

Das Forschungsrating des Wissenschaftsrats

Einige Erfahrungen und Befunde

and similar papers at core.ac.uk

provided by Publikationen der Deut

Den deutschen Soziologen und Soziologinnen wird nicht entgangen sein, dass sie vom Wissenschaftsrat im Hinblick auf die Qualität von Forschung, Nachwuchsförderung und Wissenstransfer ihrer Forschungseinrichtungen bewertet wurden. Vor zwei Jahren habe ich dieses Großunternehmen in der »Soziologie« (35. Jg., Heft 3) vorgestellt. Im April 2008 wurden die Ergebnisse von der Steuerungsgruppe des Wissenschaftsrats bekannt gemacht, und die Medien haben darüber berichtet. Sowohl der Ergebnisbericht der Steuerungsgruppe als auch der Abschlussbericht der zuständigen Bewertungsgruppe¹ (im Folgenden zitiert als AB) können bei der Geschäftsstelle des Wissenschaftsrats angefordert werden. Beide Texte sind auch im Internet einsehbar.² Ich kann mich an dieser Stelle deshalb auf die

1 Mitglieder der Bewertungsgruppe waren Andreas Diekmann (Zürich), Birgit Geissler (Bielefeld), Jürgen Gerhards (FU Berlin), Ronald Hitzler (Dortmund), Hans Joas (Erfurt/Chicago), Karl-Dieter Keim (Berlin), Siegfried Lamnek (Eichstätt-Ingolstadt), Richard Münch (Bamberg), Friedhelm Neidhardt (Berlin – Vors.), Werner Rammert (TU Berlin), Bernhard von Rosenblatt (Infratest München), Uwe Schimank (Hagen), Heike Solga (Göttingen, jetzt WZB), Peter Weingart (Bielefeld) und Reinhard Wippler (Utrecht), als fachfremdes Mitglied aus dem Wissenschaftsrat Peter Lichter (Krebsforschungszentrum Heidelberg). In der Geschäftsstelle des Wissenschaftsrats waren Rainer Lange und Moritz Mälzer für das Forschungsrating zuständig.

2 http://www.wissenschaftsrat.de/pilot_start.htm

Darstellung von Erfahrungen und Befunden beschränken, die mir besonders instruktiv erscheinen. Eine Podiumsveranstaltung, die am 8. Oktober 2008 auf dem anstehenden Soziologiekongress stattfinden soll, wird Gelegenheit zu Nachfragen und Diskussionen bieten.

Die Datenerhebung

Das Forschungsrating des Wissenschaftsrats folgt dem Prinzip des »Informed Peer Review«. Die zur Bewertung benötigten Daten werden nicht im Sinne einer »Kennziffernsteuerung« zu einem Indikatorensatz verdichtet, aus dem rein rechnerisch Bewertungen abgeleitet werden. Entscheidend ist, dass Gutachter, in diesem Falle die Mitglieder der Bewertungsgruppe, bei allen Falleinschätzungen unter anderem prüfen, ob und in welchem Maße die verfügbaren Deskriptoren auch valide Indikatoren sind, bevor sie zur Bewertung eingesetzt werden.

Die Qualität der Gutachterurteile hängt dabei aber zuerst einmal von der *Zuverlässigkeit der Daten* ab. Die damit gesetzten Ansprüche an die Datenerhebung waren im vorliegenden Falle schon dadurch nicht leicht einlösbar, dass es sich um ein Pilotprojekt zu einer Massenerhebung handelt. Das Forschungsrating sollte für die Soziologie flächendeckend sein – und in der Tat ist eine außerordentlich hohe Erfassungsquote der soziologischen Forschungseinrichtungen erreicht worden: Es gelang dank des auf die Universitäten durchgreifenden Einflusses des Wissenschaftsrats und den Bemühungen seiner Geschäftsstelle fast eine Vollerhebung. Für 57 soziologische Forschungseinrichtungen (Universitäten und drei außeruniversitäre Forschungsinstitute) mit insgesamt 254 Forschungseinheiten (Lehrstühlen, Abteilungen) liegen die einschlägigen Daten vor, und diese beziehen sich auf 376 Professoren mit 999 Mitarbeitern.

Die Bewertungsgruppe konnte zwar davon ausgehen, dass die ihr verfügbaren Datensätze bessere und reichhaltigere Informationen bieten, als dies für die bisher vorliegenden Universitätsrankings der Soziologie der Fall ist. Es sind in der WR-Pilotstudie aber gleichwohl Zuverlässigkeitsdefizite zu registrieren, die Anlass gaben, im Abschlussbericht modifizierende Verfahrensvorschläge für die Nachfolgestudien zu machen, die der Wissenschaftsrat für andere Disziplinen plant – später vielleicht noch einmal auch für die Soziologie. Sie betreffen unter anderem (a) Verbesse-

rungsmöglichkeiten bei den elektronischen Fragebögen der WR-Geschäftsstelle. Sie beziehen sich zudem (b) auf die ungleichen Informationsleistungen der Universitäten, da sich bei den Erhebungen ergab, dass viele von ihnen kein leistungsfähiges Berichtssystem besitzen, also über sich selber nicht gut Bescheid wissen.

Entwicklungsbedarf besteht auch (c) beim Service des Informationszentrums Sozialwissenschaften, das seine Publikationsdaten bisher nicht unter dem Gesichtspunkt institutionenbezogener Evaluationen erhoben hat, sich nun aber auf derartige Anfragen einzurichten beginnt.³ Als besonders aufwändig und schwierig erwies sich die für alles Weitere sehr wichtige Erstellung korrekter Publikationslisten. Da die Nachweise des IZ durchweg nicht vollständig waren, wurden für alle beteiligten Wissenschaftler über die Fachkoordinatoren, die die Universitäten für die Erhebung benannt hatten, passwortgeschützte Zugriffe auf die für sie erstellten Publikationslisten eingerichtet und Korrekturmöglichkeiten über eine Online-Maske angeboten. Es lag sicher nicht nur an vielleicht unzureichenden Einsätzen dieser Fachkoordinatoren, dass (d) ein Drittel der Professoren und insgesamt die Hälfte aller erfassten Wissenschaftler diese Kontrollchance nicht genutzt haben – zu ihrem Nachteil.

Auch wenn all diese Ungereimtheiten verschwinden würden, bliebe für das Forschungsrating ein Erhebungsproblem, nämlich die Festlegung der Bewertungseinheiten. Dazu einige Bemerkungen.

Bewertungseinheiten

Die Ergebnisse des Forschungsratings dürfen nicht als Kompetenz einschätzungen von Personen missverstanden werden; es geht bei ihnen nicht darum, wie gut Herr X und Frau Y sind. Beurteilt werden bestimmte Leistungen ihrer Einrichtungen in einem bestimmten Zeitraum, nämlich in den Jahren 2001 bis 2005. Welche Einrichtung aber ist der tatsächliche Leistungsträger für die Leistungen, die gemessen werden sollen? Welche Aggregationsebene liefert valide Zurechnungsgrößen?

³ Das IZ hat einen Bericht zur Publikationserhebung für das Forschungsrating Soziologie vorgelegt: http://www.gesis.org/Publikationen/Berichte/GESIS_Berichte/GESIS_Arbeitsbericht5.pdf

In den Forschungs- und Lehrein-schätzungen deutscher Universitäten, die man in einigen Zeitschriften (Stern, Fokus, Zeit) immer wieder lesen kann, sind Fakultäten bzw. Disziplinenbereiche in Fakultäten als Bewertungseinheiten ausgewiesen. Das mag in der Lehre angemessen sein, da deutsche Universitäten nach Lehrgesichtspunkten organisiert sind. Im Hinblick auf Forschungsleistungen aber sind »Forschungseinheiten« mit gemeinsamer Forschungsplanung und abgestimmten Projektgestaltungen in der Soziologie nur selten oberhalb der Ebene von Professuren (Lehrereinheiten) institutionalisiert. Es kann deshalb zumindest nicht hinreichend sein, die Forschungsqualität der Universitätsdisziplinen nur pauschal auszuweisen.

Eine Bestätigung dafür bieten die Ergebnisse des Forschungsratings in dem zentralen Kriterienbereich I (Forschungsqualität), in dem in der Pilotstudie nicht »Forschungseinrichtungen«, verstanden als die Summe aller Organisationseinheiten der Soziologie in Universitäten bzw. außeruniversitären Forschungsinstituten, bewertet wurden, sondern deren Unteraggregate, die sich als »Forschungseinheiten« definiert haben. In der überwiegenden Mehrheit der Fälle streuten in den Forschungseinrichtungen die Bewertungen für deren Forschungseinheiten über zwei bis drei Noten; in einem Falle waren in einer Universität die ausgewiesenen Forschungseinheiten über die gesamte fünfstellige Notenskala von »exzellent« bis »nicht befriedigend« verteilt.

Das Problem bei dieser Messung ist allerdings, dass soziologische Forschungseinheiten in den Universitäten nicht institutionalisiert sind. Sie werden in den Einrichtungen für die Zwecke des Forschungsratings nach den allgemeinen Vorgaben des WR ad hoc definiert, und dabei spielen sicher auch taktische Kalküle eine Rolle. Das lässt sich an den Daten erkennen. Einerseits bestehen drei Viertel der ausgewiesenen Forschungseinheiten aus nur einer Professur mit durchschnittlich eher zwei als drei überwiegend teilzeitbeschäftigten Mitarbeitern; und dieser Befund weist auf die tatsächlich vorhandene Kleinteiligkeit der soziologischen Forschungsinfrastruktur hin. Andererseits existieren zehn Fälle von Forschungseinheiten mit fünf und mehr Professoren, ohne dass über die eingereichten Daten nachvollziehbar gewesen wäre, ob und in welchem Maße es sich hier um Aggregate mit signifikant ausgeprägten Kooperationen und nicht um bloße Artefakte handelt. Die Bewertungsgruppe hat bei der Indikatorenbildung zumindest die Größeneffekte solcher Konstrukte neutralisieren, die mit ihnen verbundenen Differenzierungsverluste jedoch nicht identifizieren können.

Indikatorenkalküle

Die größte akademische Herausforderung bestand für die Bewertungsgruppe darin, Daten erheben zu lassen, die sich als valide Indikatoren im Hinblick auf die Leistungen interpretieren lassen, die beurteilt werden sollen. Solche Indikatoren wurden in einem »Qualitätsmodell« (AB: 75–79) für die sechs verschiedenen Kriterienbereiche (I. Forschungsqualität, II. Impact/Effektivität, III. Effizienz, IV. Nachwuchsförderung, V. Transfer, VI. Wissensvermittlung) festgeschrieben. Dabei konnten den beiden letztgenannten Kriterienbereichen nur qualitative Indikatoren zugeordnet werden, sodass sich die Bewertungsgruppe entschloss, hier nur mit einer dreistatt einer fünfstelligen Notenskala zu arbeiten.

Die unterschiedliche Indikatorenfundierung des Forschungsratings stand im Zusammenhang mit dem Prinzip, *prozessproduzierte Daten* für die eigene Urteilsbildung zu nutzen: Das Peer Review der Bewertungsgruppe sollte, wo immer möglich, den Peer-Review-Praktiken angeschlossen werden, die im normalen Wissenschaftsprozess bei der Bewertung von Projektanträgen, Manuskripten, Publikationen, Kandidaturen etc. ständig und in großer Zahl stattfinden. Auf diese Weise können in einem »Informed Peer Review« die jeweils engagierten Gutachter bei ihrer Urteilsbildung durch Daten informiert und kontrolliert werden, in denen zahlreiche Bewertungsergebnisse anderer Gutachter aggregiert sind. Das aber funktioniert nicht in jeder Hinsicht gleich gut. Die im Wissenschaftsprozess laufenden akademischen Qualitätskontrollen sind vor allem in Bereichen der Wissenschaftsanwendung relativ schwach ausgeprägt; sie liegen auch für die Nachwuchsförderung nur teilweise vor. Und wo sie, wie für die drei forschungsbezogenen Kriterienbereiche I–III, zahlreicher vorliegen, erscheinen sie nicht durchweg valide. Auch wenn sie valide für den Nachweis bestimmter Leistungen erscheinen, bleibt noch die Frage zu erwägen, ob es wünschbar ist, diese Leistungen über die normativen Effekte der Indikatorenbildung noch zusätzlich anzureizen.

Die Bewertungsgruppe hat das Für und Wider für ihre eigenen Entscheidungen über den Indikatoreneinsatz im Abschlussbericht ausführlicher dargestellt und für dessen weitere Qualifizierung Vorschläge gemacht (AB: 37–40, 58–60, 63–73). Hier nur drei Beispiele.

(1) Im Forschungsrating wurden mehrere durchaus gängige *Vernetzungsdaten* nicht erhoben, obwohl es üblich ist, dass Gastaufenthalte, Konferenzteilnahmen, Tagungsveranstaltungen in Evaluationen positiv bewertet wer-

den. Die Entscheidung der Bewertungsgruppe ergab sich aus dem Eindruck, dass es sich hier eher um Aktivitäts- als um Qualitätsausweise handelt und dass ihre Indikatorenutzung in folgenreichen Evaluationen eher einen Anreiz für Geschäftigkeit als einen Stimulus für wissenschaftliche Leistungssteigerung auslösen könnte.

(2) Die Bewertungsgruppe hat auf den Einsatz von *Drittmitteldaten* nicht verzichtet, hat diese aber nur zurückhaltend und mit bestimmten Spezifikationen benutzt – dies aus folgenden Gründen: (a) Im Hinblick auf Drittmittelabhängigkeit unterscheiden sich nicht nur Disziplinen, sondern in Disziplinen auch die diversen Forschungsbereiche, ohne dass es für diesen Zusammenhang passable Algorithmen gäbe. (b) Als Qualitätsausweis erscheinen Drittmittel nur dann, wenn ihre Vergabe einer verlässlichen Qualitätskontrolle folgt; deshalb ist es wichtig zu berücksichtigen, aus welchen Quellen Drittmittel stammen. (c) Auch bei qualitätskontrollierten Förderungen eignet sich als Indikator für die meisten Zwecke weniger die Summe von Drittmitteln als die Anzahl ihrer Bewilligungen. (d) Schließlich ist hier wie bei anderen quantitativen Indikatoren nicht anzunehmen, dass ihr Zusammenhang mit Leistungsmerkmalen linear verläuft (für Drittmittel Jansen et al. 2008: 136ff.). Die Maximierungslogik des »Je mehr, umso besser« dürfte auch in der Wissenschaft zur Qualitätssteigerung nicht unbedingt beitragen. Umso erstaunlicher ist es angesichts dieser Einschränkungen, dass in verschiedenen Bundesländern und in zahlreichen Universitäten Leistungszuschläge für die Forschung weit überwiegend in Abhängigkeit von eingeworbenen Drittmittelsummen beliebiger Herkunft verteilt werden – ohne dass die Wissenschaftsorganisationen gegen die derart umgesetzte »Tonnenideologie« (Menninghaus 2008) vernehmbar protestiert hätten.

(3) *Publikationen* wurden von der Bewertungsgruppe als instruktive Outputindikatoren im Forschungsbereich der Kriterien I–III behandelt. Aber auch deren Nutzung erscheint aus verschiedenen Gründen nicht unproblematisch (allgemein: Hornbostel 1997: 237–307). Schwierigkeiten ergeben sich vor allem aus den ungleich ausgeprägten Kontrollen der verschiedenen Publikationssorten. Von den 10.622 Publikationen, die von deutschen Soziologen für den Zeitraum 2001 bis 2005 im Forschungsrating erfasst wurden, waren 7,3% Monographien, 6,8% Sammelbände, 45,2% Sammelbandbeiträge, 34,4% Zeitschriftenaufsätze und 6,3% Rezensionen. Alle Publikationstypen wurden für die Bewertung der einzelnen Forschungseinrichtungen ausgezählt, um deren Anteil am jeweiligen Gesamtausstoß

des Faches würdigen zu können, und die jeweiligen Publikationslisten wurden auch qualitativ analysiert. Aber nur jene Publikationen erhielten ein besonderes Gewicht, für deren Veröffentlichung mit einiger Zuverlässigkeit angenommen werden konnte, dass ihr eine Qualitätsprüfung durch Peer Review vorausgegangen war. Dies war ausschließlich bei einem Drittel der Zeitschriftenaufsätze der Fall, sodass insgesamt nur eine kleine Minderheit der Publikationen bei den quantitativen Analysen durchweg und übergewichtet veranschlagt werden konnte. Das lieferte auch einen der Gründe dafür, dass im Forschungsrating Soziologie nicht mit Zitationsanalysen gearbeitet wurde (dazu ausführlicher AB: 48, 65f.). Bedauerlich erscheint vor allem, dass keine generelle Qualitätseinschätzung von Monographien und Sammelbänden möglich war (vgl. Clemens et al. 1995), da die deutschen Verlage eine fachlich korrekte Qualitätsprüfung vor der Veröffentlichung ihrer Fachbücher nicht routinisiert haben. (Dass das CHE für seine Notenbildung Publikationen pauschal und nach Seitenlänge gewichtet, erschien der Bewertungsgruppe nicht nur nicht valide, sondern auch normativ nicht vertretbar: Es erscheint nicht geboten, über die in Evaluationen gesetzten Gütemaßstäbe Anreize dafür zu setzen, dass noch mehr und noch ausladender veröffentlicht wird als jetzt schon.)

Der Bewertungsprozess

Alle Anmerkungen zum Indikatoreneinsatz weisen darauf hin, dass eine reine Kennziffernbestimmung von wissenschaftlichen Leistungen – wie in vorhandenen Rankings üblich – nicht statthaft ist. Peer Review bedarf zwar der Orientierung durch quantitative Kennziffern, aber es ist selber unverzichtbar für die gebotenen Prüfungen von deren Zuverlässigkeit und Gültigkeit (Weingart 2005: 110ff.). Solche Prüfungen sind allerdings sehr aufwändig. Auch den 15 Fachmitgliedern der Bewertungsgruppe wurde im Forschungsrating viel Einsatz abverlangt. Sie kamen im Zeitraum Februar 2006 bis März 2008 zu 11 Sitzungen mit insgesamt 15 Sitzungstagen zusammen, und sie brauchten für deren Vor- und Nachbereitung einen Zeitaufwand, der insgesamt auf jeweils zwei bis drei volle Arbeitsmonate einzuschätzen ist.

Dieser Aufwand entstand nicht zuletzt aus der Überzeugung, dass die dargestellten Publikationsprobleme einen unmittelbaren Einblick in die veröffentlichten Forschungsarbeiten der Forschungseinheiten erforderlich

machen, um die Benotungen von Forschungsqualität fundieren zu können. Die Forschungseinheiten wurden deshalb aufgefordert, aus dem Berichtszeitraum 2001 bis 2005 je nach ihrer Größe eine bestimmte Anzahl von Publikationen eigener Wahl einzureichen. Da diese von jeweils zwei Gutachtern eingesehen werden sollten, erhielt jeder Gutachter zwischen 80 und 90 Publikationen von maximal 50 Seiten zur Prüfung – dies mit erheblichen Effekten im Endresultat: Die definitive Notengebung wich aufgrund der qualitativen Einschätzungen der »Forschungsqualität« durch die Gutachter in etwa einem Drittel aller Fälle von den Notenwerten ab, die sich durch reine Kennzifferrechnung ergeben hätten.

Dieser Effekt konnte nur akzeptabel sein, weil im Bewertungsprozess *systematische Kontrollen* der jeweiligen Gutachternoten eingerichtet waren. Sie erscheinen notwendig, weil nicht wenige Untersuchungen darauf verweisen, dass »Reliabilität, Fairness und Validität des Peer-Review-Verfahrens« nicht per se zu unterstellen sind (Bornmann, Daniel 2003). Der Forschungsstand ist in dieser Hinsicht allerdings widersprüchlich. Das lässt sich vor allem damit erklären, dass die in der Wissenschaft sehr variablen institutionellen Bedingungen, unter denen Peer-Review stattfindet, in der einschlägigen Forschung kaum thematisiert und berücksichtigt werden (Neidhardt 2006). Begutachtungsprozesse unterscheiden sich im Hinblick auf die Sorgfalt, mit der Gutachter rekrutiert sowie deren Fachkompetenz und Unbefangenheit kontrolliert werden. Hinzu kommt, dass bei der Einschätzung von Personen, Texten und Projekten die teilweise dissentierenden Voten der eingesetzten Gutachter normalerweise nicht mehr als den Rang von Zwischenprodukten in den Entscheidungsprozessen über die Verteilung von Ämtern und Preisen, Forschungsgeldern und Veröffentlichungschancen besitzen. Entscheidungen treffen Komitees, Fachkollegien, Herausbergremien etc., und sie tun dies nicht zuletzt mit kritischer Nutzung von instruktivem Gutachterdissens.

Im Forschungsrating wurde diesen Bedingungen Rechnung getragen: (a) Die Mitglieder der Bewertungsgruppe wurden aus den Vorschlagslisten unter anderem der DGS und der DFG gewählt, und weil dabei (siehe Fußnote 1) auf eine Heterogenität sowohl der paradigmatischen Orientierungen als auch der fachlichen Spezialitäten der Kandidaten und Kandidatinnen geachtet wurde, darf unterstellt werden, dass die Reichweite gutachterlicher Kompetenzen insgesamt recht groß war. Wo sie für bestimmte Fälle nicht ausreichend erschien, wurden zusätzlich externe Sondergutachter herangezogen. (b) Für den Einsatz der Mitglieder der Bewertungs-

gruppe bei der Begutachtung der verschiedenen Forschungseinrichtungen galten im Übrigen streng gefasste Befangenheitsregeln.⁴ (c) Vor allem aber wurde durch eine Mehrstufigkeit des Bewertungsprozesses für die systematische Kontrolle der Gutachtervoten gesorgt (Prinzip der »Mehrinstanzlichkeit«, Hirschauer 2004: 73ff.). Dazu einige Anmerkungen.

In einer ersten Bewertungsphase waren für die Einzelfälle jeweils zwei Mitglieder der Bewertungsgruppe als Gutachter tätig. Kamen sie zu keinem übereinstimmenden Votum, wurden sie in der zweiten Bewertungsphase angehalten, ihren Dissens wenn nicht aufzulösen, so doch aufzuklären. In einer dritten Bewertungsphase wurden alle Einzelurteile in der Gesamtgruppe aufgerufen, häufig (vor allem bei unaufgelöstem Dissens) kommentiert und am Ende vom Plenum beschlossen. Nach meistens kurzer, gelegentlich aber auch längerer Diskussion ergaben sich durchweg einmütige Notenfestlegungen.⁵ (Ausführlicher zum Bewertungsprozess AB: 24ff., 50f.) Es stellte sich also beim Forschungsrating eine ähnliche Erfahrung ein wie in Herausgeber- und Gutachtergruppen (zum Beispiel der DFG): In gemeinsamen Diskussionen erreicht der kollegiale Bewertungsprozess eine elementare Ebene geteilter Gütemaßstäbe, die ein Einvernehmen bei Qualitätseinschätzungen auch jenseits aller paradigmatischen Konfessionen möglich und wahrscheinlich machen.

Ratingfolgenabschätzungen

Das mit hohem Aufwand betriebene Verfahren rechtfertigt die Vermutung, dass die Fehlermenge im Forschungsrating vergleichsweise gering ist. Es schließt solche Fehler angesichts der dargestellten Probleme der Datenerhebung und der Indikatorenbildung aber sicher nicht aus. Und die Sorgfalt des Vorgehens nimmt dem Gesamtprojekt auch nicht den Charakter eines Konstrukts, dessen Erhebungspraktiken, Datenoperationalisierungen, Indikatorenkalküle und Skalendefinitionen wie überall in der Sozialfor-

4 Betraf eine Begutachtung die eigene Forschungseinheit eines Mitglieds der Bewertungsgruppe (das war in 10 Fällen gegeben), wurden sie von mir als Vorsitzenden der Bewertungsgruppe zusammen mit zwei Sondergutachtern durchgeführt. Die Ergebnisse wurden in der Bewertungsgruppe nicht diskutiert.

5 Der einzige Ausnahmefall führte dazu, dass keine Noten vergeben wurden.

schung nicht durchgängig nach dem Muster von »richtig oder falsch« zu bestimmen waren. Das Design des Projekts ergab sich aus Güteabwägungen, die im Einzelnen auch anders hätten ausfallen können, als sie unter den wahrgenommenen Umständen im Forschungsrating balanciert wurden.

Es ist dann auch nicht überraschend, dass Ambivalenzen über die Vernünftigkeit des folgenreichen Unternehmens in der Bewertungsgruppe selber verbreitet waren. Gleichwohl entstand in der letzten Sitzung nach Abschluss aller Arbeiten in der Gruppe ein voller *Konsens* darüber, dass es vertretbar und sinnvoll sei, die Ergebnisse für die öffentliche Nutzung durch den Wissenschaftsrat freizugeben. Für die Nachvollziehbarkeit dieser Entscheidung sprechen neben der relativen Güteeinschätzung des Forschungsratings auch sehr grundsätzliche Überlegungen: Evaluationen werden aus der Wissenschaft nicht mehr verschwinden, und sie werden auch die Soziologie regelmäßig vermessen. Entscheidend wird sein, dass die betroffenen Disziplinen diesen Prozess weder den Interessen von Universitätsleitungen und Ministerialbürokratien noch den Marktabhängigkeiten von privaten Instituten und bestimmten Zeitungen überlassen. »... academics should no longer leave evaluations to others, but should invest in self-defined measures of quality, relevance, and efficiency, and in the collection and propagation of data...« (Schimank 2005: 375).

Die Bewertungsgruppe hat dem Wissenschaftsrat vorgeschlagen, Forschungsratings unter bestimmten Qualitätsbedingungen zu routinisieren (AB: 57ff.)⁶, und dafür spielte auch folgendes eine Rolle: Die Treuhänderschaft des Wissenschaftsrats würde einerseits eine gewisse Verbindlichkeit der Forschungsratings garantieren und andererseits relativ wahrscheinlich machen, dass diese im Grundsätzlichen in der Regie der Wissenschaft bleiben. Deren Einfluss wäre freilich nicht schon dadurch gesichert, dass von den Fächern bei Bedarf kleine Gruppen von Kollegen und Kolleginnen für das Evaluationsgeschäft abgestellt werden. Auch im vorliegenden Fall ist mit den Berichten der Bewertungsgruppe das Thema für die Soziologie nicht erledigt. Diese Berichte verstehen sich als Teil eines Prozesses,

⁶ Dies allerdings nicht unbedingt in den vom WR erwogenen Fünfjahresabständen. Zehnjahresabstände würden, meine ich, einerseits die Untersuchungsphasen verlängern und damit die Erhebungen auf eine breitere Datenbasis beziehen, und sie würden andererseits die durch den Erhebungsaufwand entstehenden Belästigungen der Betroffenen vermindern. Den Universitäten bliebe es im Übrigen unbenommen, zur Selbstkontrolle ihrer Fächer nach dem Muster des Forschungsratings in den Zwischenzeiten Evaluationen durchzuführen.

in den sich das Fach mit seinen einschlägigen Experten und die DGS mit ihrem wissenschaftspolitischen Auftrag einschalten müssen.

Dabei geht es einerseits um die *Kontrolle der Professionalität* des bisherigen Verfahrens und der Schlüssigkeit der Empfehlungen, die die Bewertungsgruppe selber zur weiteren Entwicklung des Verfahrens gegeben hat. Im Hinblick darauf soll eine offene Berichterstattung zum Forschungsrating des Wissenschaftsrats die kritische Urteilsbildung im Fach ermöglichen. Diese wird dadurch fundierbar werden, dass die Geschäftsstelle des Wissenschaftsrats mit dem Kölner Zentralarchiv inzwischen vereinbart hat, die Daten des Forschungsratings, angemessen anonymisiert, noch in diesem Jahr für Sekundäranalysen freizugeben. Unabhängig von den gebotenen Qualitätskontrollen geht es aber bei den Versuchen der Profession, die Dinge nicht aus der Hand zu geben, auch noch um Grundsätzlicheres, nämlich um die *Kontrolle der Steuerungseffekte*, die von den Forschungsratings ausgehen (Nelson Espeland, Sauder 2007).

Steuerungseffekte resultieren aus Evaluationen (a) dadurch, dass sie auch (und vielleicht besonders) von denen wahrgenommen werden, die die Forschung symbolisch und materiell unterstützen – mit Reputation, Status und natürlich Geld. Reagieren sie mit Ressourcenverstärkung bei denen, die in Forschungsratings gut abschneiden, und mit Ressourcenverknappung bei denen, die dabei schlecht abschneiden, dann verändert sich die Statusdifferenzierung der Disziplin im Sinne des »Matthäus-Prinzips« – mit der Folge einer self-fulfilling prophecy (Münch 2008: 73ff.). Tendenzen in diese Richtung wirken dann (b) auf die Motivationen der Forscher ein. Rationale Akteure werden ihre Forschung auf die Kriterien hin ausrichten, mit denen sie folgenreich gemessen werden. Insofern ist jede Indikatorensetzung in Forschungsratings ein normativer Akt.

Die Bewertungsgruppe hat versucht, sich solche Effekte vorzustellen und je nach ihrer Wünschbarkeit bei der Wahl und Operationalisierung von Indikatoren zu berücksichtigen. Jedoch gibt es keine instruktiven Studien über Bedingungen und Wirkungen von Steuerungen durch Evaluation (vgl. aber Gläser et al. 2008). Deshalb hat die Bewertungsgruppe den Wissenschaftsrat auch angehalten, Forschungen im Sinne von »Ratingfolgenabschätzung« anzuregen und zu unterstützen. Aber auch hier wäre es die Sache der Deutschen Gesellschaft für Soziologie, aufzupassen. In dem Maße, in dem sich die Universitäten dem Regime eines »New Public Management« anpassen, werden die Fachgesellschaften gefordert sein, die Integrität der Wissenschaft auch nach außen hin stärker zu schützen als bisher.

Literatur

- Bornmann, L., Daniel, H.-D. 2003: Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. In S. Schwarz, U. Teichler (Hg.), *Universität auf dem Prüfstand*. Frankfurt a.M.: Campus, 207–225.
- Clemens, E.S., Powell, W. W., McIlwaine, K., Okamoto, D. 1995: Careers in Print. Books, Journals, and Scholarly Reputations. *American Journal of Sociology*, Vol. 101, No. 2, 433–494.
- Gläser, J., Lange, S., Laudel, G., Schimank, U. 2008: Evaluationsbasierte Forschungsfinanzierung und ihre Folgen. In R. Mayntz, F. Neidhardt, P. Weingart, U. Wengenroth (Hg.), *Wissensproduktion und Wissenstransfer*. Bielefeld: Transcript, 145–172.
- Hirschauer, S. 2004: Peer Review Verfahren auf dem Prüfstand. Zum Soziologie-defizit der Wissenschaftsevaluation. *Zeitschrift für Soziologie* 33. Jg., Heft 1, 62–83.
- Hornbostel, S. 1997: *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westdeutscher Verlag.
- Jansen, D., Wald, A., Franke, K., Schmoch, U., Schubert, T. 2007: Drittmittel als Performanzindikator der wissenschaftlichen Forschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59. Jg., Heft 1, 125–149.
- Menninghaus, W. 2008: Tonnenideologie. Elite-Universitäten durch Masse? *Süddeutsche Zeitung*, Nr. 96 vom 26.4.2008, 14.
- Münch, R. 2008: Stratifikation durch Evaluation: Mechanismen der Konstruktion von Statushierarchien in der Forschung. *Zeitschrift für Soziologie*, 37. Jg., Heft 1, 60–80.
- Neidhardt, F. 2006: Fehlerquellen und Fehlerkontrollen in den Begutachtungssystemen der Wissenschaft. In S. Hornbostel, D. Simon (Hg.), *Wie viel (In-)Transparenz ist notwendig? Peer Review revisited*. iFQ-Paper No.1, Dezember 2006, 7–14.
- Nelson Espeland, W., Sauder, M. 2007: Rankings and Reactivities – how Public Measures recreate Social Worlds. *American Journal of Sociology*, Vol. 113, No. 1, 1–40.
- Schimank, U. 2005: »New Public Management« and the Academic Profession. Reflections on the German situation. *Minerva*, Vol. 43, 361–367.
- Weingart, P. 2005: Das Ritual der Evaluierung und die Verführung der Zahlen. In P. Weingart, *Die Wissenschaft der Öffentlichkeit*. Weilerswist: Velbrück, 102–122.