

Bibliometrie in der Forschungsevaluation. Zur Konstitution und Funktionslogik wechselseitiger Beobachtung zwischen Wissenschaft und Politik¹

Niels Taubert

Evaluation, Ranking, Rating, Leistungsmessung. Die Allgegenwart dieser Begriffe zeigt an, dass die Wissenschaft heute einer permanenten Außenbeobachtung mittels Wissenschaftsindikatoren unterliegt. Von besonderer Bedeutung für das Thema des vorliegenden Bandes sind dabei quantitative Indikatoren, die sich auf formale Merkmale der wissenschaftlichen Kommunikation beziehen, diese auswerten und so zu Aussagen über die Wissenschaft gelangen. Denn der Kontext ihrer Verwendung weist deutliche Parallelen zu manchen internet- oder onlinebasierten Strukturen auf. Ich meine hier die reflexive Figur, in der eine auf quantitativen Maßzahlen basierende Beobachtung eines sozialen Systems zur Voraussetzung, Ressource oder Rahmenbedingung für weiteres soziales Handeln im selben System wird. Diese Formen indikatorenbasierter Außenbeobachtung der Wissenschaft stehen im Mittelpunkt des Beitrags. Er fokussiert auf die Frage, welche Relevanz eine solche Art der Außenbeobachtung für die Steuerung der Wissenschaft hat und welche Form das Verhältnis von Wissenschaft und Politik hier annimmt.

In einem ersten Schritt werde ich die Entwicklung des Zitations-Index und die mit ihm verbundenen Vorstellungen über seinen Verwendungszusammenhang darstellen. Auf dem Zitations-Index basiert nur ein Teil der Indikatoren, andere werden auf der Grundlage weiterer Datenquellen berechnet.² Die Auswahl fiel aus zweierlei Grundung auf diesen Index. Erstens wertet er die formalen Merkmale der wissenschaftlichen Kommunikation – eben die Zitationen – aus und nutzt damit die Praxis von Wissenschaftlern, Leistungen anderer Autoren zu kennzeichnen und anzuerkennen, zu ganz anderen, eigenen Zwecken. Zweitens wurden mit dem Index weit reichende Vorstellungen hinsichtlich der Steuerung von Wissenschaft in Verbindung gebracht. Hier werde ich zeigen,

¹ Mein Dank gilt Matthias Winterhager, Holger Schwachheimer und Peter Weingart für Hinweise, Erläuterungen und Einblicke in den Zitations-Index und in die Praxis der Forschungsevaluation zu ungezählten Gelegenheiten.

² Klassische Wissenschaftsindikatoren sind die Summe eingeworbener Forschungsmittel (Drittmittel), die Anzahl an Patenten oder die in einer Institution beschäftigte Personenzahl.

dass im Laufe seiner Entwicklung eine mehrfache Umdeutung seines Zwecks vorgenommen wurde, so dass sich heute in ihm mehrere Zwecksetzungen überlagern.

Im zweiten Schritt interessieren mich die Eigenschaften von Beobachtungsinstrumenten, die auf dem Zitations-Index basieren. Hier argumentiere ich, dass sie Instrumente des Vergleichs sind und – da sie sich des Mediums der Zahlen bedienen – einer spezifischen Logik folgen. Im dritten Schritt wird der Fokus auf den Gegenstand verengt, und ich stelle zwei regelmäßig stattfindende Evaluationsaktivitäten vor: die UK Research Assessment Exercise und das australische Research Evaluation System. Damit verfolge ich ein doppeltes Ziel: Zum einen geht es mir darum, zu klären, welche Rolle auf Zitations-Indexen basierende Indikatoren empirisch in Prozessen der Leistungsmessung und -bewertung spielen. Wie zu zeigen ist, verändert sich die Zusammensetzung der Indikatoren fortwährend, so dass deren Bedeutung schwankt und trotz der mit zum Teil großer Vehemenz vorgebrachten Kritik an diesen Indikatoren festgehalten wird. Zum anderen geht es mir darum, zu analysieren, welche Reaktionen solche Evaluationsaktivitäten evozieren. Der vierte Schritt bildet die Theoretisierung des Phänomens. Hier wird die Frage beantwortet, welche Form das Verhältnis von Wissenschaft und Politik in Fällen indikatorenbasierter Beobachtung annimmt. Ich argumentiere, dass die gängige Perspektive, die in der indikatorenbasierten Mittelvergabe einen einseitigen Steuerungsprozess der Wissenschaft sieht, zu kurz greift: Da Anpassungsprozesse sowohl in der Wissenschaft als auch in der Politik zu beobachten sind, sollte – so mein Vorschlag – die Schaffung derartiger Instrumente als Implementation eines wechselseitigen Beobachtungsverhältnisses verstanden werden. Dabei gilt, dass das Spiel mehrere Runden hat und die Entwicklung von Indikatoren und die der Medientechnik ein treibendes Moment bilden. Der Aufsatz schließt mit einem Ausblick auf die Zukunft dieses Verhältnisses und dem Argument, dass der Weg zurück in die Zeiten vor Evaluation und Leistungsmessung verstellt ist. Ob es uns gefällt oder nicht: Wissenschaft und Politik werden sich an das wechselseitige Beobachtungsverhältnis gewöhnen müssen.

1 Umdeutungen während der Entwicklung der Bibliometrie

Sieht man einmal von der Community ab, die Bibliometrie zu Zwecken der Wissenschaftsforschung nutzt, ist aus heutiger Sicht die Bibliometrie auf das engste verknüpft mit Prozessen der Leistungsmessung und -bewertung, Evaluation und wissenschaftspolitischen Steuerung. Eine solche Zweckzuschreibung gewannen die zum Einsatz kommenden Instrumente allerdings erst allmählich im Zuge

eines mehrschrittigen Prozesses, der durch Umdeutungen und neu hinzukommende Zielsetzungen gekennzeichnet ist. Im folgenden Abschnitt rekonstruiere ich die Entwicklung dieses Prozesses am Beispiel des Zitations-Index, der innerhalb des Forschungsfeldes eine zentrale Stellung einnimmt.

1.1 Frühgeschichte³: der Zitations-Index als Rechercheinstrument

Die Entstehung der Bibliometrie, verstanden als „the application of mathematics and statistical methods to books and other media of communication“ (Pritchard 1969: 348), ist auf das engste verknüpft mit der Person Eugene Garfield. Zwar gab es vor und nach ihm Wissenschaftler unterschiedlicher Provenienz, die sich konzeptionell und in der Entwicklung bibliometrischer Methoden hervorgetan haben. Garfield kann aber insofern eine herausgehobene Stellung beanspruchen, als auf ihn die Entwicklung eines Instruments zurückgeht, das eine wesentliche Ressource für eine Vielzahl von bibliometrischen Forschungsprojekten bildete und immer noch bildet. Gemeint ist der Science Citation Index (SCI), der über mehrere Dekaden das Forschungsgeschehen im Bereich der Bibliometrie stark prägte.

Garfield beschäftigte sich seit etwa Mitte der 1950er Jahre mit der Idee des Aufbaus eines Zitations-Index für die Wissenschaft, der ab 1961 im Rahmen eines von den National Institutes of Health finanzierten Projekts systematisch entwickelt wurde. Zu diesem Zweck wurden für sämtliche Artikel, die in 563 Zeitschriften erschienen waren, Zitationen extrahiert, gesammelt und erstmalig 1963 in gedruckter Form veröffentlicht. Aufschlussreich ist die ursprüngliche Zielsetzung, die Garfield dabei verfolgte. Der Zitations-Index war von ihm keineswegs dazu gedacht, die Wissenschaft zu vermessen oder einer quantifizierenden Außenbeobachtung zugänglich zu machen. Seine Idee war zunächst vielmehr, Wissenschaftlern ein Werkzeug an die Hand zu geben, das ihnen bei der Recherche nach Literatur helfen sollte. Dieses Ziel zeigt sich im folgenden Zitat, in dem Garfield erstmalig die Idee eines Zitations-Index einer größeren Öffent-

³ Die folgende Darstellung setzt zeitlich mit der Entstehung eines Zitations-Index ein, mit dem die Bedingungen der Möglichkeit einer regelmäßigen, begleitenden Beobachtung der Wissenschaft geschaffen wurden. Die Überlegung, wissenschaftliche Aktivitäten im Allgemeinen und wissenschaftliche Publikationen im Besonderen mittels Zahlen zu messen, ist allerdings wesentlich älter. Heute noch bekannte Pioniere sind F.J. Cole und N.B. Eales, P.L.K. Gross und E.M. Gross und vor allem A.J. Lotka (Schmidmaier 1984: 404 f.). Diese frühen Untersuchungen, die sich quantitativer Mittel der Analyse von Publikationsaktivitäten bedienten, zeichnen sich dadurch aus, dass die Datengrundlage jedes Mal und entsprechend der Anforderungen des jeweiligen Erkenntnisinteresses in mühevoller und zeitraubender Arbeit zusammengetragen werden musste. Eine permanente quantifizierende Beobachtung der Wissenschaft war mit diesen Mitteln nicht möglich.

lichkeit vorstellt: „It (gemeint ist der Zitations-Index, NT) is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires.“ (Garfield 1955: 108). Der avisierte Verwendungszusammenhang, in den Garfield den Zitations-Index stellt, tritt in seinem Text auch in der Alternative zutage, die er dem Zitations-Index gegenüber stellt: Er vergleicht ihn mit einem Subject-Index, also einem Index, bei dem Publikationen bestimmten Fachgebieten zugeordnet werden. Die Gemeinsamkeit der beiden Indexe sieht Garfield in ihrem Hilfsmittelcharakter zum Auffinden von wissenschaftlicher Literatur.

1.2 Erste Umdeutung: der Zitations-Index als Forschungsinstrument

Die Erleichterung des Auffindens von Literatur durch einen Zitations-Index ist ohne Frage ein Zweck, der mit dem Systemziel der Wissenschaft, der Vergrößerung von wissenschaftlichem Wissen, im Einklang steht. Mindestens schafft eine solche Erleichterung mehr Freiraum für Forschung, im günstigen Fall stellt der Index die oben angesprochenen Querverbindungen her, die originelle Einsichten mit sich bringen und damit kreativitätsunterstützend wirken können. Sehr schnell wurde allerdings deutlich, dass ein solches Recherche-Instrument auch ganz andere Möglichkeiten bietet.

Die Idee, die Mittel der Wissenschaft auf sie selbst anzuwenden, Wissenschaft also wissenschaftlich zu beforschen, lag in der Luft⁴ und machte auch vor dem Zitations-Index nicht halt. Bereits im Zuge der Abschätzung der Praktikabilität eines solchen Index gab es die Überlegung, ihn als Forschungsinstrument einzusetzen. Hierzu Garfield:

„In fact, a comprehensive, multidisciplinary index possessed a dimension of utility that went beyond its role as a search tool. It also provided a view of the literature that threw much light on such murky and important subjects as the bounds of particular disciplines, the structure of the journal network that is the primary mechanism for exchanging information in the scientific world, the historical development of scientific thought, and the implications and impact of individual pieces of scientific work.“ (Garfield 1979: 16)

⁴ Prominentes Beispiel dafür ist Derek de Solla Price, der in dieser Zeit an seiner Untersuchung des Größenwachstums der Wissenschaft arbeitete. Zu diesem Zweck kombinierte er verschiedene Indikatoren, die sich auf die personale und organisationale Dimension sowie den Publikationsoutput der Wissenschaft bezogen. Als Ergebnis stellte er fest, die Wissenschaft wachse gemessen an der Anzahl an Wissenschaftlern, Universitäten, wissenschaftlichen Zeitschriften und Publikationen sehr deutlich und die Zahl der in der Wissenschaft tätigen Personen verdoppele sich etwa alle 15 Jahre (De Solla Price 1963: 23).

Mit dieser zweiten Zwecksetzung verändert sich die Rolle von Zitationen. Gelten im ersten Verständnis des Index Zitationen gewissermaßen als Klammern zwischen Texten, die man einzeln nutzt, um die Wirkung von Gedanken in Richtung der Gegenwart zu verfolgen oder ihre Ursprünge in Richtung der Vergangenheit zu rekonstruieren, werden Zitationen hier nun in aggregierter Form ausgewertet. Nicht das einzelne Zitat interessiert mehr, sondern die sich ergebenden Muster bei der Massendatenanalyse von Zitationen, die Rückschlüsse auf Bedeutungen von Forschergruppen, Instituten, nationalen Forschungssystemen oder die Dynamik von Forschungsfeldern zulassen. Eine solche Verwendung des Zitations-Index bewegt sich zwar im Rahmen des Bezugssystems der Wissenschaft selbst, hat aber im Vergleich zu der obigen Zweckbestimmung eine deutlich verschobene Konnotation zur Folge. Der Index bildet Mittel zur Realisierung einer Beobachtung der Kommunikation der Wissenschaft durch einen Teil der Wissenschaft selbst, nämlich der empirischen Wissenschaftsforschung.

1.3 Zweite Umdeutung: der Zitations-Index als Instrument der Wissenschaftsevaluation

Damit ist die Verschiebung der Zweckbestimmung des Zitations-Index keineswegs an ihrem Ende angekommen, denn der folgende Gedanke liegt sehr nahe: Wenn durch die Bibliometrie eine Beobachtung der Wissenschaft durch die Wissenschaft möglich ist, lässt sich dann nicht auch die Bibliometrie dazu nutzen, um die Wissenschaft von außen und unter ganz anderen Vorzeichen zu beobachten? Hinweise auf eine solche Zwecksetzung finden sich wiederum sehr früh: Garfield selbst thematisiert die Möglichkeit einer „missbräuchlichen“ Verwendung bereits in einem Aufsatz aus dem Jahr 1963, in dem er schreibt: „One purpose of this communication is to record my forewarning concerning the possible promiscuous and careless use of quantitative citation data for sociological evaluations, including personnel and fellowship selections.” (Garfield 1969 (1963): 44).

Mit der weiteren Arbeit am Zitations-Index, den immensen Kosten und der Fortschreibung des Index durch das Institute for Science Information (ISI) wichen diese vorsichtigen Warnungen dann einer aktiven Bewerbung des Potentials des Index, gerade auch im Zusammenhang von Leistungsmessung und -bewertung. So heißt es beispielsweise in den Ausführungen des sechsten, mit „A Science-Management Tool“ überschriebenen Kapitels der Monographie „Citation Indexing – Its Theory and Application in Science, Technology and Humanities“:

“In this regard, the SCI data base is being used to do such things as evaluate the research role of individual journals, scientists, organizations, and communities; refine the relationship between journals and between journals and fields of study; measure the impact of current research; provide early warnings of important, new interdisciplinary relationships; spot fields of study whose rate of progress suddenly begins accelerating; and define the sequence of developments that led to major scientific advances.“ (Garfield 1979: 62)

Der Fairness halber muss erwähnt werden, dass zwischen diesen konträren Einschätzungen der Brauchbarkeit des Zitations-Index zu Zwecken der Leistungsmessung und -bewertung anderthalb Dekaden intensiver Forschung zur Validität der Messung von ‚Qualität‘ wissenschaftlicher Publikationen mittels Zitationen liegen.

Es wäre zu kurz gegriffen, würde man die Ursachen für die neuerliche Umdeutung des Verwendungszusammenhangs allein im Ressourcenbedarf der Aufrechterhaltung und Pflege des Index und in dem von ihm bereitgestellten Potential sehen. Hinzu kommt ein großes Interesse der Politik, Wissenschaft von außen zu steuern. Grundlegende Probleme solcher Steuerungsversuche liegen in der Esoterik der Wissenschaft: Der Politik fehlen gewöhnlich die Kompetenzen, um die Leistungen unterschiedlicher Forschungseinrichtungen oder gar einzelner Wissenschaftler unter Anwendung der Exzellenzkriterien des betreffenden Fachs zu beurteilen und so förderungswürdige von weniger förderungswürdiger Wissenschaft zu unterscheiden.⁵ Vor diesem Hintergrund wurde gefragt, ob vielleicht der Zitations-Index eine Lösung für das Informationsdefizit der Wissenschaftspolitik offeriere. Eine frühe Artikulation eines solchen Interesses am Zitations-Index stammt aus einem Brief des Vorsitzenden des National Science Board der National Science Foundation an den US-amerikanischen Präsidenten. In diesem dem Bericht „Science Indicators 1972“ beigefügten Schreiben beklagt er methodische Defizite insbesondere im Bereich der Output-Indikatoren (vorrangig Publikationen) und fordert eine Weiterentwicklung mit folgender Zielsetzung:

„If such indicators can be developed over the coming years, they should assist in improving the allocation and management of resources for science and technology, and guiding the Nation’s research and development along path most rewarding for our society... The Report represents only an initial step toward a system of science

⁵ Vgl. hierzu Weinberg, 1963. Dieser unterscheidet zwischen internen und externen Kriterien für die Vergabe von Forschungsmitteln, wobei er für die internen, innerwissenschaftlichen Exzellenzkriterien bemerkt, sie können nur von Wissenschaftlern des jeweiligen Fachs gehandhabt werden (insbes. 163).

indicators. The further development of such indicators is a matter of high priority for future reports in this series.“ (Carter 1973: iii)

Dieses Zitat zeigt, dass eine neue Zwecksetzung in den Blick gekommen war. Auf der Basis des Zitations-Index soll eine Außenbeobachtung der Wissenschaft durch die Wissenschaftspolitik möglich werden und entsprechende Allokationsentscheidungen daran angeschlossen werden. Der Vorteil eines solchen Instruments liegt auf der Hand: Die politischen Entscheidungsträger gewinnen an Autonomie, da sie zwar von der bibliometrischen Forschung abhängig bleiben, sich aber von der Qualitätseinschätzungen der Vertreter des jeweiligen Fachs emanzipieren, die notorisch unter Verdacht stehen, bei Allokationsentscheidungen eigene Interessen zu verfolgen. In Bezug auf diesen Verwendungszusammenhang kann von einem gewissen Unschuldverlust des Zitations-Index gesprochen werden. Indem er in den Kontext wissenschaftspolitischer Entscheidungen gerückt wird, dient er nicht mehr nur innerwissenschaftlichen Relevanzgesichtspunkten, sondern auch politischen Prioritäten.

2 Der Zitations-Index als Vergleichsinstrument

Zum Verständnis der Umdeutung des Zitations-Index und den Konsequenzen für das Verhältnis von Wissenschaft und Politik lohnt es sich, einen Augenblick beim Instrument selbst zu verweilen und nach seinen Eigenschaften zu fragen. Für die Logik quantitativer Vergleiche interessiert sich Bettina Heintz in grundsätzlicher Weise. Vergleiche in einem ganz allgemeinen Sinne kombinieren Heintz zufolge eine Gleichheitsunterstellung, also das Konstatieren der prinzipiellen Vergleichbarkeit der verglichenen Elemente, mit einer Differenzbeobachtung, dem Abweichen des miteinander Verglichenen in Bezug auf eine bestimmte Dimension (Heintz 2010: 164). Vergleichbarkeit ist dabei keine den Objekten anhängende Eigenschaft, sondern Resultat einer sozial voraussetzungsvollen Kategorisierung, in deren Zug „einige Merkmale als relevant ausgewählt werden und alles andere ignoriert wird“ (Heintz 2007: 74). Im Fall des Zitations-Index wird die grundlegende Vergleichbarkeit hergestellt durch die implizit mitgeführte und nicht im einzelnen geprüften Annahme, es handle sich bei dem erfassten Textkorpus um eine Sammlung wissenschaftlicher Texte,⁶ die sich durch das Merkmal der Verknüpfung durch Zitationen auszeichnet. Die Konstruktionsleistung setzt sich fort in der Auswahl eines Ausschnitts aus der Gesamtmenge der

⁶ Dass es sich dabei um eine zum Teil prekäre Unterstellung handelt, zeigt sich beispielsweise im Fall von Editorials, die häufig keine Wahrheitsbehauptungen aufstellen und zu begründen versuchen. Sie werden aber sehr wohl für die im Science Citation Index erfassten Zeitschriften mit abgedeckt.

wissenschaftlichen Kommunikation, die dem ISI zufolge für die Wissenschaft von besonderer Bedeutung ist. Zusammengestellt wird der Korpus auf der Grundlage der Auswahl von Journalen und Konferenz-Proceedings und nicht etwa durch die Selektion einzelner Texte. Eine Aufnahme eines Journals in den Index findet nach einer eingehenden Prüfung statt, in der Kriterien wie Begutachtungsstandards, die Internationalität der Autorenschaft und Umfang der Zitationen in bereits im Index aufgenommenen Journalen zur Anwendung kommen.⁷ Der Science Citation Index Expanded (SCIE) erfasst derzeit Artikel aus mehr als 7.100 Journalen.⁸ Die Kategorisierung geht aber noch weiter: Neben der Auswahl der als bedeutsam erachteten Texte gibt es noch eine zweite Setzung von Relevanz. In einem Zitations-Index werden nicht die wissenschaftlichen Publikationen in ihrer Totalität berücksichtigt, sondern nur wenige Eigenschaften der Texte. Die Kategorisierung abstrahiert nahezu völlig vom Inhalt und erfasst nur die formale Eigenschaften wie Metadaten der einzelnen Publikationen (z.B. Titel, Name des oder der Autoren, Publikationsort, Verlag usw.), Abstracts sowie die Verweise zwischen den Texten (Zitationen). Für jede Zitation wird dabei die Referenz des zitierten und des zitierenden Textes erfasst.

Vergleiche zwischen Elementen des Wissenschaftssystems (wie z. B. Personen, Forschungseinrichtungen, Journale usw.) werden nun nicht direkt durch den Zitations-Index vorgenommen, sondern basieren auf Indikatoren, die auf der Grundlage des Zitations-Index errechnet werden. Das sicherlich bekannteste Beispiel für einen solchen Indikator ist der Journal Impact Factor (JIF), der eine Maßzahl zur Beurteilung des relativen Einflusses eines Journals auf ein Forschungsfeld darstellt. Er setzt dazu zwei Zahlen in Beziehung: (a) die Anzahl der Zitationen eines Jahrgangs des SCIE, die sich auf diejenigen Artikel eines Journals beziehen, die in den vorangegangenen beiden Jahren publiziert wurden, mit (b) der Anzahl der Artikel, die das betreffende Journal in diesem Zweijahres-Zeitraum publiziert hat (Garfield 2006: 90). Ein weiterer bekannter Indikator, der sich auf Personen bezieht, ist der sogenannte Hirsch-Index (h-Index). Diese sehr einfache Maßzahl besteht aus der Anzahl von Papieren (h) eines Autors, die mindestens h Zitationen auf sich gezogen haben (Hirsch 2005).

Ich führe diese beiden Indikatoren als Beispiele an, um zu zeigen, dass nicht nur durch den Zitations-Index selbst, sondern auch durch die Indikatoren weitere Relevanzsetzungen in den Vergleich eingeführt werden. Im Fall des Impact-

⁷ Eine Beschreibung der Auswahlprozedur findet sich auf der folgenden Webseite: <http://wokinfo.com/benefits/essays/journalselection/> (Stand 10/2011).

⁸ Siehe http://wokinfo.com/media/pdf/WoSFS_08_7050-1.pdf (Stand 10/2011). Beim SCIE handelt es sich nur um einen (wenngleich wichtigen) Teil des Web of Science. Neben dem SCIE umfasst das Web of Science auch noch die folgenden Zitations-Indexe: den Social Science Citation Index mit ca. 2.100 Journalen, den Arts & Humanities Index mit 1.200 Journalen und den Conference Proceedings Citation Index.

Factors ist dies beispielsweise der Referenzzeitraum von zwei Jahren, von dem unterstellt wird, er sei hinreichend, um die Qualität eines Journals einzuschätzen und zu vergleichen. Dass es sich dabei um eine Setzung handelt, die durchaus auch anders getroffen werden kann, zeigt ein Blick auf geistes- und sozialwissenschaftliche Journale: In diesen Disziplinen sind z.T. deutlich längere Rezeptionszeiten üblich, so dass der Impact-Factor in diesen Disziplinen wenig aussagekräftig ist. Im Fall des h-Index liegt eine Relevanzsetzung in der Entscheidung, einen Faktor zu bilden, in den nur die jeweils hoch-zitierten Papiere Eingang finden und beispielsweise nicht die absolute Zahl an Publikationen oder die Summe an Zitationen. Diese Setzungen führen zu dem folgenden (hypothetischen) Ergebnis: Ein Wissenschaftler, der ein einziges hoch-zitiertes und bahnbrechendes Papier publiziert, hat denselben h-Index (von 1) wie ein Autor, der eine Vielzahl von Artikeln publiziert, die jeweils ein Mal zitiert wurden, und wie der Novize, der bislang nur einen einzigen Beitrag veröffentlicht hat und der nur ein Mal zitiert wurde.

Neben dem Aspekt der Kategorisierung interessiert sich Heintz auch für das Medium solcher Vergleiche. Ihre These ist, dass ‚Zahlen‘ ebenso wie die Schrift und das Bild eine mediale Eigenlogik besitzen und es damit keineswegs beliebig ist, welches Medium für einen Vergleich genutzt wird. Im Fall des Mediums Zahlen besteht die ‚numerische Differenz‘ (oder die Besonderheit des Mediums) darin, dass sie vorgeben, eine Realität abzubilden, die außerhalb der Zahlen liegt. Dadurch verdecken Zahlen in der Tendenz die Eigenschaft von Vergleichen, selektive Konstruktionen zu sein (Heintz 2010: 171), und steigern die Wahrscheinlichkeit ihrer Akzeptanz. Die Ursachen für diese Wirkung von Zahlen sieht Heintz in den folgenden Charakteristika. Erstens sind Zahlen in Bezug auf die Erzeugungsregeln im Vergleich zur Sprache in einem wesentlich höheren Maße selbstexplikativ und eindeutiger (ebd.: 172). Zweitens besitzen sie ein hohes Maß an Selbstreferentialität und nur einen geringen äußeren Bezug. Dieser beschränkt sich auf das jeweils Gemessene des Messvorgangs (ebd.: 173). Drittens ist der Umgang mit Zahlen universell, so dass sie leichter kulturelle Grenzen überschreiten (ebd.), und viertens legen Zahlen eine externale Zurechnung nahe. Damit ist gemeint, dass zahlenmäßig ausgedrückte Befunde nicht dem eingreifenden, ordnenden oder kategorisierenden Handeln des Zahlenproduzenten, sondern den gemessenen Objekten zugerechnet werden (ebd.: 174).

Diese allgemeineren Überlegungen zu zahlenmäßigen Vergleichen liefern eine Erklärung dafür, weswegen quantifizierende bibliometrische Untersuchungen für die Politik besonders attraktiv sein dürften: Folgt man den Überlegungen von Heintz, so wäre zum einen zu vermuten, dass sie innerhalb der Politik eine hohe Überzeugungskraft haben, da ihnen ein höheres Maß an Objektivität zugeschrieben wird. Zum anderen liegt die Hypothese nahe, dass quantitative Maß-

zahlen zur Legitimierung wissenschaftspolitischer Steuerungen besonders geeignet sind. Die Vermutung wäre hier, dass sie den Eindruck besonderer Objektivität und der Sachangemessenheit wissenschaftspolitischer Entscheidungen in der Öffentlichkeit erzeugen.

3 Zwei Beispiele indikatorenbasierter Wissenschaftspolitik

Nachdem ich in den vorangegangenen beiden Abschnitten die Entwicklung des Zitations-Index und die Institutionalisierung der Bibliometrie kursorisch dargestellt, die Logik quantifizierender Vergleiche untersucht und Hypothesen zur Verwendung von quantitativen Wissenschaftsindikatoren formuliert habe, gehe ich im folgenden Abschnitt der Frage nach, welche Rolle Zitationsanalysen in der empirischen Realität der Wissenschaftspolitik spielen. Zu diesem Zweck stelle ich zwei Systeme der leistungsbezogenen Vergabe von Forschungsmitteln vor. Sie sind von besonderem Interesse, da an ihre Leistungsmessung und -bewertung wissenschaftspolitische Entscheidungen über die Vergabe von Forschungsmitteln angeschlossen werden und diese Instrumente damit unmittelbare Konsequenzen für die Forschungseinrichtungen haben. Es handelt sich zum einen um das australische System der leistungsbezogenen Vergabe öffentlicher Forschungsmittel (Australian Research Evaluation System) und zum anderen um die britische Research Assessment Exercise (RAE) und deren Nachfolger, das Research Excellence Framework (REF). Als Beispiele wurden sie gewählt, da in den beiden Ländern bereits umfangreiche Erfahrungen mit dieser Art von Instrument vorliegen. Der Fokus der Darstellung liegt hier nicht mehr nur auf den Instrumenten der Leistungsmessung selbst, sondern auch auf den Rückwirkungen dieser Aktivitäten in Wissenschaft und Politik.

3.1 Das Australian Research Evaluation System (RES) und seine Nachfolger

Das australische System der Forschungsförderung von Universitäten ist dual organisiert und setzt sich zusammen aus einer Grundfinanzierung und einer leistungsbezogenen Finanzierung. Hier ist ausschließlich die leistungsbezogene Mittelvergabe von Interesse, die sich ihrerseits aus zwei Komponenten zusammensetzt. Die beiden wichtigsten Forschungsförderer – das Health and Medical Research Council (NHMRC) und das Australian Research Council (ARC) – vergeben den Großteil ihrer Mittel in Form der Projektförderung, also als Zuwendung nach einer positiven Begutachtung von Projektanträgen. Vom Grundprinzip her entspricht dieses Verfahren im deutschen System der Mittelvergabe

durch das DFG-Normalverfahren. Der verbleibende Teil wird auf der Basis einer indikatorenbasierten Leistungsmessung vergeben. Zu diesem Zweck teilen die Universitäten seit 1993 dem Department of Education, Science and Technology (DEST) verschiedene Daten mit. Hierzu zählen die Summe der Forschungsmittel, der Forschungsoutput in Form von Publikationen sowie die vergebenen höheren akademischen Grade (Master und PhD) (Geuna/Martin 2003: 294). Die erste leistungsbezogene Allokation öffentlicher Mittel fand in Australien im Jahr 1996 statt und wird seitdem regelmäßig wiederholt (Butler 2003: 40). Während in den ersten Jahren für den Indikator ‚Veröffentlichungen‘ eine Vielzahl von Publikationsformen berücksichtigt wurden, zählen seit 1995 nur noch Bücher, Buchkapitel, begutachtete Journal-Artikel und begutachtete Beiträge in Konferenzbänden.

Die Leistungsmessung im Bereich des Forschungsoutputs folgt dabei einem vergleichsweise einfachen Prinzip: Publikationen werden schlicht gezählt, gemessen wird also die Häufigkeit der Beteiligung am wissenschaftlichen Kommunikationssystem. Unterschiede, sei es in Bezug auf die Qualität, den Umfang, den Publikationsort oder den ‚Impact‘ der Publikationen, werden nicht berücksichtigt. Aufgrund der hohen Standardisierung, dem Prinzip der Öffentlichkeit, dem hohen Maß an Transparenz und der direkten Anbindung an Mittelzuweisungen wird das RES als ‚starkes‘ Verfahren bezeichnet (Gläser/Laudel 2007: 128). Dabei lässt es die Schlichtheit der Indikatoren zu, den Wert einer Publikation in Drittmitteln direkt zu beziffern. Dieser lag im Jahr 2001 bei A\$ 3000 für einen Zeitschriftenartikel und bei A\$ 15.000 für ein Buch (Butler 2003: 40).

In der Literatur wird lebhaft über Anpassungseffekte diskutiert, mit denen sich Wissenschaftler und Universitäten auf diese Form der Forschungsevaluation einstellen. In einen Zusammenhang mit dem RES wird zwar eine Vielzahl von Phänomenen gebracht, empirisch belegt sind jedoch nur wenige Effekte. Grund dafür ist, dass Umfang und Art der Publikationsaktivitäten von einer Vielzahl von Faktoren abhängen und daher die Zurechnung der Veränderung des Publikationsaufkommens auf eine bestimmte Ursache naturgemäß schwierig ist. Interessant ist in diesem Zusammenhang eine Untersuchung von Linda Butler (2003). Sie ermittelt einen deutlichen Anstieg von Publikationen australischer Universitäten im Science Citation Index für sämtliche untersuchte Disziplinen. Für die Zurechnung des Effekts auf das RES sprechen dabei zwei Umstände. Erstens fällt der Anstieg des Publikationsaufkommens zeitlich zusammen mit der Einführung der Forschungsevaluation. Zweitens zeigt sich dieser Anstieg nur im universitären Bereich, in dem das System der leistungsbezogenen Mittelvergabe implementiert ist. In anderen Sektoren des australischen Wissenschaftssystems (wie z. B. im Bereich der außeruniversitären Forschung) findet sich das Phänomen nicht. Das oben beschriebene Wachstum lässt sich nicht bei sämtlichen

Journalen des SCIE beobachten. Weitgehend ausgenommen von diesem Trend sind prestigeträchtige Journale, die zwar hoch zitiert sind, aber auch eine hohe Ablehnungsquote aufweisen. Butler erklärt diese Ausnahme mit Verweis auf die Evaluationskriterien des RES. Das Publikationsverhalten der Wissenschaftler passt sich derart an, dass es auf eine Erhöhung der Menge von im Rahmen der Leistungsmessung gezählten SCI-Publikationen zielt, nicht aber auf die Erhöhung des Anteils an Publikationen in den Top-Journalen, da das Prestige der Zeitschriften keine Berücksichtigung findet: „With no attempt made to differentiate between the quality, visibility or impact of different journals when funding is allocated, there is little incentive to strive for publication in a prestigious journal.“ (Butler 2003: 41).

Einige Beobachter dieses Phänomens vermuten, dass die gestiegene Anzahl an Publikationen nicht als Indiz einer Verbesserung der Forschungsleistung zu interpretieren ist, sondern vielmehr ihre Ursache in einer veränderten Publikationsstrategie hat: „There is some evidence that the use of total numbers of publication as a performance measure may have led to ‚publishing inflation‘ – i.e., maximizing the numbers of articles produced by repetition, lowering quality standards, or the ‚salami slicing‘ of research into ‚least publishable units.‘“ (Geuna/Martin 2003: 283). Die Frage, weswegen sich die einzelnen Wissenschaftler an diesem Prozess beteiligen, also auf Anreize reagieren, die zunächst auf der Ebene von Forschungsorganisationen wirken, lässt sich damit erklären, dass ein Teil der Universitäten die interne Vergabe von Mitteln an die selben Maßzahlen bindet, nach denen die Leistungsmessung von außen stattfindet. Das äußere System der Leistungsmessung wird gewissermaßen innerhalb der Organisation wiederholt (Gläser/Laudel 2007: 138).

In den letzten Jahren ist Bewegung in die konzeptionelle Entwicklung des australischen Evaluationssystems gekommen. Aufgrund der Kritik am System, die sich nicht zuletzt an der Schlichtheit der ihm zugrunde liegenden Indikatoren entzündet, hat die australische Regierung im Jahr 2004 eine Expertenkommission eingesetzt. Ziel war es, ein Verfahren zu entwickeln, das auch qualitative Komponenten berücksichtigt (Butler 2007: 566). Qualitativ meint dabei, dass der Forschungoutput nicht nur durch eine einfache Zählung der Publikationen gemessen wird, sondern diese Messungen ergänzt werden durch Expertenurteile sowie quantitative Indikatoren, die das Gewicht oder die Bedeutung einer Publikation messen sollen. Für die Gruppe der Disziplinen ‚physical and biological sciences‘ wurde beispielsweise vorgeschlagen, neben der Anzahl hoch zitierter Papiere die Anzahl an Papieren in Spitzenjournalen, die Anzahl der Patente, die Erfolgsquote bei der Beantragung von Forschungsmitteln, die Anzahl an Einladungen zur Mitwirkung in Expertenkommissionen, die Höhe der von der Indust-

rie eingeworbenen Forschungsmittel sowie den Mittelumfang der im Auftrag der Industrie durchgeführten Forschungsprojekte zu erfassen (DEST 2005: 21).

Dieser Vorschlag wurde trotz weit fortgeschrittener konzeptioneller Arbeiten nie umgesetzt, sondern scheiterte an einem Regierungswechsel. Auf den Weg gebracht wird nun von der neuen Regierung ein System, das unter dem Namen ‚Excellence in Research for Australia‘ (ERA) das im RES verkörperte Prinzip einer disziplinenübergreifend einheitlichen Leistungsmessung und -bewertung aufgibt. Stattdessen werden je nach Disziplin und disziplinspezifischer Ausprägung des wissenschaftlichen Kommunikationssystems quantitative Maßzahlen mit der Beurteilung des Forschungsoutputs durch Experten kombiniert. Neben gerankten Publikationsmedien (Journale, Verlage, Konferenzen) wird Leistung – je nach Disziplin – durch Zitationsanalysen, an der Summe der eingeworbenen Forschungsmittel, der akademischen Ämter sowie der Patente und anderen intellektuellen Eigentumsrechte gemessen (ARC 2008a: 2f.).

Das Evaluationsverfahren verändert sich also in mehrerlei Hinsicht. Erstens vergrößert sich das Set an Indikatoren, die zur Leistungsmessung herangezogen werden. Zweitens werden Anstrengungen unternommen, die Qualität des Forschungsoutputs zu bestimmen. Wo immer es möglich ist, findet dies auf der Basis quantitativer Maßzahlen, wie beispielsweise Zitationsanalysen, statt. Drittens orientiert sich das Verfahren von seinem Gesamtansatz her immer noch an externen und nach Möglichkeit quantifizierenden Messungen des Forschungsinput und des Forschungsoutputs. Allerdings soll diesem Ansatz nur in Disziplinen gefolgt werden, in denen sich der jeweilige Indikator auch bewährt hat (ARC 2008b: 1). Wo dies nicht der Fall ist, soll ersatzweise eine qualitative Leistungsbewertung durch Mitglieder der betreffenden Disziplin – also *peers* – vorgenommen werden.

3.2 Die Research Assessment Exercise (RAE) und der Research Excellence Framework (REF) in Großbritannien

Auf den ersten Blick zeigen sich deutliche Ähnlichkeiten zwischen den Systemen der öffentlichen Forschungsförderung in Großbritannien und Australien. Auch das britische System ist dual angelegt. Die eine Säule des Systems wird von der projektbezogenen Forschungsförderung gebildet, die von den Grundzügen her mit dem oben beschriebenen System der australischen Projektförderung zu vergleichen ist. Die zweite Säule des Systems ist die leistungsorientierte Vergabe von öffentlichen Mitteln durch sogenannte „Funding Councils“. Diese Art der Förderung zielt auf die Herstellung und Aufrechterhaltung einer Forschungsinfrastruktur (Hare 2003: 46).

In Großbritannien begannen Versuche mit diesem Instrument der Forschungsevaluation bereits zu Beginn der 1980er Jahre, ein Einsatz des Evaluationssystems fand erstmalig im Jahr 1986 statt. Seitdem wurden 1992, 1996, 2001 und 2008 Research Assessment Exercises durchgeführt. Der Beschreibung der RAE des Jahres 2001 ist das Ziel präambelartig vorangestellt, die Qualitätsmessung von Forschung direkt mit wissenschaftspolitischen Entscheidungen zu verknüpfen: „The main purpose of the Research Assessment Exercise (RAE) is to enable the higher education funding bodies to distribute public funds for research selectively on the basis of quality.“⁹ Die Evaluation der Forschungsergebnisse beeinflusste die Mittelzuweisung an die Institute in den folgenden Jahren.

Das methodische Vorgehen der RAE fokussiert stark auf die Berücksichtigung von Qualität und weist damit deutliche Unterschiede zum ursprünglichen australischen System auf. Die Forschungsaktivitäten werden in Bereiche oder Forschungsfelder, sogenannte Units of Analysis (UoA), zusammengefasst. Für jede UoA wird ein hochkarätiges Panel erfahrener Wissenschaftler aus den entsprechenden Forschungsbereichen zusammengestellt (Hare 2003: 47, Elkin 2002: 204), das für die Qualitätsbewertung zuständig ist. Die Bewertung der Forschungsleistung basiert auf variierenden Parametern. Dazu gehörten bisher (a) das Personal der betreffenden Forschungseinrichtung, (b) Publikationen, (c) Studierenden- und Promovierendenzahlen, (d) Umfang der Forschungsmittel und (e) Informationen zum Forschungsfeld sowie zur strategischen Ausrichtung der Einrichtungen. Das Panel hat nun die Qualität für sämtliche Einreichungen auf einer mehrstufigen Skala zu beurteilen. Bei der Eingruppierung auf der Skala fungieren die erhobenen Parameter als Informationsgrundlage, führen aber nicht direkt zu einem bestimmten Platz. Die Letztentscheidung über die Eingruppierung liegt bei den Experten. An dieser Skala orientiert sich die Vergabe von Forschungsmitteln, wobei ein Großteil der Mittel auf die beste Stufe entfällt, die mittleren Stufen wenig Mittel erhalten und die Einrichtungen der schlechtesten Stufe leer ausgehen.

Sowohl die Informationsgrundlage als auch die Skala haben sich in den Durchläufen der RAE mehrfach verändert. Während die RAE 1992 den gesamten Publikationsoutput der Wissenschaftler der Forschungseinrichtungen zu erfassen suchte und in diesem Zusammenhang mindestens auch sporadisch bibliometrische Indikatoren im Hintergrund zum Einsatz kamen (Bence/Oppenheim 2004: 56), ist in den RAE von 1996 und 2001 eine Hinwendung zu qualitativen Kriterien zu beobachten (Geuna/Martin 2003: 283). Durch die Beschränkung der Zahl von vier eingereichten Publikationen pro Wissenschaftler soll es den Mit-

⁹ Siehe die Beschreibung der RAE 2001 auf der Website: <http://www.rae.ac.uk/2001/AboutUs/> (Stand 10/2011). Eine nahezu gleich lautende Passage findet sich auch in der Begründung der letzten RAE (Siehe RAE 2005: 5).

gliedern der Panels ermöglicht werden, sich durch die Lektüre der Texte ein Bild von der Leistung des betreffenden Wissenschaftlers zu machen. Der dabei entstehende Aufwand wird von manchen Beobachtern zum Teil kritisch hervorgehoben (Bence/Oppenheimer 2004: 65).

Für das Verständnis der RAE und der Anpassungsreaktionen in der Wissenschaft ist es von Bedeutung, dass sich die Evaluation auf die Einheit ‚Forschungsinstitut‘ bezieht und die Einrichtungen trotz detaillierter Vorgaben von Seiten der RAE große Spielräume haben, zu entscheiden, welche Organisationsmitglieder im Rahmen der Exercise als ‚Forscher‘ gelten. Dieser Spielraum wird von den Einrichtungen strategisch genutzt, um mit Blick auf die Leistungsmessung und die zu vergebenden Mittel möglichst gut abzuschneiden. Ein Beobachter führt hierzu aus: „So the department is invited to play games: does it put in more than thirty percent or less than thirty percent? If thirty percent will earn a 5*, that is a good multiplier but for few staff; if it put in sixty percent and only gets grade 3a that is a poor multiplier but for more staff.” (Brook 2003: 63).

Solche Optimierungsstrategien seitens der Institute hatten auf der Ebene des Systems weitgehende Effekte: Bei der Durchführung der RAE 2001 wurde bemerkt, dass nahezu alle Forschungseinrichtungen die Logik der Leistungsmessung verstanden und ihre Strategien entsprechend angepasst hatten. Zum Zeitpunkt der Meldung des Personals war absehbar, dass die Institute, die in die drei schlechtesten Kategorien fallen würden, keine Mittel bekämen. Die Einrichtungen reagierten darauf, indem sie nur die produktivsten Wissenschaftler benannten. In der Summe hatte das den Effekt, dass die Anzahl der Forscher und die Menge der Einreichungen zwar schrumpften, das Niveau der Forschung gemessen an der Skala der RAE sich insgesamt aber verbesserte:

„An unanticipated outcome in the 2001 Research Assessment Exercise was that everybody had got better. Indeed, whereas in 1996 the number for people in categories 4,5, and 5*, had been thirty-seven percent. A dire consequence was that, because the academic world had been so motivated by the Exercise and had so dramatically improved according to the criteria of the exercise, there was not enough money to give the anticipated reward... But in the end they came to the ingenious solution. All of the results should be mathematically adjusted and further modified. What this meant was if you were a 5 department you stayed just the same as you were before, but if you were a 3a department you lost about sixty percent of the money that you had been led to expect.” (Brook 2003: 63f.).

Neben der Optimierung der Ergebnisse der Leistungsmessung griffen die Forschungseinrichtungen aber auch zu weitergehenden Strategien, um ihre Chancen auf leistungsbezogene Mittel zu erhöhen. An erster Stelle zu nennen ist hier die Rekrutierung von Top-Wissenschaftlern zu dem Zweck, die Kategorie der For-

schungseinrichtung auf der Skala des RAE zu verbessern (Stephan 2008: 314, Banatvala/Bell/Symonds 2005: 459). Solche Strategien mögen zwar aus der Sicht der rekrutierenden Einrichtung rational sein, bringen aber aus der Perspektive der Wissenschaftspolitik keine Vorteile. Im Gegenteil: Sie sind eine unerwünschte Konsequenz der Leistungsmessung, da sie zu einer unproduktiven Verwendung von Forschungsmitteln führen. Daher fühlte sich wiederum die Politik aufgefordert, das Instrument der Leistungsmessung anzupassen, um solche Effekte zu unterbinden: "The resulting academic transfer market was considered to be a problematic feature of RAE 1996, so for RAE 2001, measures were taken to limit its effects notably by setting a cut of date for staff movements one year in advance of the deadline for eligible research." (Hare 2003: 53).

Die Research Assessment Exercise 2008 war die letzte, die unter diesem Namen durchgeführt wurde. An ihre Stelle tritt in Zukunft das Research Excellence Framework (REF), das zwar mit den RAE die grundsätzliche Zielsetzung teilt, sich aber in den folgenden drei zentralen Punkten unterscheidet. Erstens bleibt das Konzept der Beurteilung von Forschungsleistungen durch *peers* zwar in Kraft, die Anzahl der UoA wird aber stark geschrumpft. Vorgesehen sind nunmehr 30 Bereiche im Vergleich zu 67 UoA in der RAE 2008 (HEFCE 2009a: 2). Zweitens wird neben dem Forschungoutput und der strategischen Ausrichtung der zu bewertenden Forschungseinrichtungen nun auch die gesellschaftliche Relevanz der Forschung mit berücksichtigt. Und drittens werden erstmalig offiziell auch bibliometrische Indikatoren zur Messung von Forschungsleistungen eingesetzt. Dieser Einsatz ist freiwillig in dem Sinne, dass die verschiedenen Panels darüber entscheiden können, ob sie Zitationsanalysen nutzen wollen. Das Higher Education Funding Council für England räumt ein, dass Zitationsanalysen nicht für jedes Feld gleichermaßen geeignet sind. Die Stoßrichtung ist aber klar: Der Einsatz von Zitationsanalysen soll das Verfahren verschlanken, stärker standardisieren und den Aufwand letztlich reduzieren (HEFCE 2009a: 4).

3.3 Zwischenresümee

Die beiden Beispiele RES und der RAE sollen dazu dienen, ein Bild von der indikatorenbasierten Wissenschaftspolitik zu gewinnen. Als Zwischenresümee möchte ich die folgenden drei Punkte festhalten:

- *Bedeutung von zitationsbasierten Indikatoren:* Für die beiden Forschungsevaluationen lässt sich feststellen, dass Zitationsanalysen zwar eingesetzt werden, ihre Bedeutung aber gemessen an den eingangs beschriebenen wissenschaftspolitischen Visionen der 1970er Jahre vergleichsweise bescheiden

ist. Neben ihnen finden andere Wissenschaftsindikatoren wie zum Beispiel Publikationszählungen und die Bilanzierung von Drittmitteln Verwendung. Expertenurteile spielen zumindest im britischen System eine Rolle, im australischen Modell waren und sind sie von eher untergeordneter Bedeutung. Dennoch lässt sich beobachten, dass die Relevanz von Indikatoren, die auf Zitations-Indexen basieren, in der Tendenz steigt, und zwar in beiden Systemen. Die britischen und australischen Systeme markieren vom Ausgangspunkt ihrer Entwicklung her in gewisser Weise zwei Extrempole: Während das australische System neben anderen Indikatoren den Forschungsoutput durch eine sehr einfache, gewichtete Zählung von Publikationen zu ermitteln suchte, begann das britische System mit einer sehr aufwändigen Begutachtung durch Fachkollegen. Mittlerweile nähern sich beide Systeme einem ähnlichen Modell an, wenngleich auch aus verschiedenen Richtungen. Sie entwickeln sich in Richtung einer Leistungsmessung und -bewertung, die eine quantitative Messung des Forschungsoutputs durch ein Set von Indikatoren unter Berücksichtigung von Zitationsanalysen zur Bestimmung der Qualität von Publikationen beinhaltet. Das australische System überwindet damit die Schlichtheit einer bloßen Zählung von Publikationen, in Großbritannien scheint es dagegen vorrangiges Ziel zu sein, für die Experten die Lasten aufwändiger Urteile zu reduzieren.

- *Rückwirkungen von Systemen der Leistungsmessung:* Die empirischen Beispiele zeigen, dass die Leistungsmessung zu Ketten von Reaktionen und Gegenreaktionen führen. Auch wenn wegen der oben angesprochenen Zurechnungsprobleme kein vollständiges Bild in Bezug auf die Effekte von Leistungsevaluationen in der Wissenschaft vorliegt, finden sich in der Literatur einige Indizien dafür, dass ein Zusammenhang zwischen der Art der Leistungsmessung und der Art der Reaktionen existiert. Wegen der vergleichsweise hohen Komplexität des Instruments der Leistungsmessung reagieren englische Universitäten offenbar eher mit einer allgemeinen, unspezifischen Strategie: Sie konkurrieren vornehmlich um Spitzenwissenschaftler, die mit Blick auf verschiedene Indikatoren gut abschneiden. Die Reaktion australischer Universitäten besteht unter anderem darin, Steuerungsimpulse an die individuellen Wissenschaftler weiterzugeben. Diese verändern ihre Publikationsstrategie dahingehend, dass sie ihre Publikationen bevorzugt in einem bestimmten Segment der SCIE-Journale unterbringen.
- *Anpassungen auf Seiten der Wissenschaftspolitik:* Eine dritte Beobachtung bezieht sich auf die Reaktionen auf der Seite der Wissenschaftspolitik. Im Fall beider Evaluationssysteme sieht sich die Politik gezwungen, auf nicht-intendierte Effekte der Leistungsmessung zu reagieren. Diese Reaktionen reichen von der Veränderung des Messinstruments (Australien, UK) über

die Modifikation der Regeln, wie dieses Instrument zu handhaben ist (z.B. Einführung eines Stichtags, zu dem die in der Evaluation berücksichtigten Wissenschaftler in der Forschungseinrichtung beschäftigt sein müssen, UK) bis hin zu der Anpassung der Entscheidungsregeln zur Vergabe von Forschungsmitteln (wiederum UK). Diese Reaktionen zeigen an, dass eine indikatorbasierte wissenschaftspolitische Steuerung kein statisches Unterfangen ist, sondern einer starken Dynamik unterliegt. Und diese betrifft sowohl die Politik als auch die Wissenschaft.

4 Das Verhältnis von Wissenschaft und Politik in Zeiten indikatorbasierter Steuerung

Die Ergebnisse des Zwischenresümées führen zu der weitergehenden und auf der Ebene soziologischer Theorie liegenden Frage, wie das Verhältnis von Wissenschaft und Politik beschaffen ist, in dem Evaluationsinstrumente dazwischen geschaltet sind. Der Diskussion um indikatorbasierte Mittelvergabe liegt häufig ein einfaches Kausalmodell zugrunde, auf dessen Ursachenseite eine leistungsbewertende Politik und auf dessen Folgen- oder Wirkungsseite die Wissenschaft, in der sich Strategien und Anpassungsvorgänge vollziehen, verortet wird. Gegen eine solche Perspektive sollen hier keine prinzipiellen Einwände vorgebracht werden. Vielmehr ist diese Perspektivenbeschränkung häufig nachgerade notwendig, beispielsweise um einen empirischen Nachweis des Zusammenhangs zwischen Leistungsmessung und Wandlungsprozessen innerhalb der Wissenschaft zu erbringen und Folgen eindeutig zurechnen zu können.¹⁰ Die Absicht dieses letzten Argumentationsschritts ist aber eine andere. Es geht hier erstrangig um die Entwicklung eines übergreifenden theoretischen Rahmens, der den gesamten Bereich des Phänomens abdeckt. Daher wird im Folgenden eine Perspektive entwickelt, die den wechselseitigen Rückwirkungen in Wissenschaft und Politik Rechnung trägt.

Es bietet sich an, einen differenzierungstheoretischen Ausgangspunkt zu wählen und davon auszugehen, dass es sich bei Wissenschaft und Politik um gesellschaftliche Teilsysteme handelt, die gemäß verschiedener Logiken operieren. Diese von Weber bis in das neuere Verständnis der Systemtheorie hineinragende Figur gesellschaftlicher Differenzierung ist zu ergänzen durch die Unterscheidung verschiedener Typen von Systemen. Aufgrund der zentralen Rolle von strategischen Anpassungsprozessen, mit denen Organisationen (oder Teile von Organisationen) auf Evaluationen reagieren, liegt der Rückgriff auf eine Theorie

¹⁰ Zu den Schwierigkeiten einer solchen Zurechnung siehe ausführlicher Gläser/Laudel (2007) und Butler (2007).

nach, die über einen starken Akteursbegriff verfügt und die für solche Reaktionsmuster eine angemessenen begrifflichen Rahmen an zentraler Stelle bereithält. Aus diesem Grund wird hier einer differenzierungstheoretisch informierten Lesart des akteurszentrierten Institutionalismus der Vorzug gegeben gegenüber der ohne Frage sehr weit ausgearbeiteten und begrifflich konsistenten Systemtheorie Luhmann'scher Provenienz.

In Anknüpfung an Renate Mayntz (1988) und andere lassen sich drei Strukturdimensionen unterscheiden. Die oberste Ebene bildet die Differenzierung der Gesellschaft in funktionale Teilsysteme, die hier verstanden werden als „hochgradig generalisierte(r) sinnhafte(r) Orientierungen, die den Akteuren als allgemein verbreitete situationsdefinierende Fiktion gegenwärtig sind“ (Schimank 2007: 220) Von einer Fiktion ist zu sprechen, weil erhebliche Diskrepanzen zwischen der generalisierten sinnhaften Orientierung und konkreten sozialen Situationen existieren und die Akteure sich diesen Diskrepanzen auch durchaus bewusst sind, sich aber dennoch an den simplifizierenden Abstraktionen orientieren (Schimank 1988: 632f.). Auch wenn mit den Teilsystemen ‚Wirtschaft‘, ‚Recht‘, ‚Wissenschaft‘ und ‚Politik‘ ähnliche gesellschaftliche Makrostrukturen wie in der Systemtheorie Luhmanns adressiert werden, unterscheidet sich die Konstitutionstheorie doch in wesentlichen Punkten. Während Luhmann für die erfolgreiche Ausdifferenzierung gesellschaftlicher Teilsysteme ein Bestandsproblem des übergeordneten Sozialsystems ‚Gesellschaft‘ voraussetzt (Luhmann 1997: 747f.), sind die Annahmen des akteurszentrierten Institutionalismus deutlich schwächer. Folgt man Mayntz in diesem Punkt, handelt es sich hier um Formen sinnhafter Spezialisierung, die sich auf solche Bezugsprobleme beziehen können, es jedoch nicht zwangsläufig müssen. Zudem wird nicht grundsätzlich von einer autopoietischen Operationsweise ausgegangen, die den Funktionssystemen ein hohes Maß an Autonomie und eigenlogischem Prozessieren unterstellt. Stattdessen wird – letztlich auch mit Blick auf die empirische Verwendbarkeit der Theorie – mit einer mehr oder minder starken Spezialisierung und einem mehr oder minder hohen Grad an Ausdifferenzierung funktionaler Teilsysteme gerechnet. Teilsysteme haben dabei orientierende Wirkung auf das Handeln von Akteuren. Schimank zufolge führt die Zugehörigkeit eines Akteurs zu einem Teilsystem dazu, dass er weiß, welcher Richtung des *Wollens* er sich zuwenden kann (Schimank 2007: 220).

Fragt man im Rahmen dieser Theorie nach der Binnenstruktur gesellschaftlicher Teilsysteme, wird man auf Akteurskonstellationen verwiesen (Mayntz 1988: 24). Sie konstituieren sich aus dem Zusammenspiel mehrerer Akteure, die gewahr sind, dass sie ihre Intentionen nicht monologisch verfolgen können, sondern auf die Koordination ihres Handelns mit anderen Akteuren angewiesen sind (Schimank 2007: 221). Andere Akteure einer Konstellation wirken dabei nicht

nur restriktiv auf das Handeln, sondern häufig auch handlungsermöglichend. Sie können unterstützend wirken oder durch die Art ihrer Handlungsorientierung Hindernis oder Widerstand darstellen. In Bezug auf Akteurskonstellationen rückt daher das *Können* von Akteuren in den Fokus der Aufmerksamkeit. Die dritte Strukturdimension ist die institutionelle Ordnung oder sind die normativen Orientierungen, denen sich ein Akteur ausgesetzt sieht. Diese Erwartungen (oder Vorgaben des *Sollens*) besitzen unterschiedliche Reichweiten und Grade der Formalisierung, und umfassen einfache Verhaltenserwartungen Mitgliedschaftserwartungen von Organisationen und formalisierte Rechtsnormen (ebd.: 221).

Mit diesen Überlegungen sind nun die wesentlichen theoretischen Voraussetzungen gewonnen, um den empirischen Fall der indikatorenbasierte Leistungsmessung reflektieren zu können. Zunächst zur Seite der Politik: Im Zuge der wohlfahrtsstaatlichen Ausweitung der Zuständigkeiten des Staats ist der Umfang des Staatshaushalts stark angestiegen. Dies gilt auch für die Aufwendungen für die Wissenschaft, deren (in jüngster Zeit etwas abnehmendes) Größenwachstum über sechs Dekaden hinweg enorme Kostensteigerungen verursacht hat. Dies hat die politischen Akteure in eine Situation geführt, die durch zum Teil widersprüchliche Legitimationsanforderungen gekennzeichnet ist: Ein hohes und in der Tendenz steigendes Erwartungsniveau in Bezug auf die Leistungserbringung findet seine Ergänzung in der Forderung nach sparsamem und effizientem Umgang mit öffentlichen Mitteln. In dieser Situation ist es für die zuständigen Ministerien attraktiv, Instrumente zu nutzen, die eine Mittelallokation nach der Maßgabe der höchsten Effizienz versprechen. Der Einsatz dieses Beobachtungsinstruments hat aber auch noch einen Zusatznutzen: Die Legitimierungslasten verschieben sich von den politischen Akteuren auf das Evaluationsinstrument und auf die durchführenden Organisationen. Weiter oben wurde bereits festgestellt, dass sich für diesen Zweck ein Instrument anbietet, das sich des Mediums der Zahlen bedient.

Die Situation der wissenschaftlichen Akteure lässt sich dagegen wie folgt charakterisieren: Aufgrund ihrer Abhängigkeit von vornehmlich öffentlichen Ressourcen sind sie gezwungen, sich an der Leistungsmessung zu beteiligen und gegebenenfalls an diesem Prozess durch Bereitstellung oder Überprüfung von Daten aktiv mitzuwirken. Interessant ist dabei zu sehen, dass eine Beobachtung hier nicht nur in Richtung der Wissenschaft stattfindet, sondern auch in umgekehrter Weise. Die offenbar auf die jeweiligen Leistungsmessvorgänge abgestimmten und jeweils spezifischen Reaktionen und Anpassungsprozesse der Wissenschaft zeigen an, dass die wissenschaftlichen Akteure ihrerseits die Politik in der Handhabung des entsprechenden Leistungsmessinstruments beobachten und sich als Messobjekte mit Blick auf die Kategorisierungen, die der Messung zugrunde liegt, optimieren. Die Art und Weise der Ausgestaltung des Eva-

lutionsinstruments scheint damit nicht nur bedeutsam zu sein für das Messergebnis selbst, sondern auch für die Art und Weise, wie die Wissenschaft auf die Beobachtung der Politik reagiert.

Die oben entwickelte Unterscheidung zwischen der Ebene der strategiefähigen Akteure einerseits und der Akteurskonstellationen andererseits erweist sich für den hier analysierten Gegenstand als fruchtbar, da es mit ihr gelingt, die intentionalen Strategien und transintentionalen Effekte auseinander zu halten. Für beide Akteursgruppen lässt sich das Entstehen von weiter reichenden transintentionalen Effekten vermuten. Dazu drei Beispiele:

- *Neutralisierung der Strategien sämtlicher Akteure:* In diesem Fall blockieren sich in der Tendenz die Strategien, so dass eine Situation entsteht, die so ähnlich vermutlich auch ohne die Anwendung von Strategien eingetreten wäre. Ein Beispiel für eine solche Situation war im Zuge der britischen RAE zu beobachten, als sich eine große Zahl wissenschaftlicher Einrichtungen darum bemühte, im Rahmen der RAE einen bestimmten Rang zu erreichen, ab dem der Zufluss von Forschungsmitteln zu erwarten war. Die Politik reagierte darauf mit einer Strategie der Anpassung der Regeln zur Vergabe von Forschungsmitteln und neutralisierte diese Bestrebungen.
- *Externalitäten in der Wissenschaft:* Hier führt das Zusammenspiel der Akteursstrategien zu nicht-intendierten Effekten auf übergeordneter Ebene. Zu denken ist dabei an die Entstehung eines hoch-kompetitiven Arbeitsmarktes für exzellente Wissenschaftler durch die Konkurrenz der Forschungsorganisationen um Spitzenpersonal. Weitere mögliche Effekte wären die Überlastung der Redaktionen wissenschaftlicher Fachzeitschriften durch gezielte Einreichung von Manuskripten in einem spezifischen Segment von Journalen oder auch die Aufblähung des wissenschaftlichen Kommunikationssystems insgesamt.
- *Politisierung der Evaluation:* Sie findet auf verschiedenen Wegen statt. Hervorhebenswert sind hier Beispiele von koordiniertem Handeln wissenschaftlicher Akteure, das darauf gerichtet ist, die sachliche Angemessenheit des Evaluationsinstruments in Zweifel zu ziehen. Exemplarisch sei eine gemeinsam unterzeichnete Erklärung von Herausgebern einer beträchtlichen Anzahl an Journalen aus dem Bereich der Wissenschaftsforschung genannt, die sich gegen den Aufbau eines European Reference Index for the Humanities (ERIH) wenden (Weingart 2009: 1-5). Damit ist jedoch nur ein erster Schritt zu einer Politisierung getan, deren Ausgang ungewiss ist.

Mit einer letzten – eher spekulativen – Überlegung soll dieser theoretische Einordnungsversuch abgeschlossen werden. Führt man sich noch einmal den Zu-

sammenhang vor Augen zwischen den wechselseitigen Beobachtungen von Akteuren aus Wissenschaft und Politik durch das Evaluationsinstrument hindurch, der Entwicklung von darauf bezogenen spezifischen Strategien sowie entstehenden transintentionalen Folgen, könnte man darauf kommen, dass die Gestaltung des Beobachtungsinstrumentes Einfluss auf die Eintrittswahrscheinlichkeit solcher transintentionaler Effekte hat. Denn: Insbesondere wenn Akteure mit ihren Strategien in gleichgerichteter Weise auf Leistungsmessungen reagieren, treten derartige Effekte ein. Damit ist aber vermutlich vor allem dann zu rechnen, wenn die Instrumente der Leistungsmessung und -bewertung sehr einfach gebaut sind und den Akteuren eindeutige Strategien nahe legen, mit denen das Messergebnis beeinflusst werden kann. Dies mag unter anderem auch der Grund sein, weswegen – trotz des immer wieder beklagten hohen Aufwands von Alternativen – auf den Einsatz von sehr simplen Instrumenten der Leistungsmessung mittlerweile überwiegend verzichtet wird.

5 Forschungsevaluationen und die Zukunft des Zitations-Index

In diesem Aufsatz wurde ein Bogen gespannt von der Entwicklung des bibliometrischen Instrumentariums hin zu dessen Einsatz im Rahmen von Systemen der leistungsbezogenen Forschungsförderung. Zwar stellen bibliometrische Indikatoren nur ein Mittel neben anderen dar, um Wissenschaft zu vermessen. Sie sind aber von herausgehobenem Interesse, da sie die intertextuellen Referenzen nutzen, um auf aggregierter Ebene zumindest ihrem Selbstverständnis nach Bedeutendes von Unbedeutendem zu unterscheiden, Anschlussfähiges von Marginalem zu trennen und letztlich Förderungswürdiges gegenüber nicht Förderungswürdigem hervorzuheben. Das hohe Maß an Reaktivität der Wissenschaft einerseits aber auch die dauerhafte Unruhe in Bezug auf die Veränderung des Beobachtungsinstrumentes andererseits sind zurückzuführen auf die ‚stakes‘, die auf dem Spiel stehen: Die politischen Akteure sehen sich widersprüchlichen Legitimationsanforderungen ausgesetzt, die wissenschaftlichen Akteure bangen um ihre Ressourcen. Von daher ist es eher unwahrscheinlich, dass sich etwas an der grundsätzlichen Konstellation – wechselseitige Beobachtung wissenschaftlicher und politischer Akteure durch die Brille des jeweiligen Evaluationsinstrumentes – ändert.

Damit soll allerdings keineswegs der Eindruck geweckt werden, dass diese Beobachtungskonstellation statisch ist. Erhebliche Veränderungsimpulse ergeben sich aus der Entwicklung von Messinstrumenten und gehen auch von der Digitalisierung der wissenschaftlichen Kommunikation aus. Anders formuliert: Nicht die Beobachtungskonstellation selbst verändert sich, sondern die in ihr zum Ein-

satz kommenden Instrumente. Auch wenn es zum heutigen Zeitpunkt nicht vollständig absehbar ist, wie zukünftige Instrumente der Leistungsmessung aussehen werden, lässt sich an den derzeit noch in der Entwicklung befindlichen Messmethoden ablesen, welche Gestalt sie annehmen könnten. Dabei erwarte ich, dass sich die Digitalisierung der wissenschaftlichen Kommunikation in vielfältiger Hinsicht als wirkmächtig erweisen wird. Mit Blick auf das Feld der Zitationsanalyse lässt sich von einer Ausweitung der Beobachtungsmöglichkeiten sprechen. Daneben eröffnet das Primat der digitalen Publikation und Distribution wissenschaftlicher Texte Einblicke in Rezeptionsvorgänge. Auf den digitalen Vertriebswegen hinterlassen Rezipienten Datenspuren und diese könnten auch im Rahmen von Evaluationen genutzt werden. Ohne den Anspruch auf Vollständigkeit erheben zu wollen, werde ich mögliche Entwicklungspfade in drei Feldern skizzieren.

- *Zitations-Index*: Zu Zeiten der gedruckten Literatur gestaltete sich der Aufbau eines Zitations-Index mühsam. Die Zitationen mussten händisch aus den Publikationen herausgesucht und in eine Datenbank verbracht werden. Mit der Digitalisierung der wissenschaftlichen Kommunikation liegen Publikationen nun in elektronischer Form vor und die Zitationen lassen sich weitgehend automatisch extrahieren.¹¹ Durch die Reduzierung des Aufwands für die Bewältigung dieses Arbeitsschritts sind neben dem Web of Science weitere Zitationsdatenbanken entstanden (z.B. Scopus¹² und Citebase¹³), neue Dienste, die Zitationsdaten liefern (z.B. Google Scholar) entwickelt und Zitationsdaten in Fachdatenbanken integriert worden (z.B. MathSciNet¹⁴ oder SAO/NASA ADS¹⁵). Zwar finden sich Hinweise darauf, dass diese Alternativangebote bislang noch nicht an die Qualität des Web of Science heranreichen¹⁶, die Beispiele zeigen aber, dass die Verfügbarkeit von Zitationsdaten zunimmt. Eine zweite wichtige Tendenz im Feld der Zitationsanalysen ist die Entwicklung automatischer Evaluationstools, mit denen die Leistungen von Wissenschaftlern oder wissenschaftlichen Einrich-

¹¹ Siehe zur Beschreibung eines solchen Verfahrens exemplarisch Brody (2003) für die Zitationsdatenbank Citebase.

¹² <http://www.scopus.com/home.url> (Stand 10/2011).

¹³ <http://citebase.org> (Stand 10/2011).

¹⁴ Siehe hierzu exemplarisch MathSciNet, eine der beiden großen Fachdatenbanken der Mathematik. Diese hält neben Abstracts und den Metadaten von Publikationen auch für 450 Journale die Zitationen bereit. (Siehe: <http://www.ams.org/mathscinet/help/about.html>, Stand 10/2011).

¹⁵ Siehe für die Fachdatenbank MathSciNet <http://www.ams.org/mathscinet/help/about.html> und für SAO/NASA Astrophysics Data System http://doc.adsabs.harvard.edu/help_pages/citations.html (Stand 10/2011).

¹⁶ Vgl. hierzu Jasco 2005.

tungen verglichen werden können. Mit SciVal¹⁷ (Elsevier) und InCite¹⁸ (Thomson/Reuters) haben beide Anbieter der großen Zitationsdatenbanken solche Tools im Angebot. Die Verfügbarkeit von Zitationsdaten und automatischen Evaluationstools spricht dafür, dass die Rolle von Zitationsanalysen bei der Beobachtung der Wissenschaft künftig eher zunehmen wird. Auch wenn die Qualität von Analysen vermittelt automatischer Evaluationstools eher zweifelhaft ist, scheint es wahrscheinlich zu sein, dass Zitationsanalysen in Zukunft nicht mehr nur im Rahmen größer angelegter Evaluationsaktivitäten eingesetzt werden, sondern auch ad hoc, beispielsweise im Rahmen von Berufungsverfahren oder bei der Beantragung von Drittmitteln.

- *Usage-based Metrics*: Während der erste Entwicklungspfad die mit dem Zitations-Index geschaffenen Beobachtungsmöglichkeiten von ihrer Form her reproduziert, gehen so genannte Usage-based Metrics von anderen Elementen – nämlich Clicks oder Downloads – aus. Im Zuge der Rezeption eines wissenschaftlichen Textes findet eine Zitation erst zu einem recht späten Zeitpunkt statt, dem üblicherweise ein mehrschrittiger Prozess vorausgeht. Dieser beinhaltet die Recherche nach Literatur, die Sichtung der Rechercheergebnisse, die Beschaffung des betreffenden Textes und schließlich dessen Lektüre. Erst auf dieser Grundlage findet bei der Anfertigung eines weiteren Textes möglicherweise eine Zitation der ersten Publikation statt. Die bei der elektronischen Beschaffung von Publikationen anfallenden Daten können genutzt werden, um Indikatoren zu entwickeln, die etwas über die nachgefragte Publikation, deren Autor oder den Publikationsort aussagen. Dies geschieht üblicherweise durch eine Auswertung der Downloadaktivitäten aus Publikationsdatenbanken oder Online-Repositoryen. Brody, Harnad und Carr (2006) nutzen beispielsweise die Downloadraten von Volltexten, um einen ‚Usage Impact‘ zu ermitteln und zu prüfen, inwieweit dieser geeignet ist, den späteren ‚Citation Impact‘ vorauszusagen. Der Vorteil einer solchen Maßzahl liegt auf der Hand: Sie kann relativ schnell ermittelt werden während der ‚Citation Impact‘, gerade im Fall von Fächern mit langsamer Rezeption und langen Rezeptionsdauern, erst nach einigen Jahren zuverlässig ermittelt werden kann. Eine Alternative zum bereits weiter oben erwähnten ‚Journal Impact Factor‘ entwickeln Bollen und van de Sompel (2008) auf der Grundlage von Nutzerdaten. Diese basiert ebenfalls auf Downloadraten von Volltexten und setzt – analog zum JIF – die Anzahl an Publikationen eines Journals in einem bestimmten Zeitraum in Beziehung zur Anzahl an Downloads dieser Publikationen. Bollen und van de Sompel machen darauf

¹⁷ <http://www.info.scival.com/> (Stand 10/2011).

¹⁸ <http://researchanalytics.thomsonreuters.com/incites/> (Stand 10/2011).

aufmerksam, dass der Usage Impact Factor eine eigene Dimension von ‚Impact‘ messe. Während der JIF auf einer Datengrundlage ermittelt wird, die durch Rezeptionshandlungen ausschließlich von Wissenschaftlern zustande kommt (eben Zitationen), bezieht der Usage Impact Factor auch die Rezeption durch andere Akteure mit ein. Usage Impact bezeichnet daher eine eigenständige Wirkungsdimension. Eine Umfrage zeigt, dass bei Wissenschaftlern, Bibliotheken und Verlagen eine große Aufgeschlossenheit gegenüber Indikatoren besteht, die auf Nutzungsdaten basieren (Shepherd 2007). Ob die Usage-based Metrics künftig im Rahmen von Evaluationsprozessen genutzt werden, scheint mir auch davon abhängig zu sein, ob es gelingt, die Manipulierbarkeit von Nutzerdaten weitgehend auszuschließen.¹⁹

- *Bookmark based metrics*: Eine weitere Datenquelle, die Einblicke in Rezeptionsvorgänge gibt, sind so genannte Online Reference Manager, also Literaturverwaltungsdienste wie CiteULike²⁰. Ebenso wie Publikationsdatenbanken und Online-Repositoryen lassen sie es mit vergleichsweise geringem Aufwand zu, Nutzerdaten als Nebenprodukt ihrer Verwendung zu sammeln: Während Nutzer mit ihrer Rezeption jeweils eigene Zwecke verfolgen, erzeugen sie große Mengen von Daten, die in aggregierter Form für die Beobachtungs- und Bewertungsprozesse geeignet sind. Im Fall von Literaturverwaltungsdiensten beziehen sich diese Daten auf Rezeptionsvorgänge wie die Speicherung von Dokumenten, die Annotierung oder Kategorisierung von Publikationen oder das Setzen von Lesezeichen (*bookmarks*). An letztgenannter Art der Rezeptionshandlung setzt Taraborelli an, um die Möglichkeit der Indikatorenentwicklung zu diskutieren. Unter anderem schlägt er einen „Popularity Index“ vor, der Auskunft darüber gibt, wie viele Personen ein Lesezeichen für einen bestimmten Text gesetzt haben (Taraborelli 2008: 6). Bei diesem Indikator sieht er das Potential, im Rahmen von Evaluationsaktivitäten zur Leistungsmessung beizutragen. Die Vorteile gegenüber anderen Indikatoren, die auf Nutzerdaten wie den eben angesprochenen Downloads basieren, sind in einer höheren Robustheit gegenüber Manipulationen und einer höheren Eindeutigkeit jedes einzelnen Vorgangs auf der Plattform zu sehen. Eine höhere Robustheit ist gegeben, weil die Plattform eine Benutzerregistrierung erfordert und die Manipulation der Daten durch automatische Programme erschwert wird. Von einer größeren Eindeutigkeit kann Taraborelli zufolge gesprochen werden, weil es möglich ist,

¹⁹ Downloadzahlen können beispielsweise von den Autoren erhöht werden, indem sie ihre Artikel mehrfach herunterladen. Mit softwaretechnischen Hilfsmitteln ließe sich ein solches „self-promotion gaming“ (Taraborelli 2008: 9) optimieren.

²⁰ <http://www.citeulike.org/> (Stand 10/2011).

das Setzen eines Lesezeichens einer individuellen Handlung eines Nutzers zuzurechnen (ebd.: 9).

Die hier skizzierten neuen Möglichkeiten einer Indikatorkonstruktion schließen sich keineswegs gegenseitig aus. So machte Stevan Harnad den Vorschlag, Publikationen, Metadaten und Nutzerdaten in einem komplexen Indikator zu vereinen und im Rahmen von Forschungsevaluationen zu nutzen. Neben anderen Maßzahlen könnten der Journal Impact Factor, Zitationen, Co-Zitationen, Downloads und Co-Downloads zu einem komplexen Index verknüpft werden (Harnad 2009: 154). Sollte ein solcher Index in Zukunft genutzt werden, darf man gespannt sein, mit welchen Strategien die auf solche Weise vermessenen Akteure reagieren. Aber auch wenn diese Informationen nicht zu Evaluationszwecken genutzt werden würden, hätte die dem Index zugrunde liegende Datenbank noch einen anderen Zweck, den Harnad anspricht: die Nutzung „for search and navigation by research users“ (Harnad 2009: 155). In diesem Sinne hätte sich ein Kreis geschlossen, denn diese Form der Nutzung entspricht der ursprünglichen Zwecksetzung des Zitations-Index: als ein Instrument der Recherche des Werdegangs von Ideen.

Literatur:

- ARC, Australian Research Council 2008a: ERA Indicator Descriptors. Als Online Dokument: http://www.arc.gov.au/pdf/ERA_Indicator_Descriptors.pdf (Stand 10/2011).
- ARC, Australian Research Council 2008b: ERA Indicator Principles. Als Online Dokument: http://www.arc.gov.au/pdf/ERA_Indicator_Principles.pdf (Stand 10/2011).
- Banatvala, Jangu; Bell, Peter; Symonds, Malcom 2005: The Research Assessment Exercise is bad for UK medicine. In: The Lancet, Vol. 365, Feb. 5, S. 458-460.
- Bence, Valerie; Oppenheimer, Charles 2004: The Influence of Peer Review on the Research Assessment Exercise. In: Journal of Information Science, Vol. 30(4), S. 347-368.
- Bollen, Johann; van de Sompel, Herbert 2008: Usage Impact Factors: The Effects of Sample Characteristics on Usage-Based Impact Metrics. In: Journal of the American Society for Information Science and Technology, Vol. 59(1), S. 136-149.
- Brook, Richard 2003: Research Survival in an Age of Evaluation. In: Max-Planck-Gesellschaft/Deutsche Forschungsgemeinschaft: Science Between Evaluation and Innovation: A Conference on Peer Review. München: Max-Planck-Gesellschaft. S. 61-66.
- Brody, Tim 2003: Citebase Search: Autonomous Citation Database for e-Print Archives. sinn03 Conference on Worldwide Coherend Workforce. Als Online-Dokument:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.9125&rep=rep1&type=pdf> (Stand 10/2011).

- Brody, Tim; Harnad, Stevan; Carr, Leslie 2006: Earlier Web Usage Statistics as Predictors of Later Citation impact. *Journal of the American Society for Information Science and Technology*, Vol 57(8), S. 1060-1072.
- Butler, Linda 2003: Modifying Publication Practices in Response to Funding Formulas. In: *Research Evaluation*, 12(1), S. 39-46.
- Butler, Linda 2007: Assessing University Research. A Plea for a Balanced Approach. In: *Science and Public Policy* Vol. 34(8), S. 565-574.
- Carter, H. E. 1973: Letter of Transmittal. In: *National Science Board, 1973: Science Indicators 1972. Report of the National Science Board 1973*. Washington: U.S. Government Printing Office.
- De Solla Price, Derek J. 1963: *Little Science, Big Science. Von der Studierstube zur Großforschung*. Frankfurt: Suhrkamp.
- DEST, Department of Education, Science and Training 2005: *Research Quality Framework: Assessing the Quality and Impact of Research in Australia*. Issue Paper. Endorsed for release for public consultation by the Expert Advisory Group for an RQF. Als Online-Dokument: <http://www.maths.anu.edu.au/other/ncms/rqf-research-quality-issues-paper.pdf> (Stand 04/2010).
- Elkin, Judith 2002: The UK Research Assessment Exercise. In: *Libri*, Vol. 52, S. 204-208.
- Garfield, Eugene 1955: Citation Indexes for Science. A New Dimension in Documentation Through Association of Ideas. In: *Science*, Vol. 122, S. 108-111.
- Garfield, Eugene 1963 (1969): Citation Indexes in Sociological and Historical Research. In: *Essays of an Information Scientist*, Vol. 1 p.43-46. Reprint from: *American Documentation* 14(4), S. 289-291.
- Garfield, Eugene 1979. *Citation Indexing – Its Theory and Application in Science, Technology and Humanities*. New York, et al: John Wiley & Sons.
- Garfield, Eugene 2006: The History and Meaning of the Journal Impact Factor. In: *JAMA*, Vol. 295(1), S. 90-93.
- Geuna, Aldo; Martin, Ben R. 2003: University Research Evaluation and Funding: An International Comparison. In: *Minerva*, 41(4), S. 277-304.
- Gläser, Jochen; Laudel, Grit: 2007: Evaluation without Evaluators: The Impact of Funding Formulae on Australian University Research. In: Whitley, Richard; Gläser, Jochen (Eds.): *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*. *Sociology of the Science Yearbook* 26. Springer: Dordrecht. S. 127-151.
- Hare, Paul G. 2003: The United Kingdom's Research Assessment Exercise: Impact on Institutions, Departments, Individuals. In: *Higher Education Management and Policy*, Vol. 15(2), S. 43-62.
- Harnad, Stevan 2009: Open Access Scientometrics and the UK Research Assessment Exercise. In: *Scientometrics*, Vol. 79(1), S. 147-156.
- HEFCE (Higher Education Funding Council for England) 2009a: *The Research Excellence Framework: A brief guide to the proposals*. Als Online-Dokument: <http://www.hefce.ac.uk/research/ref/resources/REFguide.pdf> (Stand 10/2011).

- HEFCE (Higher Education Funding Council for England) 2009b: Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework. Als Online-Dokument: http://www.hefce.ac.uk/pubs/hefce/2009/09_39/ (Stand 10/2011).
- Heintz, Bettina, 2007: Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven. In: Mennicken, Andrea; Vollmer, Hendrik (Hrsg.). Zahlenwerk. Kalkulation, Organisation und Gesellschaft. Wiesbaden: VS Verlag. S.65-86.
- Heintz, Bettina 2010: Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. In: Zeitschrift für Soziologie, Vol. 39(3), S. 162-181.
- Hirsch, Jorge E. 2005: An index to quantify an individual's scientific research output, Proceedings of the National Academy of Sciences of the United States of America, Vol. 102(46), S. 16569–16572.
- Jasco, Peter 2005: As we may search – Comparison of major features of the Web of Science, Scopus and Google Scholar citation based and citation-enhanced databases. In: Current Science, Vol. 89(9), S. 1537-1547.
- Lawrence, Peter A. 2003. The politics of publication. In: Nature, Vol. 20(20th March), S. 259-261.
- Luhmann, Niklas 1997: Die Gesellschaft der Gesellschaft. Band II. Frankfurt: Suhrkamp.
- Mayntz, Renate 1988: Funktionelle Teilsysteme in der Theorie sozialer Differenzierung. In: Mayntz, Renate; Rosewitz, Bernd; Schimank, Uwe; Stichweh, Rudolf (Hrsg.): Differenzierung und Verselbständigung. Zur Entwicklung gesellschaftlicher Teilsysteme. Frankfurt/New York: Campus.
- Pritchard, Alan 1969: Statistical Bibliography or Bibliometrics? In: Journal of Documentation, Vol. 25, S. 348-349.
- RAE 2001: Research Assessment Exercise 2001. Guidance to Panel Chair and Members: Criteria and Working Methods. Als Online-Dokument: www.rae.ac.uk/2001/PanGuide/guide/guide.rtf (Stand 10/2011).
- RAE 2005: RAE 2008. Guidance on Submissions. Als Online-Dokument: <http://www.rae.ac.uk/pubs/2005/03/rae0305.pdf> (Stand 10/2011).
- Schimank, Uwe 1988: Gesellschaftliche Teilsysteme als Akteursfiktionen. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Vol. 40(4), S. 619-639.
- Schimank, Uwe 2007: Theorien gesellschaftlicher Differenzierung. 3. Auflage. Wiesbaden: VS Verlag.
- Schmidmaier, Dieter, 1985: Zur Geschichte der Bibliometrie. In: Zentralblatt für Bibliothekswesen, Vol. 98(9), S. 404-406.
- Shepherd, Peter T. 2007: The Feasibility of Developing and Implementing Journal Usage Factors: a Research Project Sponsored by UKSG. Serials, Vol 20(2), S. 117-123.
- Stephan, Paula E. 2008: Science and the University: Challenges for Future Research. In: CESifo Economic Studies, Vol. 54(2), S. 313-324.
- Taraborelli, Dario 2008: Soft peer review: Social software and distributed scientific evaluation. In: Proceedings of the 8th International Conference on the Design of Cooperative Systems (COOP '08), Carry-Le-Rouet, May 20-22, 2008, S. 99-110.
- Weinberg, Alvin M. 1963: Criteria for Scientific Choices. In: Minerva, Vol. 1(3), S. 159-171.
- Weingart, Peter 2009: Editorial for Issue 47/1. In: Minerva, Vol. 47(1), S. 1-5.