Das Forschungsrating des Wissenschaftsrats für die Soziologie in Deutschland revisited

Von Katrin Auspurg, Andreas Diekmann, Thomas Hinz und Matthias Näf

Zusammenfassung: Die Bewertung wissenschaftlicher Leistungen gehört zum Alltagsgeschäft – auch in der Soziologie. Der bislang umfassendste, gleichwohl umstrittene Versuch, die Forschungsleistung der deutschen Soziologie-Standorte zu messen, war das im Jahr 2008 abgeschlossene Rating des Wissenschaftsrats (WR). Die eingesetzte Evaluationsgruppe aus 16 anerkannten Soziologinnen und Soziologen hatte etwa 250 Forschungseinheiten an mehr als 50 Forschungsstandorten mit fünf Kategorien von „exzellent" bis „nicht befriedigend" zu bewerten. Die Literaturauswertung basierte auf über 10.000 quantitativ erfassten Veröffentlichungen und etwa 700 eingereichten, zu lesenden Buchkapiteln und Aufsätzen. Wie lässt sich dieses ambitionierte Pilotprojekt rückblickend einschätzen? In welchem Verhältnis stehen Aufwand und Ertrag? Der vorliegende Artikel versucht auf Grundlage der Daten des Wissenschaftsrats die Bewertungen auf der Ebene von Forschungseinheiten zu rekonstruieren. Im Ergebnis zeigt sich, dass die Qualitätsbewertung maßgeblich Outputgrößen wie die Zahl der Aufsätze in Zeitschriften mit *peer review* berücksichtigt. Zudem zeigt sich, dass ein großer Anteil der Varianz der Urteile der Bewertungsgruppe mit einfachen quantitativen Indikatoren erklärt werden kann. Auch die Größe der Einheiten wurde von der Bewertungsgruppe offenbar berücksichtigt.

1. Einleitung

Evaluationen wissenschaftlicher Leistungen gehören zum Alltagsgeschäft aller wissenschaftlichen Disziplinen. Die große Mehrzahl der anzufertigenden Gutachten und Bewertungen bezieht sich auf die erbrachte bzw. in Aussicht gestellte Leistung einzelner Personen oder Personengruppen im Wissenschaftssystem. Daneben besteht aber oft auch ein Bedarf an Urteilen zu ganzen wissenschaftlichen Einrichtungen und Fächern. Diese Bewertung von wissenschaftlicher Leistung ist ganz überwiegend als Prozess gegenseitiger Expertenurteile organisiert: Im sogenannten *peer review* werden Entscheidungen über Publikationen, Projekte und Personalauswahlen vorbereitet. Das Wissenschaftssystem alloziert seine Belohnungen ganz wesentlich nach den Urteilen der *peers*, auch wenn dieses Verfahren mit vielfältigen Problemen verbunden ist und immer wieder kritisiert wird (Bornmann / Daniel 2003).[1] Da in Deutschland wie in vielen anderen Ländern die Grundlagenforschung zu großen Teilen aus öffentlichen Steuermitteln finanziert wird, hat die Politik ein großes Interesse daran, die Mittelvergabe an Universitäten, Forschungsinstitute und innerhalb der Einrichtungen an Arbeitsgruppen und Einzelpersonen – wie man heute sagt – evidenzbasiert zu organisieren. Mit dem Wissenschaftsrat (WR) existiert in Deutschland eine Institution, in der Fragen der Leistungsfähigkeit des Wissenschaftssystems evidenzbasiert unter Beteiligung der das System tragenden finanziellen Körperschaften (Bund und Länder) und anerkannten Akteuren aus der Wissenschaft diskutiert werden. Viele wichtige Strukturentscheidungen werden in Deutschland vom WR vorbereitet.

Mitte der letzten Dekade startete der WR eine Reihe von fachbezogenen Evaluationen universitärer und außeruniversitärer Forschung. Die zugrunde gelegten Verfahren zielten dabei auf maximale Transparenz und Akzeptanz in den jeweiligen Fächern. Das angestrebte Forschungsrating sollte den verantwortlichen Akteuren möglichst hilfreiche Informationen

1 Zweifel bestehen an der Objektivität von Leistungsmessungen und damit auch an einer häufig uneinheitlichen Gutachtenlage. Weiterhin steht die Evaluation von Wissenschaft unter Verdacht, vornehmlich Mainstream-Forschung zu unterstützen, innovatives Querdenken falle im *peer review* leicht durchs Raster (s. z.B. Hirschauer 2004; Münch 2007).

Soziale Welt 66 (2015), S. 177 – 191

zur Qualität der geleisteten Forschung liefern. Ausdrücklich war kein Forschungsranking beabsichtigt. Bevor das Rating auf alle Fächer angewandt werden kann, sollten für ausgewählte Fächer zunächst wissenschaftsadäquate Methoden und Verfahrensweisen erprobt werden, mit denen wissenschaftliche Leistungen möglichst objektiv, reliabel und valide gemessen werden, unter Berücksichtigung der unterschiedlichen Wissenschaftsdisziplinen und ihrer Leistungskriterien. Von 2006 bis 2011 wurden die Fächer Soziologie, Chemie, Elektro- und Informationstechnik sowie Anglistik/Amerikanistik evaluiert. Weiterhin wurden Vorüberlegungen zur Anwendbarkeit des Forschungsratings in den Geisteswissenschaften angestellt (Wissenschaftsrat 2010).

Vorliegender Beitrag rekonstruiert auf der Grundlage der vom WR zur Verfügung gestellten Daten das Forschungsrating für das Fach Soziologie auf der Ebene von 229 wissenschaftlichen Einheiten in 53 Universitäten. In den Analysen steht zunächst die Frage nach dem Verhältnis von Input- und Outputgrößen und ihrer Gewichtung im Mittelpunkt. Anschließend können nachträgliche Zitationsanalysen der im Verfahren bewerteten Literaturauswahl darüber informieren, ob und inwieweit die erzielten Gutachterurteile zur Bewertung der Forschungsleistungen mit späteren Zitationen der begutachteten Literatur übereinstimmen (prädiktive Validität).

2. Das Forschungsrating Soziologie

Die Soziologie wurde im Jahr 2006 als sozialwissenschaftliches Fach für eine Pilotstudie ausgewählt (Neidhardt 2006, 2008, 2009).[2] Das eingesetzte Evaluationsverfahren bestand in einem sogenannten *informed peer review*. Dies bedeutet, dass in der Bewertungsgruppe unter Vorsitz von Friedhelm Neidhardt nicht nur Kennzahlen heranzuziehen waren, sondern diese Kennzahlen (etwa die Anzahl der Publikationen und Drittmitteleinwerbungen) durch ein Expertengremium (nachfolgend: die „Bewertungsgruppe“) reflektiert und diskutiert werden sollten. Die Mitglieder dieser Bewertungsgruppe sollten die von den begutachteten Einrichtungen eingesandten Schriften lesen und ihre Leseeindrücke bei der anschließenden Diskussion der Forschungsqualität einbringen. Insgesamt deckte das Forschungsrating in der Soziologie sechs verschiedene Dimensionen für den Zeitraum 2001 bis 2005 ab: Forschungsqualität, Impact/Effektivität, Effizienz, Nachwuchsförderung, Transfer, Wissensvermittlung. Als wissenschaftsimmanent sind vor allem die ersten beiden Kriterien der Forschungsqualität und des erzielten Impacts zu sehen, wobei der Impact als „Beitrag zur Entwicklung der Wissenschaft im Fachgebiet und darüber hinaus“ definiert war (Wissenschaftsrat 2006). Zur Bewertung der Forschungsqualität wurden vor allem die Veröffentlichungen, zur Bewertung des Impacts zusätzlich die Einwerbung wissenschaftsgesteuerter Drittmittel, die Zahl nicht-deutschsprachiger Publikationen, die Zahl der Veröffentlichungen in Fachzeitschriften außerhalb der Soziologie sowie Forschungspreise, Auszeichnungen und Ämter in wissenschaftlichen Institutionen und Gremien herangezogen. Der Impact, so wie er vom WR definiert wurde, ist mehrdimensional und nicht deckungsgleich mit dem, was gemeinhin als Impact von Veröffentlichungen verstanden und in der Regel durch die Anzahl der Zitationen gemessen wird (s. dazu z.B. Bornmann et al. 2008).

Eine zentrale operative Frage des Forschungsratings lautete, auf welcher organisatorischen Ebene die Evaluation angesetzt werden sollte. Man unterschied letztlich *Forschungseinrichtungen*, also Universitäten oder außeruniversitäre Forschungsinstitute, und die feinere Ebene der *Forschungseinheiten*, welche in der Regel Professuren oder Arbeitsgruppen innerhalb dieser Einrichtungen darstellten. Die konkrete Definition der zu bewertenden For-

2 Die zusammenfassende Darstellung folgt Hinz (2015), für weitere Einzelheiten vgl. Steuerungsgruppe (2008).

schungseinheiten überließ man den Forschungseinrichtungen, was zu einer gewissen Uneinheitlichkeit führte.

Die Pilotstudie für die Soziologie stand vor weiteren Herausforderungen. Im Fach Soziologie ist die Veröffentlichungskultur bekanntlich heterogen. Vor allem jüngere und international orientierte Wissenschaftler/innen versuchen, ihre Forschungsergebnisse in Zeitschriften mit *peer review* Verfahren zu publizieren. Daneben sind jedoch auch weiterhin Buchpublikationen oder Beiträge für thematisch orientierte Sammelbände relevant. Gängige Datenbanken für Publikationen, etwa das *Web of Science* oder *Scopus*, basieren nur auf Zeitschriften, und hiervon auch nur auf einer Auswahl. Alle führenden Zeitschriften der Soziologie in Deutschland werden allerdings im *Social Sciences Citation Index* (SSCI) erfasst; damit waren die Veröffentlichungen in diesen Zeitschriften Teil der Datengrundlage des Forschungsratings. Um die Arbeit der Bewertungsgruppe vorzubereiten, wurde eine Literaturdatenbank erstellt, die allerdings deutliche Mängel aufwies: Die Einträge in den Publikationsdatenbanken der Sozialwissenschaften deckten nur einen Teil aller Veröffentlichungen ab und enthielten überdies auch Beiträge, die dem Kriterium eines ernsthaften *peer review* nicht oder nur partiell genügten. So war als empirische Grundlage für das Forschungsrating eine ergänzende Erfassung derjenigen Publikationen nötig, die über die in den Datenbanken enthaltenen Zeitschriftenbeiträge hinauswiesen. Dazu gab die Bewertungsgruppe allen begutachteten Einheiten die Möglichkeit, die vom WR vorbereitete Liste mit den in Datenbanken recherchierten Publikationen eigenhändig zu ergänzen. Von Zitationsanalysen wurde damals abgesehen. Für Bücher und Beiträge in Sammelbänden wären solche Zitationsanalysen möglich, soweit sie in Zeitschriften zitiert werden. Dies ist allerdings mit Zusatzaufwand verbunden. Das Verfahren des informierten *peer review* sollte sich jedenfalls nicht auf die Beurteilung von Datenbankeinträgen beschränken: Jede Forschungseinheit sollte (größenabhängig) eine gewisse Anzahl von Veröffentlichungen an die Bewertungsgruppe einsenden, die dann von den zugewiesenen Fachgutachter/innen gelesen wurden.

Insgesamt waren 16 Fachgutachter/innen zwischen Februar 2006 bis März 2008 mit der Evaluation beschäftigt. Es waren insgesamt 254 Forschungseinheiten in 57 Forschungseinrichtungen zu begutachten. Für die wissenschaftsimmanenten Dimensionen wurde eine fünfstufige Ordinalskala eingesetzt (mit den Kategorien „exzellent", „sehr gut", „gut", „befriedigend" und „nicht befriedigend"). Jedes Mitglied der Bewertungsgruppe hatte bis zu 80 Einzelbewertungen von Forschungseinheiten abzugeben. Die Gruppe traf sich zu elf Sitzungen an insgesamt 15 Sitzungstagen. Insgesamt wurde pro Mitglied der Bewertungsgruppe eine Arbeitslast von zwei bis drei Arbeitsmonaten angegeben (Neidhardt 2008).

Nachfolgend werden für 229 Forschungseinheiten an 53 Universitäten die Evaluationen der Bewertungsgruppe nachvollzogen. Wir konzentrieren uns auf die Bewertung der Forschungsqualität, die auf der Ebene der Forschungseinheiten erfolgte. Damit unterscheidet sich die vorliegende Auswertung auch von der auf die Forschungseinrichtungen (Universitäten) bezogenen kritischen Re-Analyse von Riordan et al. (2011). Wir verwenden die beim WR für Sekundäranalysen erhältlichen Daten und haben außeruniversitäre Forschungsinstitute (das DIW mit dem SOEP, das Max-Planck-Institut für Gesellschaftsforschung und das Wissenschaftszentrum Berlin) sowie eine Einrichtung, die soziologische Forschung eher randständig betrieb (Deutsche Sporthochschule in Köln), ausgeschlossen. Einige Forschungseinheiten wurden nicht bewertet oder es konnten keine gültigen Angaben bei erklärenden Variablen ermittelt werden, daher verringert sich die Fallzahl in der Analyse von 257 auf 229 Forschungseinheiten. Zusätzlich zu den WR-Daten haben wir im Jahr 2012 Zitationsanalysen zu den sechs Jahre zuvor zur Lektüre der Bewertungsgruppe eingesandten Schriften vorgenommen. Dazu wurden für alle identifizierbaren Dokumente, auch für Buchveröffentlichungen und Beiträge zu Sammelbänden, die Zitationen in einschlägigen Daten-

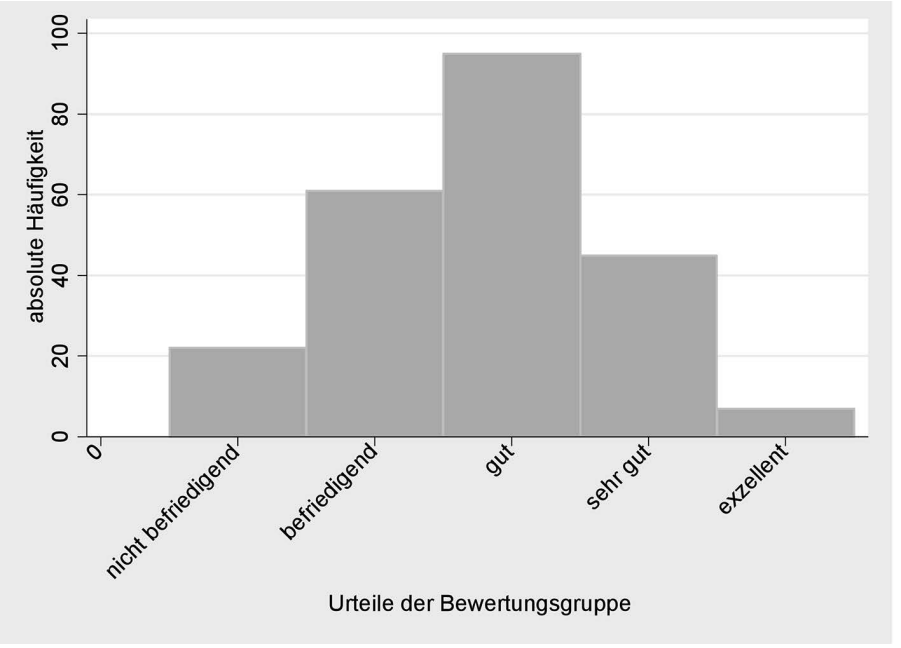
banken wie dem *Web of Science* recherchiert. Dies ist mit einer weiteren, geringfügigen Reduktion der analysierbaren Fälle verbunden (227 statt 229).

3. Input-Output-Analysen

Die Diskussion um wissenschaftlichen Impact, so wie ihn der WR definierte, ist vom Verhältnis von Input- zu Outgrößen bestimmt. Mit anderen Worten: Die Bewertung wissenschaftlicher Leistungen muss sinnvollerweise das Verhältnis des Outputs zum gegebenen Input berücksichtigen. Der Input wird maßgeblich durch die zur Verfügung stehenden Ressourcen geprägt, welche sich durch Grundmittel und Drittmittel zusammensetzen. Im Zuge der rückläufigen Grundfinanzierung und einer damit einhergehenden stärkeren Drittmittelorientierung der Universitäten ist insbesondere die Rolle der eingeworbenen Drittmittel genauer zu thematisieren. Drittmittel, die durch ein *peer review* Verfahren zugeteilt wurden, unterliegen zwar einer Qualitätsprüfung, sagen aber per se noch wenig über die Forschungsqualität aus. Erst wenn sich drittmittelgeförderte Projekte auch in Output, also etwa in Publikationen, niederschlagen, können sie die Forschungsqualität einer Einrichtung verbessern. Es gibt sogar skeptische Stimmen, die eine erhöhte Drittmittelorientierung mit geringerer Produktivität in der Forschung verbinden (etwa: Münch 2007).

Konkret werden wir als Inputfaktor die Größe der Einheiten beim wissenschaftlichen Personal (in Vollzeitäquivalenten) berücksichtigen. In den Unterlagen, die der Bewertungsgruppe zur Verfügung standen, hieß es: „Die Bewertung ist nicht von der Größe abhängig, d.h. kleine Einheiten können die höchste Evaluationsstufe erreichen." Aber auch: „Bei der Bewertung der Forschungsqualität sollte die Größe der Einheit berücksichtigt werden." Zusammengenommen stellte dies einen Hinweis dar, dass die Forschungsqualität nur unter Berücksichtigung der Größe – und damit der zur Verfügung stehenden Ressourcen – angemessen beurteilt werden kann. Bei der Berücksichtigung der Größe kann man verschiedene Maße verwenden: Personalressourcen (in Vollzeitäquivalenten) mit oder ohne Drittmittelprojekte. Bei Personen, die auch Lehrverpflichtungen haben, sind schließlich nur Stellenanteile zur Berechnung der Größe heranzuziehen. Wir verwenden nachfolgend die Personalressourcen der Grundausstattung (ohne Drittmittelstellen) und unter Berücksichtigung eines Lehranteils bei Professor/innen und Mitarbeiter/innen von 50%. Die Personalausstattung hat eine Spannweite von 0,5 bis 11,5 Vollzeitäquivalenten. Der Median liegt bei 1,75 Stellen.

Abbildung 1: Urteile der Bewertungsgruppe



Quelle: Daten des WR, *N* =229

Als zweite Variable für den Input findet die Anzahl der von 2001 bis 2005 erfolgreich eingeworbenen DFG-Projekte Eingang in die Analyse. Sie korreliert stark mit der Anzahl aller wissenschaftsgesteuerten Drittmittelprojekte, dürfte aber das Potential für grundlagenorientierte Forschung besonders trefflich abbilden. Hier reicht die Spannweite von 0 bis 11. Der Medianwert beläuft sich auf ein DFG-Projekt in fünf Jahren. Als Messung für den Output findet die Anzahl von Aufsatzpublikationen mit *peer review* Verwendung. Es wurde bereits auf die möglichen Unschärfen dieser Maßgröße hingewiesen, insbesondere dürften einige der gezählten Aufsätze nicht wirklich durch einen *peer review* Prozess begutachtet worden sein. Erwartungsgemäß ist bei dieser Variable die Spannweite am größten: Sie reicht von 0 bis 62, wobei der Median (wiederum für den Fünfjahreszeitraum) bei drei Aufsätzen in Zeitschriften mit *peer review* liegt. Die Verteilung ist rechtsschief. Zu beachten ist, dass diese Werte nicht größenstandardisiert sind. Die Analysen zum Zusammenhang von Output und Input sind der Gegenstand der folgenden Auswertungen. Zu rekonstruieren gilt es dann wie gesagt die Urteile der Bewertungsgruppe. Die Forschungsqualität wurde wie erwähnt in fünf Stufen erfasst: „exzellent“ (5), „sehr gut“ (4), „gut“ (3), „befriedigend“ (2) und „nicht befriedigend“ (1).

Sieben Forschungseinheiten wurden als „exzellent“ bewertet (3,0 Prozent; s. Abbildung 1). Die Bewertung „sehr gut“ wurde 45mal vergeben (19,6 Prozent). 95 Forschungseinheiten erhielten das Urteil „gut“ (41,3 Prozent), 61 Forschungseinheiten wurden als „befriedigend“ eingeschätzt (26,5 Prozent). Immerhin 22 Forschungseinheiten waren aus Sicht der Bewertungsgruppe „nicht befriedigend“ (9,6 Prozent).

Wir gehen nun in drei Schritten vor, um eine einfache Input-Output-Analyse durchzuführen: (1) Zunächst betrachten wir den Zusammenhang von Personalressourcen in Vollzeit-

äquivalenten (Grundausstattung, Input) mit der Anzahl erfolgreich eingeworbener Drittmittelprojekte bei der Deutschen Forschungsgemeinschaft (DFG). (2) Die Anzahl der Aufsatzpublikationen (Output), wie sie von der WR Bewertungsgruppe gezählt wurde, wird mit beiden Inputfaktoren (Personalressourcen *und* DFG-Drittmittel) korreliert. (3) Schließlich betrachten wir den Zusammenhang des Urteils der Bewertungsgruppe mit allen drei Faktoren.

Tabelle 1: Input-Output Analysen (OLS Regressionen)

	(1) DFG-Projekte	(2) Aufsätze	(3) Urteil
Personal	0,535***	1,122**	-0,0813*
	(0,0876)	(0,386)	(0,0329)
DFG-Projekte		1,730**	0,0939
		(0,633)	(0,0561)
Aufsätze			0,0787***
			(0,0142)
Konstante	-0,0888	0,831	2,454***
	(0,172)	(0,837)	(0,0983)
N	229	229	229
R^2	0,404	0,348	0,387

Robuste Standardfehler in Klammern.
* $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$.

Wie zu erwarten (Tabelle 1), zeigt sich ein deutlicher Zusammenhang zwischen Grundausstattung und eingeworbenen DFG-Projekten.[3] Je höher die Grundausstattung der Forschungseinheiten (Personal ohne Drittmittelstellen), desto größer die Anzahl der eingeworbenen DFG-Projekte. Beide Input-Faktoren korrelieren positiv mit der Anzahl der wissenschaftlichen Aufsätze. Das Urteil der Bewertungsgruppe zur Forschungsqualität der Einheiten wird dann vornehmlich von der Anzahl der wissenschaftlichen Aufsätze bestimmt. Bei Berücksichtigung von Output und DFG-Drittmitteln findet sich sogar ein negativer Zusammenhang zwischen Personalressourcen und dem Urteil der Bewertungsgruppe. Dies ist ein Hinweis darauf, dass die Urteile der Bewertungsgruppe nicht völlig unabhängig von der Größe der Einheiten erfolgten.

Für die Forschungseinheiten in der deutschen Soziologie zeigen sich also insgesamt und nicht überraschend folgende Zusammenhänge: Die Anzahl der DFG-Projekte ist von der Größe einer Einheit, gemessen an den Personalressourcen, abhängig. Der Input in Form von Forschungsmitteln wiederum beeinflusst den Output, d.h. die Anzahl begutachteter Fachartikel. Diese haben schließlich einen positiven Einfluss auf die Urteile der Bewertungsgruppe. Interessant und nicht ganz selbstverständlich ist, dass die Bewertungsgruppe die Größe einer

3 Aufgrund der hierarchischen Datenstruktur (mehrere Einheiten gehören zu einer Einrichtung) werden in allen Regressionen auf Ebene der Einheiten geclusterte Standardfehler verwendet (Rogers 1993). Ähnliche Regressionsschätzungen wie Modell 3 wurden bereits von Rainer Lange (WR) durchgeführt und den Mitgliedern der Bewertungsgruppe vorgelegt.

Einheit bei der Urteilsbildung berücksichtigt und das Urteil für die Größe der Einheit quasi diskontiert.

4. Rekonstruktion der Urteile der Bewertungsgruppe zur Forschungsqualität

Der Bewertungsgruppe standen – wie geschildert – systematisch aufbereitete Informationen zu den wissenschaftlichen Leistungen der zu bewertenden Einrichtungen zur Verfügung. Für die Bewertung der Forschungsqualität sollten die Selbstberichte der Einheiten (mit Stärken-Schwächen-Analysen), Ergebnisse der Literaturrecherche, die Aktivität bei der Einwerbung von Drittmitteln und die Leseeindrücke in das gemeinsame Qualitätsurteil einfließen. Weiterhin sollten vorgelagerte Qualitätsbewertungen von *peers* einfließen. Dies sind etwa die Anzahl an Artikeln in Zeitschriften mit *peer review*, die Anzahl der Projekte, welche ein Begutachtungsverfahren durchlaufen haben, und sonstige Drittmittelprojekte. Die höchste Bewertungsstufe („exzellent") sollte nur dann erreicht werden, wenn die Forschungseinheit nach internationalen Maßstäben zu den führenden Forschungseinheiten gehört.

Das Forschungsrating zielte also auf eine *absolute* Einschätzung der Forschungsqualität der Forschungseinheiten ab, ein vergleichendes Ranking war explizit nicht beabsichtigt. Allerdings bestand ein Dilemma des Forschungsratings – wie bei jeder zusammenfassenden Bewertung absoluter Kriterien – darin, insbesondere die herausragenden und stark abfallenden Leistungen und Qualitäten nur im Vergleich unterschiedlicher Leistungsträger ermitteln zu können. Daher gab es gewisse Daumenregeln zur Bewertung der Anzahl an Publikationen. Die Mitglieder der Bewertungsgruppe sollten ihr Urteil auf quantitative Indikatoren stützen (Perzentile bei der Anzahl der Artikel mit *peer review* sowie bei den wissenschaftsgestützten Drittmittelprojekten). Die Daumenregel lautete: Die Perzentilwerte 100-90 entsprechen der Bewertung „exzellent", 90-70 sind „sehr gut", 70-30 „gut", 30-10 „befriedigend" und Werte im Bereich 10-0 „nicht befriedigend". Abweichungen im Urteil um mehr als eine Kategorie waren möglich, mussten aber eigens begründet werden, mit Bezugnahme auf die gewonnenen Lektüreeindrücke und andere qualitative Aspekte (z.B. die Selbstberichte der Forschungseinheiten).

Zur Rekonstruktion der Qualitätsurteile zur Forschungsqualität der untersuchten Einheiten verwenden wir nachfolgend die Perzentilwerte, welche die Forschungseinrichtungen im Hinblick auf die Anzahl der Aufsätze in Zeitschriften mit *peer review* sowie im Hinblick auf die Anzahl der eingeworbenen DFG-Projekte erreichten. Zunächst geht es um die reine Wirkung dieser quantitativen Indikatoren (s. Tabelle 2).

Tabelle 2 zeigt, dass etwa 60 Prozent der Varianz der Qualitätsurteile auf Unterschiede in den Perzentilwerten für Veröffentlichungen in Zeitschriften mit *peer review* zurückgehen. Vergleichsweise Berechnungen mit Daten des zeitgleich abgelaufenen Evaluationsverfahrens in der Chemie ergeben, dass in der Naturwissenschaft der Wert erklärter Varianz nicht dramatisch höher liegt (69 Prozent) falls lediglich die beschriebenen einfachen Indikatoren herangezogen werden. Zusätzliche (nicht berichtete) Tabellenanalysen führen in der Soziologie nur zu etwa sechs Prozent fehlklassifizierter Forschungseinheiten, was die Reliabilität der Rekonstruktion mit den zwei einfachen Indikatoren unterstreicht. „Fehlklassifiziert" heißt dabei, dass das Urteil der Bewertungsgruppe in allen Fällen als korrekt angenommen wird, was in der Praxis natürlich nicht gegeben sein muss. Das Modell 2 aus Tabelle 2 belegt den positiven Zusammenhang von Drittmittelprojekten und Urteilen der Expertengruppe, wobei allerdings die Varianzerklärung deutlich kleiner ausfällt. Schließlich zeigt sich im Gesamtmodell (Modell 3), dass nur noch der Outputfaktor und nicht mehr der Inputfaktor das Qualitätsurteil bestimmt. Dies entspricht durchaus den Erwartungen, die man vernünftigerweise an Evaluationen wissenschaftlicher Qualität stellen kann. Das Forschungsrating des WR berücksichtigt vor allem den wissenschaftlichen Output.

Tabelle 2: Rekonstruktion der Bewertung

	(1) Urteil	(2) Urteil	(3) Urteil
Perzentil: Aufsätze[a]	0,0233***		0,0221***
	(0,00141)		(0,00163)
Perzentil: DFG-Projekte		0,0117***	0,00242
		(0,00142)	(0,00137)
Konstante	1,723***	2,378***	1,693***
	(0,0688)	(0,0870)	(0,0698)
N	229	229	229
R^2	0,597	0,182	0,603

Robuste Standardfehler in Klammern. [a] Aufsätze mit *peer review*.
* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Wie verändern sich die Zusammenhänge, wenn größenstandardisierte Werte für den Input (DFG-Projekte) und Output (Aufsätze) betrachtet werden? Für die in Tabelle 3 dargestellten Analysen wurden die Perzentilwerte neu berechnet – unter Berücksichtigung der zur Verfügung stehenden gesamten Personalressourcen. Konkret wurden das Verhältnis von Publikationen pro Vollzeitäquivalent und anschließend die Perzentile dieser größengewichteten Variable bestimmt. Genauso wie in den oben angestellten Input-Output-Analysen werden wieder nur die Vollzeitäquivalente ohne Einbezug des Drittmittelpersonals herangezogen.

Tabelle 3: Rekonstruktion der Bewertung (größengewichtet)

	(1) Urteil	(2) Urteil	(3) Urteil
Perzentil: Aufsätze [a] (größengewichtet)	0,0231***		0,0213***
	(0,00125)		(0,00133)
Perzentil: DFG-Projekte (größengewichtet)		0,0106***	0,00501***
		(0,00135)	(0,00113)
Konstante	1,703***	2,382***	1,590***
	(0,0619)	(0,0859)	(0,0633)
N	229	229	229
R^2	0,590	0,169	0.624

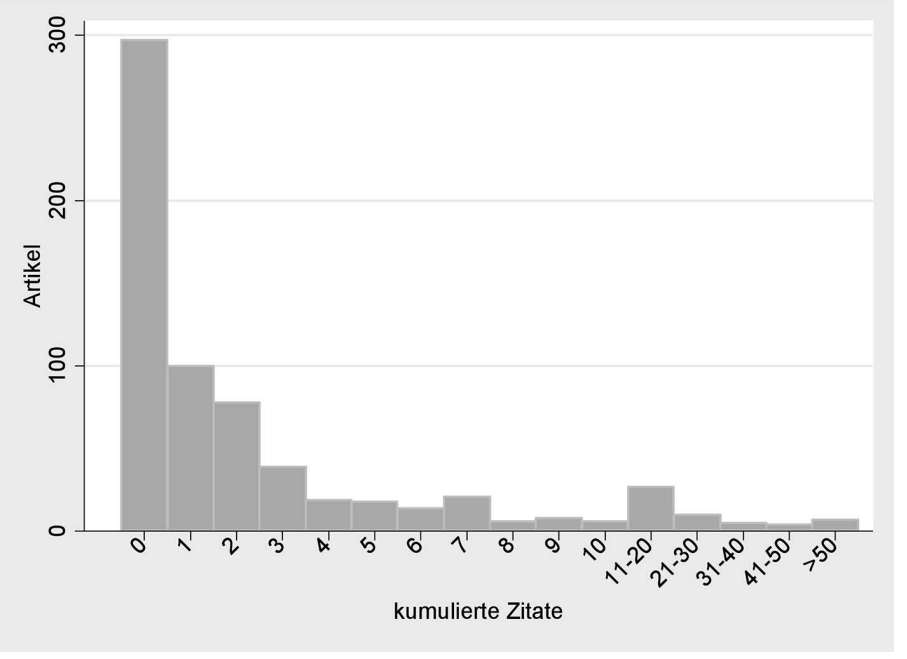
Robuste Standardfehler in Klammern. [a] Aufsätze mit *peer review*.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Nur die Ergebnisse des Gesamtmodells unterscheiden sich substanziell von den Analysen, die in Tabelle 2 vorgestellt wurden. Bei größenstandardisierter Betrachtung beeinflusst die relative Anzahl der DFG Projekte (pro Vollzeitäquivalent) die Einschätzung der For-

schungsqualität deutlich positiv. Offenbar haben die Mitglieder der Bewertungsgruppe ihre Einschätzungen unter Berücksichtigung der Personalressourcen der Forschungseinheiten getroffen. Da die Perzentilwerte beider Maße auf den Wertebereich von 0 bis 100 normiert sind, können die Effektgrößen direkt miteinander verglichen werden. Die relative Publikationsleistung wird etwa als viermal bedeutsamer eingeschätzt als die relative Anzahl von DFG-Drittmittelprojekten. Zu beachten ist schließlich, dass im Gesamtmodell nun sogar 62 Prozent der Varianz der Urteile zur Forschungsqualität erklärt werden. Dabei muss man aber im Auge behalten, dass sich die Mitglieder der Bewertungsgruppe vorgängig geeinigt hatten, die Perzentile der Anzahl begutachteter Fachartikel der Urteilsbildung zugrunde zu legen.

In den Veröffentlichungen des Sprechers der Bewertungsgruppe wird der besondere Einfluss der Leseeindrücke hervorgehoben. Diese Leseeindrücke können naturgemäß nicht nachträglich und extern rekonstruiert werden. Möglich ist allerdings eine Prüfung der prädiktiven Validität der Lektüre durch die Bewertungsgruppe. In den Daten des WR finden sich 659 Datenbankeinträge für die zur Lektüre eingesandten Schriftproben, darunter 317 Artikel (in Zeitschriften), 216 Beiträge in Sammelbänden, 109 Monographien, 16 Herausgeberschaften und ein nicht zuzuordnender Residualfall. Durch sorgfältige Datenbankrecherchen konnten die Zitationen in wissenschaftlichen Datenbanken (*Web of Science*) immer fünf Jahre nach Erscheinen der jeweiligen Publikationen ermittelt werden.

Abbildung 2: Zitationshäufigkeit fünf Jahre nach Erscheinen (ohne Selbstzitate)



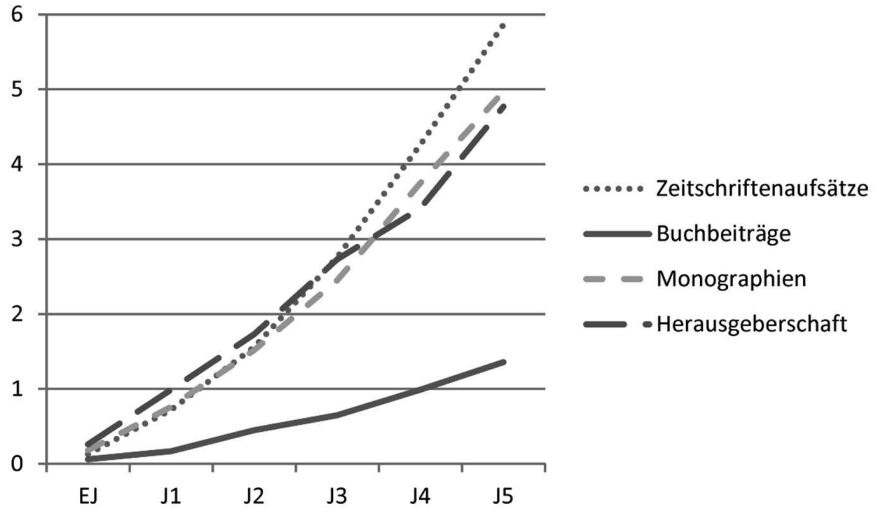
Quelle: WR Daten, eigene Recherchen, *Web of Science*.

Die Mehrheit der eingesandten Schriften wurde gar nicht ($N = 297$; 45,1 Prozent) oder nur ein einziges Mal ($N = 100$; 15,2 Prozent) in fünf Jahren zitiert. Sehr wenige eingesandte Schriften wurden häufiger als 10-mal zitiert ($N = 53$; 8,0 Prozent) und ein verschwindend

kleiner Anteil öfter als 50-mal ($N = 7$; 1,1 Prozent). Insgesamt ergibt sich der Eindruck, dass die eingereichten Arbeiten der Forschungseinheiten überwiegend nur einen sehr geringen Impact in Form von Zitationen durch andere Wissenschaftler/innen aufweisen.

Es ist weiterhin möglich, die durchschnittlichen Zitationshäufigkeiten im Zeitverlauf nach Erscheinen und nach Art der Schrift zu analysieren (vgl. Abbildung 3). Leicht zu erkennen ist, dass sich die Zitationshäufigkeit von Zeitschriftenaufsätzen über den gesamten Beobachtungszeitraum von fünf Jahren am besten entwickelt. In den ersten drei Jahren nach Erscheinen dominieren noch die herausgegebenen Bände. Allerdings sind über den gesamten Zeitraum die Unterschiede der durchschnittlichen Zitationen zwischen Aufsätzen, Monographien und herausgegebenen Sammelwerken eher gering. Nach fünf Jahren (J5) liegen die Durchschnittswerte dieser Publikationsarten zwischen 4,8 und 5,9 Zitationen insgesamt. Dagegen fallen Buchbeiträge in Sammelbänden deutlich ab. Auch fünf Jahre nach Erscheinen werden solche Veröffentlichungen im Durchschnitt nur ein einziges Mal (1,4) zitiert.

Abbildung 3: Durchschnittliche (kumulierte) Zitationen nach Typ der eingereichten Veröffentlichungen im Erscheinungsjahr (EJ) und bis fünf Jahre nach Erscheinen (J1-J5)



Quelle: WR Daten, eigene Recherchen, *Web of Science,* ohne Selbstzitate.

Als Maß für den Zitationsimpact einer Forschungseinheit kann die durchschnittliche Anzahl von Zitationen pro eingereichter Veröffentlichung verwendet werden. Für fast ein Drittel der Forschungseinheiten ist dieser Wert für den Zitationsimpact 0, der Median beträgt 1 und der Maximalwert 57. Solche extrem schiefen Verteilungen sind aus Zitationsanalysen wissenschaftlicher Veröffentlichungen durchaus bekannt (Bornmann et al. 2008). Analog zur Anzahl der Aufsätze und DFG-Projekte kann ein Perzentilwert für den Impact berechnet werden, der in die Regressionsmodelle aufgenommen wird (vgl. Tabelle 4).

Die Zitationen der eingereichten Veröffentlichungen (fünf Jahre nach ihrem Erscheinen) erklären immerhin 26 Prozent der Varianz der Urteile zur Forschungsqualität. Die Zahl der Zitationen zum Zeitpunkt der Begutachtung war den Mitgliedern der Bewertungsgruppe in der Regel nicht bekannt. Dennoch ist die Korrelation zwischen dem Urteil und den Zitationen beträchtlich. Gelesene Literatur mit einem höheren Impact, d.h. Artikel, die sich seit

Erscheinen als einflussreich erwiesen haben oder noch erweisen sollten, beeinflussen das Gutachterurteil positiv und gehen im Durchschnitt mit einer besseren Bewertung einher als selten zitierte Literatur. Fachgutachter erkennen, welcher Artikel oder welches Buch in der Zunft Eindruck machen wird und berücksichtigen dies bei Ihrem Urteil (vgl. Diekmann / Näf / Schubiger 2012). Natürlich ist die Korrelation alles andere als perfekt. Zitationen sind nicht unbedingt ein Maß für wissenschaftliche Qualität. Sie sind eher ein Maß für die Stärke des „Echos", der Aufmerksamkeit, die eine Publikation in der *scientific community* erhält. Auch das Gutachterurteil ist nicht zweifelsfrei ein Maß für die Qualität. Dennoch korrelieren Urteil und nachfolgend ermittelte Zitationsrate relativ stark miteinander (r = 0,51). Stärker noch hängt das Urteil von der Anzahl begutachteter Artikel ab. Durch die Hinzufügung dieses Merkmals in der Urteilsregression steigt die erklärte Varianz auf 0,603 (Modell 2 in Tabelle 4). Beide Werte bleiben signifikant, wenn zusätzlich die Perzentilwerte für DFG-Projekte berücksichtigt werden. Letztere tragen zur Varianzerklärung des Urteils nicht mehr bei.

Tabelle 4: Rekonstruktion der Bewertung mit Zitationen

	(1)	(2)	(3)
	Urteil	Urteil	Urteil
Perzentil: Zitate	0,0148***	0,00484***	0,00454**
	(0,00163)	(0,00135)	(0,00131)
Perzentil: Aufsätze [a]		0,0203***	0,0193***
		(0,00158)	(0,00175)
Perzentil: DFG-Projekte			0,00236
			(0,00133)
Konstante	2,131***	1,638***	1,617***
	(0,0864)	(0,0761)	(0,0782)
N	220	220	220
R^2	0,262	0,603	0,609

Robuste Standardfehler in Klammern. [a] Aufsätze mit *peer review*.
* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Die gelesene Literatur wird also im Hinblick auf ihren zukünftigen Impact im Mittel nicht vollkommen falsch eingeschätzt, gleichwohl besteht diesbezüglich nur eine mäßige prädiktive Validität. Allerdings erkennt man in Modell 2 (Tabelle 4), dass die Anzahlen der Aufsatzpublikationen wesentlich bedeutsamer sind als ihre Zitationen. Die Qualität der Schriften, welche im Rating durch die Leseeindrücke gemessen wurde, wird hier also alternativ über die Hilfskonstruktion der Zitationen in die Untersuchung einbezogen. Mit dem zentralen Ergebnis, dass die relative Zitationsrate als Erklärungsfaktor wichtiger als die Anzahl von DFG-Projekten ist, jedoch die Bewertung der Forschungsleistung immer noch wesentlich von der reinen Anzahl an Publikationen dominiert wird. Dies repliziert frühere Ergebnisse, dass für die Evaluation von Forschungsleistungen die Quantität an *peer reviewed* Publikationen weitaus zentraler ist als ihr Impact (s. z.B. für Berufungschancen von Wissenschaftler/innen auf Professuren: Long et al. 1993). Die Berechnungen demonstrieren zudem, dass Zitationsanalysen nicht nur für Zeitschriftenartikel sinnvoll sind. Auch bei anderen

Publikationsformaten, insbesondere für Bücher, lässt sich der Einfluss auf die publizierte Fachliteratur mittels Zitationsanalysen durchaus belegen.

5. Schlussfolgerungen

Das Forschungsrating Soziologie des Wissenschaftsrats war ein ambitioniertes Pilotprojekt. Der Vorsitzende Friedhelm Neidhardt (2006, 2008, 2009) und die Mitglieder der Bewertungsgruppe haben mit großem Arbeitsaufwand und Engagement erstmalig und systematisch versucht, umfangreiche Daten über die soziologische Forschung an deutschen akademischen Institutionen zu gewinnen und die beteiligten 254 Forschungseinheiten und 57 Einrichtungen durch einen Indikatoren gestützten *informed peer-review* zu bewerten. Eine Besonderheit der Evaluation war, dass die Forschungseinheiten ausgewählte Publikationen einreichen konnten, die von den Gutachtern gelesen und bewertet wurden. In einer einmaligen Anstrengung wurden so von den Juroren 666 eingereichte Publikationen gelesen und bei der Bewertung der Forschungsleistung berücksichtigt. Im Ergebnis liegt nicht nur ein Bericht des WR zur Forschungslage der Soziologie, sondern auch ein Datenbestand vor, der über zahlreiche forschungsbezogene Merkmale der Einheiten informiert und dessen Analyse für empirische, wissenschaftssoziologische Untersuchungen von hohem Interesse ist. Überdies wurden durch das Pilotprojekt Erfahrungen gewonnen, die von der Auswahl der Einrichtungen und der Diskussion der Bewertungsskala über die Bestimmung der Indikatoren und ihrer konkreten Operationalisierung mittels Perzentilen bis hin zur Aggregation der Informationen zu einem Urteil reichten.

Die von der Bewertungsgruppe und dem WR erhobenen Daten haben wir einer Sekundäranalyse unterzogen. Es zeigt sich, dass die Forschungsleistung der Forschungseinheiten (mit im Mittel etwa 2,3 Vollzeitäquivalenten) durch bibliometrische Verfahren, und somit deutlich einfacher zu erhebenden Daten, bereits recht gut bestimmt werden kann. Neben der bloßen größengewichteten Anzahl von Veröffentlichungen konnten hier auch die Zitationen für die eingesandten und gelesenen Forschungsbeispiele ermittelt werden. Davon wurde in der Pilotstudie selbst abgesehen – aus damals nachvollziehbaren Gründen. Allerdings könnte die Forschungsqualität der Einheiten durch eine gemeinsame Bewertung von größenstandardisiertem Publikationsoutput und dessen Einfluss noch adäquater beschrieben werden. Eine weitere Verbesserung ist zu erhoffen, wenn der von der Bewertungsgruppe beklagte Missstand einer in Teilen lückenhaften und unzuverlässigen Publikationsdatenbank behoben wird.

Das Forschungsrating der Soziologie wurde im Fach überwiegend skeptisch aufgenommen (Baier 2011; Münch 2009). Zwar wurde im Vergleich zu (noch) stärker umstrittenen Erhebungen (etwa das sog. CHE Ranking) die sorgfältige und sehr transparente Durchführung gewürdigt. Gleichzeitig wurde besondere Kritik an der faktischen Eindimensionalität geübt. Das Forschungsrating stelle eine einseitige und auf die Publikation in Fachzeitschriften beschränkte Konstruktion von soziologischer Exzellenz dar und schaffe erst eine Stratifikation in einem zuvor durch gleichberechtigte Diskursteilnehmer/innen gekennzeichneten Feld (Münch / Baier 2009). Eine einmal etablierte Stratifikation tendiere dann zur Selbstreproduktion, weil sich die Akteure zukünftig an dieser Konstruktion orientierten. Diese Kritik müsste unserer Meinung nach – wenn sie denn überhaupt einen wahren Kern besitzt (was wir an dieser Stelle nicht diskutieren können) – auf das Wissenschaftssystem insgesamt bezogen werden, in dem sich die sog. Matthäus-Effekte nicht erst seit dem Forschungsrating des WR zeigen (s. z.B. Allison et al. 1982; Merton 1968, 1988). Die Eindimensionalität des Forschungsratings erscheint letztlich im Hinblick auf das angestrebte Ziel des WR Ratings, eine vergleichbare Einschätzung von Forschungsqualität zu erreichen, sogar sinnvoll.

Der WR selbst gelangte übrigens – wenig überraschend – zu einer positiven Gesamtbilanz. Trotz des großen Aufwands wird eine Fortführung und breitere Anwendung des Forschungsratings empfohlen (Wissenschaftsrat 2013). Gelobt wird am eigens entwickelten Verfahren die Verbindung aus hoher methodischer Qualität, fachspezifischer Anpassungsfähigkeit und potenziell breitem Nutzen. Wenn damit die Empfehlung zu einer unveränderten Wiederholung verbunden ist, können wir uns dieser Einschätzung allerdings nicht ganz anschließen. Aufgrund des unbestritten enormen Aufwands der Bewertungsgruppe, insbesondere auch des großen zeitlichen Aufwands für das Lesen der eingereichten Literatur, dürfte das Forschungsrating in der durchgeführten Form kaum wiederholbar sein.

Die Reaktion der Universitäten auf die Ergebnisse des Forschungsratings in der Soziologie war zudem geringer, als der WR vermutete. Es ist uns nicht bekannt, dass nach Vorstellung der Ergebnisse eine Diskussion über „nicht befriedigend" oder „befriedigend" evaluierte Einheiten stattgefunden hätte, geschweige denn, dass Universitäten oder Ministerien irgendwelche Konsequenzen gezogen hätten. Immerhin 60 Forschungseinheiten hatten ja in fünf Jahren keinen einzigen Artikel in einer referierten Zeitschrift veröffentlicht, fast ebenso viele nicht ein einziges wissenschaftsgesteuertes Drittmittelprojekt eingeworben.

Sollte der WR die Forschungsratings wiederholen wollen, erschiene uns im Sinne eines wirklichen Mehrwertes an Erkenntnissen zur effektiven Ausgestaltung von Forschungsinstitutionen eine Orientierung an den größeren Forschungseinrichtungen zielführender (vgl. auch Diekmann 2007; Hinz 2015). Zudem müsste eine wirklich verlässliche Publikationsdatenbank geschaffen werden, wie es auch von der Bewertungsgruppe gefordert wurde. Dass in der Soziologie Forschungsergebnisse nicht nur in Form von Artikeln in begutachteten Journalen, sondern auch als Monographien veröffentlicht werden, sollte nicht aus dem Auge verloren werden. Auf der Ebene der Einrichtungen könnte auf Grundlage zunächst standardisierter Daten eine informierte Diskussion unter *peers* zur Forschungsausrichtung, einer erkennbaren Profilierung, zu genutzten bzw. ungenutzten Synergien und zur Einbindung in internationale Forschungszusammenhänge erfolgen. All dies macht nur auf der Ebene von Forschungseinrichtungen Sinn. Hinzu käme eine enorme Vereinfachung des Verfahrens, da anstelle von 254 Forschungseinheiten nur noch 57 Einrichtungen (Institute) zu bewerten wären. Die *peers* wären mit ihrer Expertise dann effizienter eingesetzt, als wenn sie die Veröffentlichungen von Kolleg/innen (im Fall von *peer review* gestützten Zeitschriften: *nochmals*, im Fall mancher Sammelwerke vielleicht *erstmals*) lesen und bewerten müssten. Das „qualitative" Element des Lesens von Veröffentlichungen war ein besonderes Merkmal des Verfahrens, könnte dann aber auf Konfliktfälle begrenzt werden. Auch die sonstigen Leistungsfaktoren – etwa die Einwerbung von wissenschaftsgesteuerten Drittmitteln – wurden in der Regel bereits vor dem Forschungsrating durch *peers* evaluiert. Aber was bislang fehlt, ist eine Gesamtschau dazu, wie die einzelnen Forschungseinheiten zusammenpassen und wie zielführend und kreativ sie zusammenarbeiten. Dazu könnte der Sachverstand eines informierten *peer review* bei einer eventuellen Neuauflage des Forschungsratings gewinnbringend genutzt werden.

Literatur

Allison P.D. / J.S. Long / T.K. Krauze (1982): Cumulative Advantage and Inequality in Science, in: American Sociological Review 47, S. 615-625.

Baier, C. (2011): Wissenschaft regieren. Eine diskursanalytische Studie zum Forschungsrating des Wissenschaftsrates, in: Soziale Welt 62, S. 371-390.

Bornmann, L. / H.P. Daniel (2003): Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zu Reliabilität, Fairness und Validität des Peer-Review-Verfahrens, in: S. Schwarz / U. Teichler (Hrsg.), Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung, Frankfurt / Main, S. 207-225.

Bornmann, L. / R. Mutz / C. Neuhaus / H.D. Daniel (2008): Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results, in: Ethics in Science and Environmental Politics 8, S. 93-102.

Diekmann, A. (2007): Vorschlag für die Vereinfachung und Verbesserung der Evaluation – Lehren aus dem Pilotprojekt, Rundschreiben an die Mitglieder der Bewertungsgruppe, ETH-Zürich.

Diekmann, A. / M. Näf / M. Schubiger (2012): Die Rezeption (Thyssen-)preisgekrönter Artikel in der „Scientific Community", in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 64, S. 563-581.

Hinz, T. (2015): Weitgehend überraschungsfrei, folgenlos und (so) nicht wiederholbar: Das Forschungsrating der Soziologie in Deutschland, in: Die Evaluation der Soziologie – Kritik und Perspektiven, Bulletin der Schweizerischen Gesellschaft für Soziologie 147/148, S. 16-21.

Hirschauer, S. (2004): Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation, in: Zeitschrift für Soziologie 33, S. 62-83.

Long, J.S. / P.D. Allison / R. McGinnis (1993): Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity, in: American Sociological Review 58, S. 703-722.

Merton, R.K. (1968): The Matthew Effect in Science, in: Science 159, S. 56-63.

Merton, R.K. (1988): The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property, in: ISIS 79, S. 606-623.

Münch, R. (2007): Die akademische Elite. Zur sozialen Konstruktion wissenschaftlicher Exzellenz, Frankfurt / Main.

Münch, R. (2009): Die Konstruktion soziologischer Exzellenz durch Forschungsrating, in: Soziale Welt 60, S. 63-89.

Münch, R. / C. Baier (2009): Die Konstruktion der soziologischen Realität durch Forschungsrating, in: Berliner Journal für Soziologie 19, S. 295-319.

Neidhardt, F. (2006): Forschungsrating der deutschen Soziologie durch den Wissenschaftsrat, in: Soziologie 35, S. 303-308.

Neidhardt, F. (2008): Das Forschungsrating des Wissenschaftsrates. Einige Erfahrungen und Befunde, in: Soziologie 37, S. 421-432.

Neidhardt, F. (2009): Stärken und Schwächen der Soziologie in Deutschland, in: Soziologie 38, S. 40-48.

Riordan, P. / C. Ganser / T. Wolbring (2011): Zur Messung von Forschungsqualität. Eine kritische Analyse des Forschungsratings des Wissenschaftsrats, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 63, S. 147-172.

Rogers, W.H. (1993): Regression Standard Errors in Clustered Samples, in: Stata Technical Bulletin 3, S. 19-23.

Steuerungsgruppe (2008): Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Soziologie. Steuerungsgruppe der Pilotstudie Forschungsrating im Auftrag des Wissenschaftsrates, online abrufbar unter: http://www.wissenschaftsrat.de/download/Forschungsrating/ Dokumente/Pilotstude_Forschungsrating_Soziologie/pilot_ergeb_sozio.pdf, letztes Abrufdatum: 20.6.2015.

Wissenschaftsrat (2006): Bewertungsmatrix Soziologie, 14.10.2006, online abrufbar unter: http://www.wissen-schaftsrat.de/download/Forschungsrating/Dokumente/Pilotstude_Forschungsrating_Soziologie/Bewertungsmatrix_Soz.pdf, letztes Abrufdatum: 20.6.2015.

Wissenschaftsrat (2010): Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften, Drs 10039-10, online abrufbar unter: http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf, letztes Abrufdatum: 20.6.2015.

Wissenschaftsrat (2013): Empfehlungen zur Zukunft des Forschungsratings, Drs 3409-13, online abrufbar unter: http://www.wissenschaftsrat.de/download/archiv/3409-13.pdf, letztes Abrufdatum: 20.6.2015.

Prof. Katrin Auspurg
Goethe-Universität Frankfurt a.M.
Institut für Soziologie
Theodor-W.-Adorno-Platz 6
D-60323 Frankfurt am Main
Auspurg@soz.uni-frankfurt.de

Prof. Andreas Diekmann
Matthias Näf
ETH Zürich
Professur für Soziologie
Clausiusstrasse 50
CH-8092 Zürich
andreas.diekmann@soz.gess.ethz.ch
matthias.naef@soz.gess.ethz.ch

Prof. Thomas Hinz
Universität Konstanz
Fachbereich Geschichte und Soziologie
Universitätsstr. 10
D-78464 Konstanz
Thomas.Hinz@uni-konstanz.de