierung (insbes. empirische Forschung) betrieben wird. Wie der folgenden Abbildung 2 zu entnehmen ist, stellen in einem Zeitraum von 3 Jahren in den naturwissenschaftlichen und technischen Disziplinen deutlich unter 20% der Professoren keine Drittmittelanträge, dafür aber 15% bis 30% der Professoren mehr als 10 Drittmittelanträge. Rechtswissenschaften und Architektur sind hingegen fast drittmittelabstinent; in den Rechtswissenschaften wird Drittmittelforschung denn auch nur in wenigen Spezialgebieten betrieben. Aus diesem Grund macht es z.B. keinen Sinn, die Drittmiteleinwerbungen in der Rechtswissenschaft als Indikator für Forschungsleistungen zu benutzen.

Das wird für die Erziehungswissenschaft nicht in gleichem Maße gelten, aber man wird sich Gedanken darüber machen müssen, wie sich der eher auf geisteswissenschaftliche Reflexion orientierende Teil der Forschung, der eher sozialwissenschaftlich-empirisch ausgerichtete und der eher auf praktisch relevante Dienstleistungen spezialisierte Teil angemessen in Indikatoren repräsentieren lässt.

Abbildung 2:

Drittmittelanträge von Universitätsprofessoren in den letzten 3 Jahren
(Prozent der befragten Professoren)



Quelle: Hornbostel, CHE Studientführer 1999 und 2000

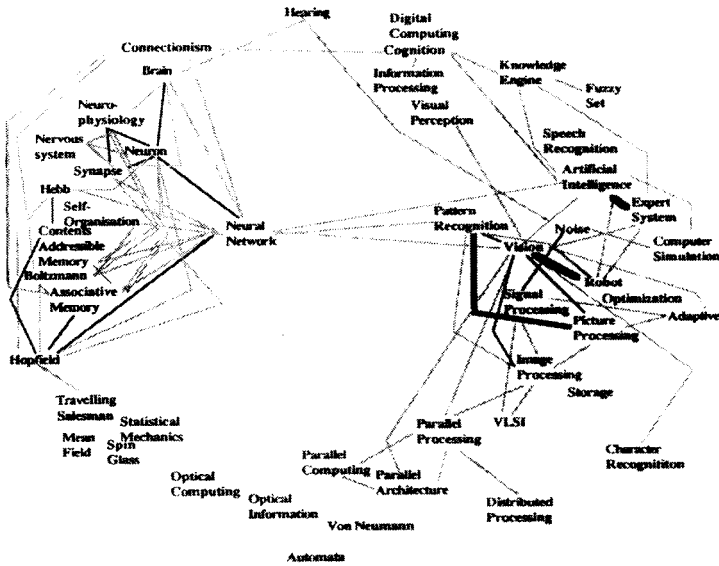
Erhebungszeitraum: Masch.bau, Elektrotechnik, Architektur 1999, alle anderen 1998

Sind solche basalen Fragen geklärt, lassen sich auch komplexere Beschreibungsverfahren einsetzen. Strukturindikatoren zielen dabei weniger auf eine Outputmessung als auf eine Strukturbeschreibung. Die Grundlage solcher Indikatoren sind entweder Publikationsanalysen, die beispielsweise erlauben, Kooperationsbeziehungen von Autoren oder Institutionen darzustellen, Zitationsanalysen, die die Identifikation von aktuellen „Forschungsfronten“ ermöglichen, oder sog. Co-Word-Analysen, die aus Titeln, Abstracts oder ganzen Texten Clusterungen von Texten nach inhaltlichen Gemeinsamkeiten vornehmen. Möglich sind ebenso Analysen, die Interdisziplinarität oder Internationalität zum Gegenstand haben.

Häufig entstehen dabei nicht unmittelbar Indikatoren, sondern Kartierungen („Mapping“) von Forschungsgebieten (vgl. Abbildung 3). Allerdings lassen sich solche Abbildungen von Strukturen in der Forschungslandschaft auch unter Leistungsgesichtspunkten interpretieren. Die Teilhabe an sich schnell entwickelnden, aktuellen Forschungsschwerpunkten kann z.B. als besondere Leistung gewertet werden. Die Abbildung zeigt die thematische Vernetzung eines Forschungsschwerpunktes. Die Karte beruht auf dem gemeinsamen Auftreten von Schlüsselbegriffen im Abstract, Titel oder Text eines Artikels und bezeichnet die „Nähe“ der Forschungscluster zu einander. Innerhalb dieser Cluster können nun wiederum sehr aktive Forschergruppen identifiziert werden.

Abbildung 3:

Co-word map of neural networks research 1985-1988



Quelle: VAN RAAN, A.F.J. (1994)

Ein praktisches Beispiel dafür, wie sich derartige Instrumente mit Verfahren der Experteneinschätzung (Delphi-Studien) zu einer „Forschungsprospektion“ kombinieren lassen, ist die Pilotstudie des WISSENSCHAFTSRATES zu einer „Prospektion der Forschung anhand ausgewählter Gebiete“ (1998). Dort wird für die „Materialwissenschaft“ nicht nur eine international vergleichende Bestandsaufnahme versucht, sondern es werden auch Entwicklungsperspektiven und Potentiale bestimmter Forschungsfelder ausgeleuchtet und die Möglichkeiten für Schwerpunktbildungen aufgezeigt.

Derartige empirisch fundierte Beschreibungen, die aus der Beobachtung des wissenschaftlichen Kommunikationssystems gewonnen werden, sind auch für die Erziehungswissenschaften in ersten Ansätzen vorhanden. Einen hervorragenden Überblick liefert dazu E. KEINER (1999). Allerdings ist der Kenntnisstand bisher doch eher fragmentarisch: Das hat einerseits mit Widersprüchlichkeiten und Uneindeutigkeiten in der Selbstbeschreibung der Disziplin zu tun, wohl auch mit Schwierigkeiten bei der Bestimmung des Zusammenhangs von allgemeiner Pädagogik und teildisziplinären Spezialitäten (z.B. Sozialpädagogik) und ebenso mit der starken Praxisorientierung in die unterschiedlichsten Handlungsfelder hinein (immerhin verfolgen über 80% der Hochschullehrer in den Erziehungswissenschaften mit ihrer wissenschaftlichen Arbeit praktische Anliegen; vgl. BAUMERT & ROEDER 1990). Entsprechend scheinen die fachlichen Kommunikationszusammenhänge zwischen For-

schung und Dogmatik, zwischen sozialwissenschaftlicher Empirie und geisteswissenschaftlicher Reflexion zu oszillieren, was alsdann die Frage aufwirft, ob Erziehungswissenschaft als Disziplin, im Sinne eines selbstreferentiellen Kommunikationszusammenhangs oder eher als „composite area of study“ verstanden werden kann. Die empirischen Befunde dazu zeigen immerhin erhebliche nationale Variationen. Für Deutschland kann man eine disziplinäre Selbstbezüglichkeit feststellen, von der allerdings unklar bleibt, ob sie aus einem starken paradigmatischem Kern erwächst oder aus einem Mangel an Austausch mit den Nachbardisziplinen. Andererseits verweisen viele Befunde auf eine „Normalisierung“ der Erziehungswissenschaften im Vergleich zu Nachbardisziplinen. Das betrifft die bekannten Konzentrationsprozesse im Bereich von Autorschaft, die soziale und statusmäßige Zusammensetzung der Profession, die Entwicklung von Koautorschaften und die Entwicklung typischer Zitationsmuster (vgl. KEINER 1999; SCHRIEWER & KEINER 1993).

Diese Unschärfen wirken sich auf eine indikatorengestützte Beschreibung aus: Es wird dann nämlich kompliziert, Indikatoren zu identifizieren, die das Leistungsspektrum des Faches in toto zu repräsentieren vermögen. Je diffuser die Definition disziplinärer Aufgaben ausfällt, desto heterogener wird auch die Definition der Leistungen, die diesen Aufgaben entsprechen, und es ist zugleich nicht nur mit einer Vielzahl von Bewertungskriterien für diese Leistungen zu rechnen, sondern auch mit einer Vielzahl von höchst unterschiedlichen Foren, in denen Erkenntnisansprüche verhandelt werden (vgl. WHITLEY 1982). Es wird also schwierig, konsensual gesicherte Beurteilungskriterien auszumachen.

Davon ist auch die Qualität der Indikatoren betroffen. Indikatoren sind ja per se keine Beschreibung einer wie immer gearteten Realität. Sie ähneln vielmehr sozialen Konventionen darüber, was als maßgeblich gelten soll. Deshalb operieren Indikatoren mit einer modellhaften Abstraktion des Forschungsprozesses, die voraussetzt, dass ein Mindestmaß an gemeinsamen Ziel- und Aufgabendefinitionen in einer Disziplin oder Subdisziplin vorhanden sind und ebenso ein Mindestmaß an Konventionen über die Beurteilung von Relevanz und Qualität der Forschungsbeiträge bzw. die Art und Weise, wie diese Beurteilung innerhalb des wissenschaftlichen Kommunikationssystems zum Ausdruck gebracht wird.

Fehlen derartige Konventionen, ist weiterhin die kognitive Struktur einer Disziplin unscharf, und besteht schließlich – wie im Falle der Erziehungswissenschaft – die Möglichkeit, sehr heterogene Praxisfelder als Beurteilungsauditorien zu adressieren, dann bedarf es einer sehr genauen Prüfung der Reichweite und Aussagekraft von Indikatoren.

In diesem Sinne lässt sich mit Hilfe von Strukturindikatoren und Expertise eine Art Kartographie einer Disziplin oder zumindest von Teilen einer Disziplin anfertigen. Zugleich ist dies aber auch die Ausgangsbasis für die Entwicklung eines anderen Typus von Indikatoren, nämlich Leistungs- und Effizienzindikatoren, die zum Vergleich von Forschergruppen, Fachbereichen, Forschungseinrichtungen etc. genutzt werden können. Der Übergang von Struktur- zu Leistungsindikatoren ist aller-

dings fließend. Auch Strukturindikatoren lassen sich evaluativ interpretieren: Die Kennzeichnung von „hot spots“ der Forschung, von Gebieten mit steigender oder sinkender Aktivität, hohen oder geringen Vernetzungen mit anderen enthalten genauso solche bewertenden Elemente.

Leistungs- und Effizienzindikatoren

Strukturindikatoren knüpfen nur bedingt an Leistungsdimensionen an und beschreiben statt dessen stärker die spezifischen Eigenschaften des Output- oder aber des Erstellungsprozesses. Sie bemühen sich um Deskription der Forschungstätigkeiten einer Untersuchungseinheit im Hinblick auf thematische Verortung, auf Kooperationsstrukturen, Interdisziplinarität, die Teilnahme an aktuellen Forschungsfronten etc.

Leistungsindikatoren werden hingegen meist vergleichend eingesetzt, um Aktivität, Leistungsstand oder Effizienz von Institutionen zu beurteilen. Sie gehören damit (unabhängig von ihrem faktischen Einsatz) in jedem Fall zu den Instrumenten einer professionellen Selbstbeschreibung. Dass sie auch von Dritten für Steuerungsprozesse im weitesten Sinne genutzt werden können, ändern daran nichts. Grob lassen sich diese Indikatoren danach unterscheiden, ob sie Inputgrößen beschreiben, Aktivität oder Partizipation messen, Qualität, Sichtbarkeit oder Akzeptanz wiedergeben, Strukturen und Prozesse abbilden oder aber subjektive Einschätzungen.

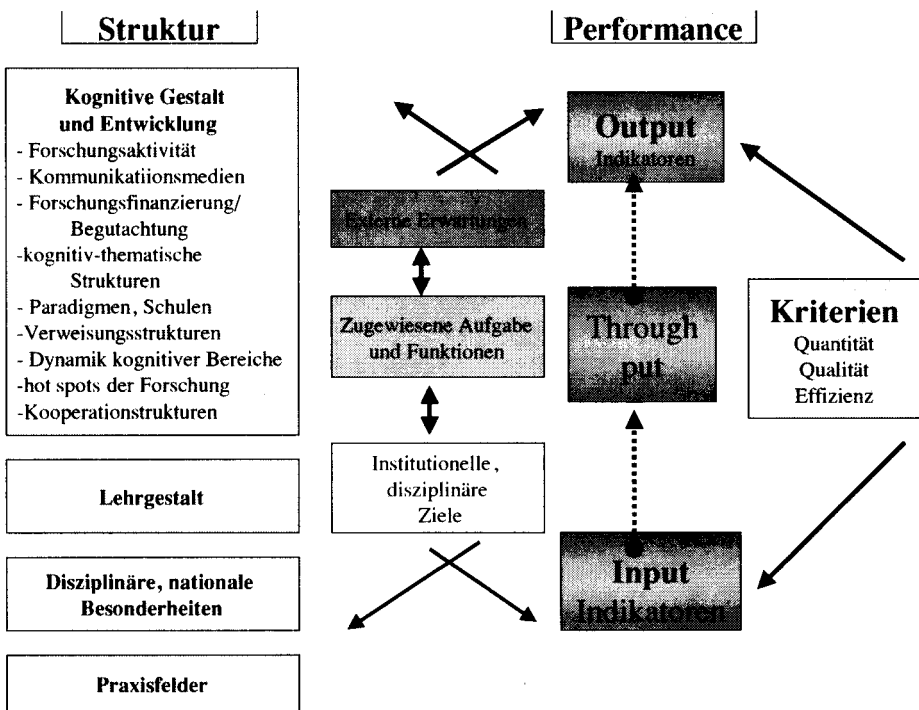
Aktivitätsindikatoren, z.B. Publikationszählungen, Drittmittelanträge, Patentanmeldungen berichten zunächst nur über die Präsenz im wissenschaftlichen Kommunikationssystem bzw. über die Teilhabe an Transferprozessen in das ökonomische System. Qualitätsindikatoren: z.B. Zitationsanalysen, Lizenanalysen, Bewilligungen von Drittmitteln greifen auf Bewertungsprozesse in der scientific community zurück oder auf Bewertungen von Nutzern von Forschungsergebnissen. Dabei geht es meist weniger um ein methodologisches Qualitätskonzept, als um Resonanz, Akzeptanz bestimmter Befunde oder Anknüpfung/Fortführung von Forschungslinien. Auf subjektive Einschätzungen wird man vor allen Dingen dann zurückgreifen müssen, wenn man mit dem klassischen Indikatorenkonzept nicht weiter kommt. Das dürfte häufig auf wissenschaftliche Aktivitäten in den Praxisfeldern zutreffen.

Auch für Leistungsindikatoren gilt, dass sie auf eine Zieldefinition oder ein Auftragsverständnis zurückgreifen müssen. Die Interpretation eines solchen Auftrages durch die Institution und die externen Erwartungshaltungen müssen dabei keineswegs völlig kompatibel sein. Teil eines Evaluationsprozesses wird daher immer auch die genaue Klärung der Ziele und angestrebten Entwicklungen (mission) sein müssen. Das fällt bei einem hoch spezialisierten Forschungsinstitut leichter als bei einer Einrichtung, in der Lehre, Service-, Forschungs- und Entwicklungsaufgaben gleichzeitig erfüllt werden sollen.

Leistungsindikatoren werden überwiegend zur Charakterisierung von Institutionen eingesetzt (Hochschulen, Fachbereiche, Forschergruppen, Forschungsinstitute etc.).

Damit stellt sich zunächst einmal die Frage nach Beurteilungskriterien: Was ist wesentlich für die Institution? Herausragende Forschungsleistungen, effiziente Mittelverwendung, breite Präsenz in einem Fachgebiet, interdisziplinäre Kontakte, internationale Kooperationen, Mitarbeit an Forschungsfronten oder zentralen fachlichen Themen, Zufriedenheit externer Klienten, Erfüllung definierter Qualitätsstandards? Die Liste lässt sich fast beliebig erweitern. Die einzelnen Punkte schließen sich auch nicht aus, aber sie setzen Akzente, die am Ende jeweils unterschiedliche Indikatorenkonstrukte nahe legen. Was ein geeigneter Indikator ist, bestimmt sich also zunächst einmal durch die Definition einer Leistungsdimension und dann durch eine Spezifikation des Leistungsbegriffs (vgl. Abbildung 4).

Abbildung 4:



Traditionell hat man in Deutschland den Forschungsbereich vor allen Dingen mit Input-Indikatoren beschrieben (vgl. etwa die Forschungsberichte der Bundesregierung). In den letzten Jahren haben sich aber immer stärker Output-Indikatoren durchgesetzt. Input-Indikatoren spielen aber insofern weiterhin eine wichtige Rolle, als sie zur Beurteilung von Aufwand/Ertrag-Relationen und zum Vergleich der „Effizienz“ von Institutionen notwendig sind. Was jeweils genau zum Input zu rechnen ist bzw. welche Anteile von Ressourcen zu berücksichtigen sind, ist in der Praxis häufig gar nicht so einfach zu bestimmen und wird nicht selten über Schätzungen

ermittelt (z.B. Vollzeitäquivalente des Personals für Forschung und Entwicklung (FuE) an den Hochschulen). Diese Unsicherheitszonen machen insbesondere Effizienzvergleiche anfällig für Fehleinschätzungen (vgl. DOCHY u.a. 1990).

Schwieriger zu ermitteln ist der Output. Zunächst muss dazu das „Produkt“ (s.o.) näher bestimmt und ggf. abgegrenzt werden (Forschung, Service, Lehre, Nachwuchsförderung etc.), sodann müssen geeignete Indikatoren zur Messung dieses Outputs bestimmt werden. Das geht nicht ohne ein gewisses Maß an konsensualer Festlegung von Konventionen (wie er in paradigmatisch verfestigten Forschungsgebieten - z.B. der Physik - durchaus existiert). Der ganz überwiegende Teil der Forschungsindikatoren greift dazu Bewertungsprozesse auf, die innerhalb der Wissenschaft ohnehin anfallen, und verdichtet sie zu numerischen Charakterisierungen. Derartige Bewertungen fallen in der Manuskriptbeurteilung, in der Beurteilung von Forschungsförderungsanträgen, in der Rezeption wissenschaftlicher Erkenntnisse und ihrer Dokumentation (Zitate), in der organisierten Honorierung von Leistungen (Preise, Ehrungen) und vielen anderen Formen der Begutachtung statt. Einige wenige Indikatoren versuchen auch die Wirkungen von Forschung in anderen gesellschaftlichen Subsystemen (Wirtschaft, Kultur) messbar zu machen.

Aussagefähigkeit von Indikatoren

War bisher die Rede davon, dass Indikatoren Phänomene wie Wissensstruktur, Kooperation, Fortschritt, Qualität, Relevanz, Aktivität etc. immer nur partiell unter Rückgriff auf die empirisch zugänglichen Phänomene des Kommunikationssystems modellieren können, soll im folgenden zumindest cursorisch auf die eher „technischen“ Probleme der Indikatorenbildung eingegangen werden.

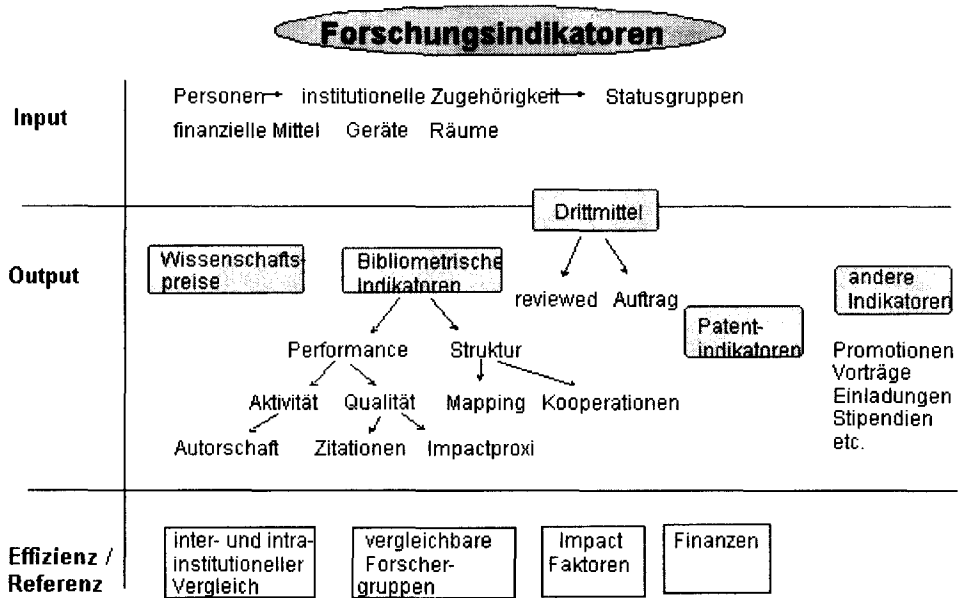
Bei der Entwicklung von Indikatoren sind immer auf mehreren Stufen Entscheidungen zu treffen, die Einfluss darauf haben, welcher Realitätsausschnitt modelliert wird.

Zunächst ist zu klären, welche Leistungen ein Indikator erfasst bzw. erfassen soll. Bereits auf dieser Stufe sind häufig Gewichtungen und Bewertungen notwendig, deren Gelingen entscheidend davon abhängt, inwieweit disziplinäre Konventionen über Wertigkeiten von Forschungsergebnissen vorhanden sind.

Am Beispiel von bibliometrischen Analysen und Drittmittelwerbungen seien die einzelnen Entscheidungsschritte und Konsequenzen kurz dargestellt (vgl. Abbildung 5): Zunächst muss entschieden werden, welche Publikationstypen berücksichtigt werden sollen. In der Physik fällt eine solche Entscheidung relativ leicht. Wie eingangs gezeigt, ist mit Artikeln in Fachzeitschriften und Kongressbeiträgen der ganz überwiegende Teil der Publikationen erfasst. Ebenso ist international die Bedeutung einzelner Fachzeitschriften zumindest in einer groben Reputationshierarchie unstrittig. In anderen Fachgebieten gilt das keineswegs. Daher sind bereits bei der Datensammlung Entscheidungen zu treffen. Keine Datenbank enthält vollständige

Informationen über die Publikationen im Fach, selbst die Bibliographien der Autoren sind häufig unvollständig, auf jeden Fall höchst heterogen. Um eine Gesamtbibliographie zu erstellen sind also Entscheidungen über die aufzunehmenden Publikationsgattungen, ihre Vollständigkeit bzw. die Selektionskriterien notwendig, ggf. sind auch Gewichtungen vorzunehmen. Praktisch bedeutet dies meist eine intensive Prüfung der verfügbaren Datenbanken.

Abbildung 5:



Ist eine Liste von Publikationen erstellt, müssen weitere Entscheidungen getroffen werden. Die wichtigste ist dabei die Zuordnung von Publikationen zu den Untersuchungseinheiten: Das können Personen, Forschergruppen, Labore, Abteilungen, Institute, Universitäten, nationale Forschungssysteme oder auch weltweite Entwicklungen sein.

Sobald es um kleinere Aggregate geht, wird die Zuordnung von Texten zu Personen oder Institutionen in dem Augenblick schwierig, wo es sich um mehrere Autoren handelt. In den Naturwissenschaften ist Mehrfachautorschaft inzwischen die Regel, in Extremfällen tauchen mehr als 100 Autoren auf. Eine einfache Zählung führt dann zu Ergebnissen, wonach die Spitzenschreiber etwa alle 4 Tage eine Publikation veröffentlichen. Der Grund liegt in institutionell durchaus unterschiedlichen Konventionen dafür, wer als Autor erscheint: Ob der Laborchef auf jeder Publikation als Autor geführt wird, ob die technischen Mitarbeiter als Koautoren auftauchen und ob die Autoren alphabetisch genannt werden; all das ist nicht wirklich konventionalisiert (vgl. HORNBOSTEL 1997).

Tab. 2: Koautorschaften von Artikeln in ausgewählten erziehungswissenschaftlichen Zeitschriften

Zeitschrift	Jahrgänge 1997, 1998, 1999		
	1 Autor	2 Autoren	3 u. mehr Aut.
	Zeilen%	Zeilen%	Zeilen%
Bildung und Erziehung	69,6%	30,4%	
Die Deutsche Schule	76,7%	17,8%	5,5%
Grundschule	85,6%	11,4%	3,0%
Pädagogik	89,4%	9,5%	1,1%
Pädagogische Rundschau	94,4%	4,6%	1,0%
Z.f. Sozialisationsforsch. u. Erziehungssoz.	57,4%	27,7%	14,9%
Zeitschrift für Erziehungswissenschaft	88,1%	9,5%	2,4%
Zeitschrift für Pädagogische Psychologie	50,9%	26,3%	22,8%

In den Erziehungswissenschaften ist mit solchen Verhältnissen zwar nicht zu rechnen. Aber bereits eine grobe Analyse der Häufigkeit der Koautorschaft von Artikeln in ausgewählten erziehungswissenschaftlichen Zeitschriften zeigt, dass zwar insgesamt Artikel mit nur einem Verfasser dominieren, aber gerade jene Zeitschriften mit fachlicher Nähe zur Soziologie und zur Psychologie einen deutlich höheren Anteil von Artikeln mit Mehrfachautorschaft aufweisen - ein Indiz dafür, dass teildisziplinäre Spezialisierungen sich auch in den Publikationsmodalitäten niederschlagen. Verschärfend kommt hinzu, dass Artikel, die in Mehrfachautorenschaft verfasst wurden, in der Regel höhere Zitationsraten erreichen. Entscheidet man sich für das Prinzip „one paper is one paper“ und rechnet den einzelnen Autoren - entsprechend der Gesamtzahl der Autoren einer Publikation - nur Bruchteile zu, dann schlagen jedoch insbesondere jene die Sichtbarkeit erhöhenden und forschungspolitisch erwünschten Kooperationsbeziehungen negativ zu Buch.

Hat man diese Probleme gelöst bzw. sich für ein Verfahren entschieden, verfügt man über eine Art Aktivitätsindikator (der allerdings u.U. auch die impliziten oder expliziten Qualitätsbeurteilungen der Datenbankbetreiber reflektiert oder durch Gewichtungen implizite Qualitätszuweisungen enthält); wohlgermerkt, einen disziplinspezifischen Indikator, denn die Länge von fachwissenschaftlichen Beiträgen und die Schreibhäufigkeit variieren erheblich zwischen den Forschungsgebieten.

Will man von hier zu einer Qualitätsgewichtung kommen, benötigt man eine Einschätzung der wissenschaftlichen Bedeutsamkeit. Üblich ist eine solche Einschätzung einzelner Artikel im Rahmen der Manuskriptbegutachtung. Eine derartige Bewertung durch Peers ist jedoch für die enormen Mengen von Publikationen, die im Rahmen bibliometrischer Analysen verarbeitet werden müssen, nicht möglich. Das bekannteste Verfahren, dem einzelnen Artikel ein Gewicht - unabhängig von neuerlichen Bewertungen - zuzuweisen, ist die Ermittlung erhaltener Zitationen. Zitationsanalysen sind (wie übrigens alle Qualitätsgewichtungen) umstritten:

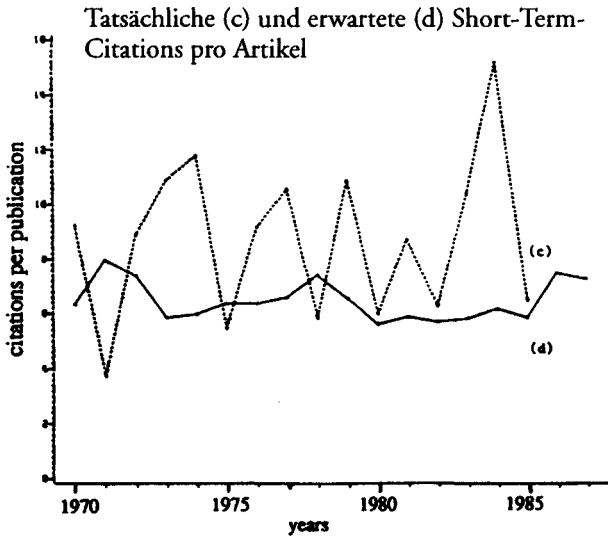
Zunächst einmal lassen sich nicht alle Zitate auffinden. Meist werden nur Referenzen aus Zeitschriften, die in einer der Datenbanken des ISI geführt werden, ausgewiesen. Zweitens wird immer wieder die Frage gestellt, wofür ein Zitat eigentlich steht. Idealerweise dokumentiert ein Zitat den Einfluss einer bestehenden Arbeit auf einen neuen Forschungszusammenhang. Tatsächlich reflektieren Zitationen natürlich neben Einfluss allerlei anderes. Z.T. figurieren sie als sog. Standard Symbols und bezeichnen einfach ein komplexes Verfahren oder eine Theorielinie, sodann wirkt das Matthäus-Prinzip, d.h. bekannte Autoren werden tendenziell häufiger zitiert; auch haben bestimmte Textgattungen (Überblicksartikel, Methodendarstellungen) häufig bessere Zitierchancen. Schließlich können auch innerhalb einer Disziplin sehr unterschiedliche Subkulturen hinsichtlich der Zitations- und Publikationsfrequenzen bestehen. Auf der anderen Seite werden in der Literatur allerdings immer wieder hohe Übereinstimmungen zwischen Peer-Urteilen, Auszeichnungen und anderen Forschungsindikatoren und Zitationsanalysen mitgeteilt. Beide Messungen gehen zwar nicht ineinander auf, weisen aber deutliche Korrelationsbeziehungen auf. Insgesamt kann man sich wohl am ehesten darauf einigen, dass Zitationen die Sichtbarkeit von Publikationen abbilden. Zitationen liefern keine Qualitätsbewertung im methodologischen Sinne, sondern Informationen über die Wahrnehmungen anderer Wissenschaftler.

Problematisch ist es weiterhin zu entscheiden, wie tief die Retrospektive angelegt werden soll. Publikationen haben fachspezifisch sehr unterschiedliche Halbwertszeiten. Eine wichtige Publikation in der Mathematik wird häufig über Jahrzehnte zitiert, während gerade in neuen, sehr dynamischen Forschungsgebieten die Literatur sehr schnell veraltet und dann nicht mehr zitiert wird. Ein sogenanntes Zitationsfenster, das angibt, wie lange nach Erscheinen einer Publikation die Zitate registriert werden, muss also so definiert werden, dass einerseits auch einigermaßen aktuelle Publikationen erfasst werden können, andererseits die Zitationsgewohnheiten der Disziplin berücksichtigt werden. Schließlich sind ggf. Selbstzitationen zu definieren und aus der Analyse zu entfernen.

Beim interdisziplinären Vergleich ist weiterhin zu beachten, dass die Zitierhäufigkeiten nach Forschungsgebieten sehr unterschiedlich ausfallen. Damit ist die Frage berührt, was die Messlatte für die Beurteilung der Indikatorwerte ist oder, anders formuliert, wie sich derartige Indikatoren normalisieren lassen. Häufig wird es notwendig sein, eine Bezugsgröße zu definieren. Das kann z.B. das durchschnittliche Zitationsaufkommen der jeweiligen Zeitschrift sein oder das weltweite Zitationsaufkommen der Subdisziplin (wenn sie denn klar zu begrenzen ist). Es ist aber auch möglich, erhaltene Zitate anhand der vergebenen Referenzen zu normalisieren.

Die folgende Abbildung zeigt z.B. einen Vergleich zwischen den (externen) Zitaten, die Publikationen aus einem Physik-Department der Universität Leiden erhielten (tatsächliche Zitationen) im Vergleich zu den Zitaten, die die übrigen Artikel in derselben Zeitschrift im Durchschnitt weltweit erhielten (erwartete Zitationen). Damit lässt sich zumindest grob identifizieren, ob eine Publikation durchschnittlich bzw. über- oder unterdurchschnittliche Resonanz in der scientific community erhielt.

Abbildung 6:



Quelle: VAN RAAN, A.F.J. (1994)

Ein anderer Weg zur Qualitätseinschätzung benutzt pauschale Beurteilungen von Artikeln derart, dass von Wissenschaftlern die Wichtigkeit oder das Qualitätsniveau einzelner Zeitschriften erfragt wird und auf dieser Grundlage dann ein Zeitschriften-set ausgewählt wird oder den einzelnen Artikeln Gewichtungen zugewiesen werden. Solche pauschalen Gewichtungen lassen sich auch mit Hilfe bibliometrisch gewonnener Gewichtungsfaktoren bewerkstelligen. Grundsätzlich werden dazu die vergebenen Referenzen und erhaltenen Zitate einer ganzen Zeitschrift verrechnet. Die daraus entwickelten Gewichtungsfaktoren unterscheiden sich nur in den Details ihrer Konstruktion (Impact Factor, Influence Weight). Das Problem solcher Gewichtungen liegt vor allen Dingen in der Varianz innerhalb der Zeitschriften: Auch sehr angesehene Zeitschriften weisen erhebliche Unterschiede in der Zitationshäufigkeit der einzelnen Beiträge auf. Dem einzelnen Beitrag wird man daher mit einer Zeitschriften-gewichtung in der Regel nicht gerecht. Bei größeren Publikationsmengen (über 100) nähern sich jedoch individuelle Zitationsgewichte und Zeitschriftengewichte an, wie auch verschiedene Gewichtungsverfahren dann tendenziell konvergieren.

Die Auswirkungen von Entscheidungen für oder gegen bestimmte Gewichtungs- und Auswahlverfahren werden besonders deutlich sichtbar, wenn aus den ermittelten Daten Effizienzmaße konstruiert werden. Die folgende Tabelle zeigt, dass, je nachdem ob man nur Publikationen zählt, Gewichtungen anhand von erhaltenen Zitaten vornimmt oder berücksichtigt, wie viele der Autoren tatsächlich aus dem analysierten Fachbereich stammen, mal die Konstanzer Physiker erfolgreicher sind, mal die Kölner, die sehr viel mehr Artikel in Kooperation mit Wissenschaftlern anderer Universitäten verfasst haben.

Tabelle 3: Vergleich verschiedener Effizienzmaße

Publikationen (1983-88) und Zitationen (in den ersten 3 Jahren nach Veröffentlichung) aus dem Forschungsgebiet Physik der Universitäten Köln, Konstanz und Bremen								
Universi- tät	I Zahl der Publikatio- nen *	II Mit Zitaten gewichtete Publikationen **	III Mit Zitaten ge- gewichtete Auto- renanteile ***	IV Personal am Fach- bereich ****	Relationen			
					I/IV	II/IV	III/IV	
Köln	662	1.455	440,9	47,3	14,0	30,8	9,3	
Konstanz	316	725	441,9	24,8	12,7	29,2	17,8	
Bremen	82	143	82,1	22,7	3,6	6,3	3,6	

alle Publikationen der Datenbank Scisearch, die als institutionelle Adresse den Fachbereich Physik oder entsprechende Institute und Lehrstühle aufweisen.

** Gewichtungsfaktoren: 0-1 Zitat: 1; 2-5 Zitate: 2; 6-10 Zitate: 3; 11-15 Zitate: 4; 16-20 Zitate: 5; >20 Zitate: 6

*** Gewichtungsfaktor: Autoren aus dem Fachbereich / Gesamtzahl der Autoren

**** Professoren wurden mit 1, Mittelbaustelle mit 1/2 und Qualifikationsstellen mit 1/3 gewichtet
Teilzeitstellen wurden als halbe Stellen gerechnet.

Quelle: HORNPOSTEL 1997

Steuerungsprozesse: Indikatorengestützter Wettbewerb

Zumindest drei unterschiedliche Verwendungskontexte sollten unterschieden werden: Der erste betrifft das Verhältnis von Institution und Öffentlichkeit oder öffentlichen Finanziers. Hier werden Indikatoren für eine Art Rechenschaftslegung eingesetzt. Geht die Initiative von der Institution aus, ist sie relativ frei in der Bestimmung der Indikatoren, kommt die Initiative von außen, wird sie nolens volens auf bestimmte Informationsanforderungen reagieren müssen. Diese Art von „accountability“ ist nur bedingt ein disziplinärer oder professioneller Selbstaufklärungsprozess. Er wird häufig von Außenstehenden angestoßen. Wichtig ist allerdings in diesem Zusammenhang eine Art rekursiver Wirkung der Verwendung von Indikatoren. Sind bestimmte Indikatoren einmal etabliert, werden sie als Konvention akzeptiert, geraten die Akteure in die Situation, auf eben jenen Indikatoren akzeptable Werte erreichen zu müssen. Bekannt sind insbesondere die unerwünschten Nebeneffekte eines solchen Verhaltens: publish or perish, die Zergliederung einer Publikation in viele kleine, Trainingsveranstaltungen für Doktoranden, in denen erläutert wird, mit welchen Strategien sich Zitationsraten steigern lassen etc. Zu den weniger sichtbaren positiven Effekten gehört, dass seit der Etablierung einer Rankingdebatte in Deutschland zum ersten Mal Fachbereiche ihre Stärken und Schwächen ernsthaft analysieren und Maßnahmen diskutieren. Wenn man im Hinblick darauf einmal den Studienführer von CHE und Stern/Start – der auf absehbare Zeit wohl umfassendste, regelmäßig erscheinende und vergleichende Evaluationsbericht – ansieht, stellt man fest, dass viele Aspekte des Evaluationskataloges der DGfE dort auftauchen (vgl. HORNPOSTEL

1999b, 2001; BARZ u.a. 2000). Auf eine derartige Berichterstattung wird man sich einstellen müssen. Und auch darauf, dass diese Informationen, die einen sehr breiten Leserkreis erreichen, Steuerungswirkungen im Sinne eines Reputationswettbewerbs entfalten. Angesichts der zunehmenden Erosion der Gleichheitsfiktion deutscher Hochschulen und eines enormen Informationsbedarf bei den Studienanfängern, ist durchaus zu erwarten, dass ein wachsender Teil von Studienanfängern jene durch Indikatoren abgebildete Performance von Fachbereichen in ihr Entscheidungsverhalten aufnimmt - wenngleich bisher nur eine kleine, allerdings sehr leistungsorientierte Gruppe von Studienanfängern andere Kriterien als die Nähe der Universität zum heimatlichen Herd in ihrem Entscheidungsprozess berücksichtigt (vgl. DANIEL 2000, HIS 1997).

Das Gegenstück zu solchen öffentlichen Leistungsvergleichen ist ein internes „monitoring“. Hier ist man weitgehend frei von externen Erwartungen. Die Fragen, die sich auf dieser Ebene stellen, beziehen sich eher darauf, wie ein aussagefähiges controlling system mit vertretbarem Erhebungsaufwand aussehen kann. Es geht dort nicht um einmalige großangelegte Evaluationen, sondern - neben eher informellen, nicht-öffentlichen Vergleichen (z.B. bench marking) - um eine kontinuierliche indikatorengestützte Selbstbeobachtung. Derartige systematische Selbstbeobachtungen ganzer Disziplinen oder auch einzelner Institutionen sind bisher eher die Ausnahme. Sie werden aber in einer Situation, in der Hochschulen zukünftig weitaus stärker um Studenten, öffentliche Mittel, Forschungsmittel und vor allen Dingen private Sponsorengelder konkurrieren müssen, zunehmend wichtig werden. Im Gegensatz zur „Rechenschaftslegung“ ist man hier relativ frei in der Bestimmung von Relevanzen und Indikatoren.

Wenn man auf die Universitäten, aber auch auf die Allokationspolitiken mancher Länder schaut, wird ein dritter Kontext deutlich, nämlich die Verbindung von Leistungs-, Belastungs-, oder Performanceindikatoren mit Mittel- oder Personalzuweisungen. Hier zeigt sich eine andere Eigenschaft von Indikatoren. Sie sind nämlich - einmal etabliert - so etwas wie institutionalisierte Dritte. Sie entlasten Zuweisungsmechanismen von Dauerverhandlungen. Sie etablieren sozusagen eine eigene Faktizität. Zugleich wird damit allerdings auch voluntaristisches Umsteuern schwieriger. Die Wahrscheinlichkeit, dass Zuweisungsmechanismen - wie in anderen europäischen Ländern längst üblich - auch in Deutschland mit Performanceindikatoren verknüpft werden, hat etwas mit basalen Funktionsdefiziten der akademischen Selbstverwaltung zu tun. Sie ist, obwohl formell auf Mehrheitsentscheidungen gegründet, faktisch durch „Zurückhaltung auf Gegenseitigkeit“ strukturiert. Unter Bedingungen einer extrem starken Absicherung professoraler Autonomie durch das Beamtenrecht und einer kaum hierarchisierten Organisationsstruktur mit hoher vertikaler Autonomie und geringen organisatorischen Interdependenzbeziehungen (hohe horizontale Autonomie), versacken Steuerungsimpulse häufig ohne jeden erkennbaren Effekt. Sie werden letztendlich im bekannten Gießkannenprinzip neutralisiert. Das ist, obwohl auch von den Beteiligten meist als suboptimal empfunden, nicht der Faulheit, Trägheit oder Desinteressiertheit der Professorenschaft geschuldet, sondern den organisatorischen Strukturen akademischer Selbstverwaltung. Indikatorgestützte

Allokationsmechanismen im Verein mit veränderten Leitungsstrukturen und stärkerem Wettbewerb sind ein Weg, unendliche und meist ergebnislose Gremienentscheide zu übersteuern. Sinnvoll wird eine solche Steuerung allerdings erst, wenn klare Zielvorstellungen definiert sind und die Steuerungswirkung solcher formelgebundenen Zuweisungen selbst kritisch verfolgt wird (vgl. WEINGART 1995).

Wollte man nun abschließend eine Antwort auf die im Titel gestellte Frage versuchen, dann müsste sie wohl lauten, dass die Frage falsch gestellt ist. Selbstbeobachtung jenseits eines sporadischen akademischen Interesses tritt erst auf, wenn - wie LUHMANN es formuliert - die Umwelt das System nicht ohne weiteres so akzeptiert, wie es sich selbst verstanden wissen möchte. Werden in einer solchen Situation Instrumente für eine systematische Selbstbeschreibung entwickelt, lässt sich ihr Gebrauch nicht auf eine folgenlose intellektuelle Übung begrenzen. Anstelle der Alternative „Selbstbeobachtung“ oder „Verteilungskampf“ muss man festhalten, dass Indikatoren für beides eingesetzt werden. Wichtiger als diese Unterscheidung scheint mir die Notwendigkeit, die Aussagekraft von Leistungsindikatoren anhand möglichst dichter Analysen der kognitiven und sozialen Struktur der jeweiligen Disziplinen zu prüfen und ggf. zu justieren.

Literatur

- BAUMERT, J. & ROEDER, P. M. (1990): Expansion und Wandel der Pädagogik. Zur Institutionalisierung einer Referenzdisziplin. In: Professionswissen und Professionalisierung. Braunschweig, S. 79-128.
- BARZ, A.; BUHR, P.; GIEBISCH, P.; HORNBOSTEL, S.; & MÜLLER-BÖLING, D. (2000): Studienführer 1999: Jura, Informatik, Mathematik, Physik. Vorgehensweise und Indikatoren. In: WEISHAUPT, H. (Hrsg.): Qualitätssicherung im Bildungswesen. Erfurter Studien zur Entwicklung des Bildungswesens 13, Pädagogische Hochschule Erfurt.
- DEUTSCHE FORSCHUNGSGEMEINSCHAFT (1997): Bewilligungen nach Hochschulen. Bewilligungsvolumen 1991 bis 1995. Anzahl kooperativer Projekte im Jahr 1996. Bonn.
- DEUTSCHE FORSCHUNGSGEMEINSCHAFT (2000): Bewilligungen an Hochschulen und außeruniversitäre Forschungseinrichtungen – 1996 bis 1998. Bonn.
- DANIEL, H. D. (2000): Folgeschwere Urteile. Der Einfluss der Spiegel-Rangliste auf Abiturienten. In: Unispiegel 2/2000.
- DOCHY, F.J.R.C.; SEGERS, M.S.R. & WIJNEN, W.H.F.W (Eds.) (1990): Management Information and Performance Indicators in Higher Education. Assen, Maastricht.
- GEGENWORTE. Heft 5, Frühling 2000.
- GILLET, R. (1989): Research Performances Indicators Based on Peer Review. A Critical Analysis. In: Higher Education Quarterly, 43, pp. 20-38.
- FRÖHLICH, G. (1999): Gegen-Evaluation: Das Peer-Review-System auf dem Prüfstand der Wissenschaftsforschung. Gesellschaft für Wissenschafts- und Technikforschung Jahrestagung 1999, Hamburg.
- HORNBOSTEL, S. (1997): Wissenschaftsindikatoren. Bewertungen in der Wissenschaft. Opladen.
- HORNBOSTEL, S. (1999a): Interaktionsprozesse zwischen Öffentlichkeit und Wissenschaft. Das SPIEGEL Ranking deutscher Universitäten und die Folgen. In: GERHARDS, J. & HITZLER, R. (Hrsg.): Eigenwilligkeit und Rationalität sozialer Prozesse. Opladen.
- HORNBOSTEL, S. (1999b): Evaluation und Ranking - Führen sie zu mehr Transparenz und Vergleichbarkeit? In: HRK: „Viel Lärm um nichts?“, Beiträge zur Hochschulpolitik 4, S. 77-80.

- HORNPOSTEL, S. (2001): Der Studienführer des CHE – ein multidimensionales Ranking. In: ENGEL, U. (Hrsg.): Hochschulranking. Zur Qualitätsbewertung von Studium und Lehre. Frankfurt/M. HIS (1997): Kurzinformation A5/97.
- INTERNATIONALE KOMMISSION zur Systemevaluation der DFG und der Max-Planck-Gesellschaft (Hrsg.) (1999): Forschungsförderung in Deutschland. Hannover.
- KEINER, E. (1999): Erziehungswissenschaft 1947-1990. Eine empirische und vergleichende Untersuchung zur kommunikativen Praxis einer Disziplin. Weinheim.
- KIESER, A. (1999): Über Marktmechanismen nachdenken. In: *Forschung & Lehre*, 6, S. 284-285.
- KREMPKOW, R. (1999): Ist "gute Lehre" meßbar? Untersuchungen zur Validität, Zuverlässigkeit und Vergleichbarkeit studentischer Lehrbewertungen, Marburg.
- OTTO, H.-U.; KRÜGER, H.-H.; MERKENS, H.; RAUSCHENBACH, Th.; SCHENK, B.; WEISHAUPT, H. & ZEDLER, P. (2000): Datenreport Erziehungswissenschaft. Befunde und Materialien zur Lage und Entwicklung des Faches in der Bundesrepublik. Opladen.
- RÖBBECKE, M. & SIMON, D. (1999): Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren im Wandel. Wissenschaftszentrum Berlin, P99-003.
- SCHIMANK, U. (1999): Professorenautonomie und Selbststeuerung der deutschen Hochschulen. In: *Das Hochschulwesen*, S. 11-14.
- SCHRIEWER, J. & KEINER, E. (1993): Kommunikationsnetze und Theoriegestalt: Zur Binnenkonstitution der Erziehungswissenschaft in Frankreich und Deutschland. In: SCHRIEWER, J.; KEINER, E. & CHARLE, Ch. (Hrsg.): *Sozialer Raum und akademische Kulturen*. Frankfurt/M., S. 277-341.
- TOULMIN, S. E. (1983): *Kritik der kollektiven Vernunft*. Frankfurt/M.
- VAN RAAN, A.F.J. (1994): Assessment of Research Performance with Bibliometric Methods. In: BEST, H. u.a. (Hrsg.): *Informations- und Wissensverarbeitung in den Sozialwissenschaften*. Opladen, S. 499-524.
- VORSTAND DER DEUTSCHEN GESELLSCHAFT FÜR ERZIEHUNGSWISSENSCHAFT (DGfE) (1998): Grundzüge eines Evaluationskatalogs für die Erziehungswissenschaft. In: *Erziehungswissenschaft*, 9, H. 17, S. 96-104.
- WEINGART, P. (1995): Forschungsindikatoren: Instrumente politischer Legitimierung oder organisatorischen Lernens. In: MÜLLER-BÖLING, D. (Hrsg.): *Qualitätssicherung in Hochschulen*. Gütersloh, S. 73-84.
- WISSENSCHAFTSRAT (1985): *Empfehlungen zum Wettbewerb im deutschen Hochschulsystem*. Köln.
- WISSENSCHAFTSRAT (1996): *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Berlin.
- WISSENSCHAFTSRAT (1998): *Pilotstudie zu einer Prospektion der Forschung anhand ausgewählter Gebiete*. Köln.
- WISSENSCHAFTSRAT (2000): *Thesen zur künftigen Entwicklung des Wissenschaftssystems in Deutschland*. Berlin.
- WHITLEY, R.D. (1982): The Establishment and Structure of the Sciences as Reputational Organisations. In: ELIAS, N. u.a. (Hrsg.): *Scientific Establishments and Hierarchies*. Dordrecht, pp. 313-357.