

**Pilotstudie Forschungsrating Soziologie  
Abschlussbericht der Bewertungsgruppe**

<u>Inhalt</u>	<u>Seite</u>
Vorbemerkung .....	3
Kurzfassung.....	5
A. Ausgangslage.....	9
A.I. Ablauf der Pilotstudie .....	9
A.II. Operationalisierung des Forschungsratings für das Fach Soziologie .....	10
II.1. Definition des Fachs Soziologie .....	11
II.2. Definition der Kriterien und Skalen .....	12
II.3. Zuordnung von Indikatoren.....	13
II.4. Definition der Forschungseinheiten .....	14
II.5. Entwicklung von Fragebögen .....	15
II.6. Rahmenbedingungen der Publikations- und Zitationsanalyse.....	16
A.III. Datenerhebung .....	16
III.1. Organisation und Ablauf .....	16
III.2. Erfassung der Forschungseinheiten .....	17
III.3. Datenerhebung.....	19
III.4. Publikationserhebung und bibliometrische Analyse.....	20
III.5. Aufbereitung der Daten.....	23
A.IV. Bewertungsphase .....	24
IV.1. Organisation und Ablauf .....	24
IV.2. Analyse der Bewertungsergebnisse .....	28
A.V. Veröffentlichung und Reaktionen auf das Forschungsrating.....	31
V.1. Veröffentlichung der Ergebnisse.....	31
V.2. Reaktionen auf das Forschungsrating .....	32
B. Empfehlungen .....	35
B.I. Organisation und Ablauf .....	35
B.II. Zu Einzelaspekten des Forschungsratings .....	37

II.1. Zur fachspezifischen Operationalisierung des Verfahrens .....	37
II.2. Zur Erfassung der Forschungseinheiten.....	40
II.3. Zur Datenerhebung .....	42
II.4. Zur Publikationserhebung und bibliometrischen Analyse .....	45
II.5. Zur Datenaufbereitung.....	48
II.6. Zur Bewertungsphase .....	50
II.7. Zur Veröffentlichung .....	52
B.III. Zum weiteren Vorgehen.....	53
III.1. Aufwand und Ertragserwägungen .....	53
III.2. Qualitätsbedingungen des Forschungsratings.....	57
Anhang .....	61

## Vorbemerkung

Der Wissenschaftsrat hat im Juli 2005 beschlossen, das von ihm in seinen „Empfehlungen zu Rankings im Wissenschaftssystem“ vom November 2004<sup>1</sup> entworfene Verfahren eines Forschungsratings in einer Pilotstudie in den Fächern Chemie und Soziologie zu erproben. Die vom Wissenschaftsrat mit der Durchführung der Pilotstudie beauftragte Steuerungsgruppe hat auf Basis von Vorschlägen der großen Wissenschaftsorganisationen und der Deutschen Gesellschaft für Soziologie die Mitglieder einer Bewertungsgruppe für die Soziologie berufen. Diese Bewertungsgruppe hat das Verfahren an die Bedürfnisse der Soziologie angepasst und auf Basis der erhobenen Daten die Forschungsleistungen deutscher Universitäten und außeruniversitärer Forschungseinrichtungen in der Soziologie vergleichend bewertet. Die Ergebnisse dieser Bewertungen sind von der Steuerungsgruppe im April 2008 veröffentlicht worden.<sup>2</sup>

Im vorliegenden Abschlussbericht fasst die Bewertungsgruppe Soziologie ihre Erfahrungen aus der Pilotstudie zusammen und gibt Empfehlungen dazu ab, wie das Verfahren optimiert werden sollte, falls der Beschluss gefällt wird, es nach einigen Jahren erneut, durchzuführen.

Der vorliegende Abschlussbericht wurde von der Bewertungsgruppe Soziologie (Mitgliederliste s. Anhang 3) am 15. Februar 2008 verabschiedet.

---

<sup>1</sup> Wissenschaftsrat: „Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung.“ in Empfehlungen und Stellungnahmen 2004, Bd. I. Köln 2005, S. 159-220.

<sup>2</sup> Steuerungsgruppe der Pilotstudie Forschungsrating im Auftrag des Wissenschaftsrates: Forschungsleistungen deutscher Universitäten und außeruniversitärer Forschungseinrichtungen in der Soziologie. Ergebnisse der Pilotstudie Forschungsrating des Wissenschaftsrates, Köln 18.4.2008.



## **Kurzfassung**

Der Wissenschaftsrat hat im Juli 2005 beschlossen, das von ihm im November 2004 empfohlene Verfahren zu einem Forschungsrating zunächst in einer Pilotstudie in zwei Fächern, Chemie und Soziologie, zu erproben. Ziel der Pilotstudie war es auszuloten, ob und in welcher Weise der in den Empfehlungen vorgegebene Rahmen eine Operationalisierung für verschiedene Fächer zulässt und dann praktisch umsetzbar ist. Die für die fachliche Umsetzung in der Soziologie eingesetzte Bewertungsgruppe hat zunächst das Verfahren für ihr Fach angepasst und dann die vergleichende Bewertung der Forschungsleistung deutscher Universitäten und außer-universitärer Einrichtungen vorgenommen, deren Ergebnisse am 18. April 2008 veröffentlicht wurden.

Die Bewertungsgruppe Soziologie kommt zu dem Ergebnis, dass das gewählte Verfahren einer Gutachterbewertung auf der Basis zahlreicher quantitativer und qualitativer Indikatoren und Kontextinformationen einfacheren Verfahren, etwa einer rein statistischen Indikatorenauswertung, eindeutig vorzuziehen ist. Der höhere Aufwand erscheint durch den Ertrag einer qualifizierteren Bewertung gerechtfertigt. Dieser Aufwand lässt sich bei Folgeprojekten gegenüber der Pilotstudie zwar reduzieren. Aber auch dann werden akzeptable Forschungsratings einen beträchtlichen Einsatz sowohl auf der Seite der Forschungseinrichtungen als auch bei denen mit sich bringen, die ihre Leistungen korrekt und mit förderlichen Effekten für die Wissenschaft feststellen und bewerten sollen. Gute Forschungsratings sind nicht billig zu haben. Erst eine umfassende Zusammenstellung quantitativer und qualitativer Aspekte und deren Abwägung durch Fachgutachter ermöglichen eine fundierte Bewertung. Diese sollte aber nicht unmittelbar an staatliche Mittelzuweisungen gekoppelt werden, sondern den betroffenen Einrichtungen in Verbindung mit anderen Instrumenten der Qualitätssicherung als Grundlage für ihre strategischen Planungen dienen.

Um eine Verlässlichkeit der Bewertung zu gewährleisten und zugleich den Aufwand, wo immer möglich, begrenzt zu halten, macht die Bewertungsgruppe eine Reihe von Vorschlägen für den Fall, dass die Soziologie in einigen Jahren erneut bewertet werden soll. Sie werden im Projektbericht differenzierter ausgeführt, hervorgehoben seien an dieser Stelle folgende Punkte:

- Zu den Vorzügen des gewählten Verfahrens zählt seine Mehrdimensionalität. Diese ermöglicht es, Einrichtungen mit unterschiedlichen Aufgabenstellungen nach gleichen Maßstäben zu bewerten und ein individuelles Profil ihrer Stärken und Schwächen für sie aufzuzeigen. Die Bewertungsgruppe Soziologie hält die den drei Dimensionen „Forschung“, „Nachwuchsförderung“ und „Wissenstransfer“ zugeordneten Kriterien Forschungsqualität, Impact/Effektivität, Effizienz, Nachwuchsförderung, Transfer in andere gesellschaftliche Bereiche und Wissensvermittlung und -verbreitung grundsätzlich für geeignete Kriterien zur vergleichenden Bewertung der Leistungsfähigkeit in der Forschung, empfiehlt aber in einer Hinsicht, Möglichkeiten der teilweisen Zusammenfassung zu prüfen. Vorgeschlagen wird, dass die beiden Transferkriterien künftig für Fächer wie die Soziologie einerseits zusammengefasst werden, andererseits aber dadurch differenziert werden, dass sie nicht nur auf der Aggregatebene der Forschungseinrichtungen, sondern auch auf der Ebene ihrer diversen Forschungseinheiten bewertet werden. Der Erhebungs- und Analyseaufwand für die drei Kriterien der Dimension Forschung könnte überdies dadurch reduziert werden, dass Bewertungen von Impact/Effektivität und Effizienz überwiegend rechnerisch vom Kriterium Forschungsqualität her bestimmt werden.
- Da sich in der Soziologie die überwiegend kleinen Forschungseinheiten innerhalb einer Einrichtung sowohl im Hinblick auf Forschungs- als auch Transferleistungen häufig deutlich unterscheiden und deshalb in diesen beiden Dimensionen nicht nur als Erhebungs-, sondern auch als Analyseeinheiten behandelt werden sollten, müssen im Hinblick auf ihre Bestimmung durch die Forschungseinrichtungen Vorgaben gemacht bzw. Kontrollen durchgeführt werden, um zu verhindern, dass taktische Überlegungen bei der Erhebung zum Ausweis künstlich gebildeter Aggregate führen. Nach mehrheitlicher Auffassung der Bewertungsgruppe kann für die Analyse von Leistungen der Nachwuchsförderung auf Bewertungen auf der Ebene der Forschungseinheiten verzichtet werden, da zentrale Leistungen der Nachwuchsförderung in starkem Maße auf der Ebene der Einrichtungen beschlossen und durchgeführt werden.
- Für die Bewertungsgruppe ist es im Hinblick auf ihre Empfehlung von Forschungsratings entscheidend, dass an der gutachterlichen Lektüre ausgewählter Publikationen durch Mitglieder der Bewertungsgruppe trotz des hohen

Aufwandes festgehalten wird, da nur auf diese Weise eine qualitative Basis zur Einschätzung der quantitativen Publikationsdaten gebildet wird.

- Die in der Pilotstudie erhobenen Indikatoren (siehe dazu Anhang 1 und 2) sind nach Ansicht der Bewertungsgruppe im Wesentlichen beizubehalten. Für die meisten Kriterien sind in der Soziologie vor allem Publikationen und deren Merkmale zentral. Drittmittel sollten nach Auffassung der Bewertungsgruppe für Qualitätsurteile in Fächern wie der Soziologie zurückhaltender eingesetzt werden, als es in vielen Leistungsbilanzen gegenwärtig üblich ist.
- Eine erhebliche Verringerung des Begutachtungsaufwandes lässt sich erreichen, indem die Qualität der Daten, die aufgrund der Datenanforderungen der Bewertungsgruppe von den Forschungseinrichtungen und ihren betroffenen Wissenschaftlern geliefert werden, verbessert wird. Das Wissenschaftssystem ist auf diesen Bedarf offenkundig noch nicht hinreichend eingestellt. Für eine Routinisierung von Forschungsratings sowie auch für die Zwecke der Selbststeuerung der Forschungseinrichtungen erscheint es sowohl erforderlich als auch möglich, dass die Forschungseinrichtungen in größerem Maße verlässlichere Daten vorhalten. Dass diese bei der Datenerhebung für Forschungsratings korrekt eingegeben werden, hängt aber auch davon ab, dass durch eine gute Kooperation zwischen Bewertenden und Bewerteten Vertrauen in das Verfahren hergestellt wird.
- Bei der Arbeit der Bewertungsgruppe sind Erfahrungen im Hinblick auf die Spezifizierung der Datennachfrage und die Praktiken der Datenanalyse entstanden. Im Projektbericht sind in dieser Hinsicht deshalb Möglichkeiten einer zunehmenden Qualifizierung des Forschungsratings angesprochen worden. Dazu gehört nicht zuletzt eine weitere Standardisierung sowohl von Fragestellungen als auch von Indikatorenbildungen. In welchem Maße deren quantitative Gewichtung für die Urteilsbildung fixiert werden kann, blieb in der Bewertungsgruppe strittig.

Die im Abschlussbericht näher ausgeführten Vorschläge zum weiteren Verfahren ergeben sich aus den Erfahrungen, die die Bewertungsgruppe in der Pilotstudie gewonnen hat. Die Offenlegung nicht nur der ermittelten Befunde, sondern auch der dabei eingesetzten Methoden und der angefallenen Zwischenergebnisse dienen dem Zweck, das Forschungsrating für speziell Betroffene und allgemein Interessierte nachvollziehbar und kritisch diskutierbar zu machen. Die Pilotstudie sollte nur als ers-

ter Schritt einer Entwicklung verstanden werden, die die Bewertungsgruppe als unumkehrbar ansieht. „Rating“ wird als Teil evaluativer Prozesse im Wissenschaftssystem nicht verschwinden. Die nachhaltige Qualität der Verfahren und die Förderlichkeit ihrer Effekte werden davon abhängen, dass sie Teil eines Lernprozesses sind, an dem auch die Bewerteten mit ihren Fachgemeinschaften selber kritisch mitwirken. Dafür bietet der Wissenschaftsrat nach Einschätzung der Bewertungsgruppe Soziologie bessere Voraussetzungen als ministerielle oder kommerzielle Träger des Verfahrens.



## **A. Ausgangslage**

Die Pilotstudie Forschungsrating geht zurück auf die „Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung“, die der Wissenschaftsrat im November 2004 verabschiedet hat. Darin empfiehlt der Wissenschaftsrat, die Forschungsleistungen von Universitäten und außeruniversitären Einrichtungen vergleichend zu bewerten, um die Einrichtungen bei strategischen Entscheidungen zu unterstützen und durch mehr Transparenz den Wettbewerb zu fördern. Der Wissenschaftsrat hat sich in diesen Empfehlungen kritisch mit bestehenden Ranking-Verfahren auseinandergesetzt und angesichts der internationalen Erfahrungen mit vergleichenden Bewertungsverfahren eine ausschließlich auf quantitativen Indikatoren basierende Forschungsbewertung ebenso ausgeschlossen wie eine reine Reputationsmessung. Das von ihm vorgeschlagene Forschungsrating ist demgegenüber gekennzeichnet durch:

- das Prinzip des „Informed Peer Review“, d.h. eine Bewertung durch Gutachter auf der Basis standardisierter, quantitativer und qualitativer Daten, die fachspezifisch bestimmt werden;
- Mehrdimensionalität, d.h. Unterscheidung verschiedener Leistungskriterien, die nicht zu einer Gesamtbewertung verrechnet werden, um so unterschiedlichen Aufgabenstellungen verschiedener Einrichtungstypen gerecht zu werden;
- Verzicht auf Ranglistenbildung, um irreführende Scheingenauigkeit und dadurch erzeugte Fehlsteuerungseffekte zu vermeiden.

In den Empfehlungen war vorgesehen, das vorgeschlagene Verfahren möglichst in einer Pilotstudie zu erproben. Im Juli 2005 wurde beschlossen, eine solche Pilotstudie in den Fächern Chemie und Soziologie durchzuführen.<sup>3</sup>

### **A.I. Ablauf der Pilotstudie**

Für die Durchführung der Pilotstudie ist eine vom Wissenschaftsrat eingesetzte Steuerungsgruppe verantwortlich. In ihr sind neben Mitgliedern der Wissenschaftlichen Kommission und weiteren Sachverständigen die großen Wissenschaftsorganisationen Deutsche Forschungsgemeinschaft (DFG), Fraunhofer-Gesellschaft (FhG),

---

<sup>3</sup> Die folgenden Ausführungen beziehen sich – sofern nicht anders ausgewiesen – immer nur auf die Pilotstudie im Fach Soziologie.

Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF), Hochschulrektorenkonferenz (HRK), Max-Planck-Gesellschaft (MPG) und Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) durch ex officio-Mitglieder, in der Regel durch Vizepräsidenten, vertreten. Sechs Länder, der Bund und der Verband der Chemischen Industrie (VCI) sowie die Geschäftsstellen der großen Wissenschaftsorganisationen entsenden Gäste in die Steuerungsgruppe.

Für die Operationalisierung des Verfahrens und die Bewertung der Forschungsleistung im Fach Soziologie wurde eine Bewertungsgruppe eingesetzt, die sich aus insgesamt 16 Fachvertretern zusammensetzte. Die Mitglieder der Bewertungsgruppe Soziologie sind von der Steuerungsgruppe aus Vorschlägen der großen Wissenschaftsorganisationen sowie der Deutschen Gesellschaft für Soziologie ausgewählt worden. Bei der Auswahl der Gutachter wurde darauf geachtet, möglichst viele paradigmatische Orientierungen und fachliche Teilgebiete der Soziologie durch die Expertise der Bewertungsgruppe abzudecken. Zugleich sollten auch ausländische Gutachter beteiligt werden, die allerdings in der Lage sein mussten, in deutscher Sprache verfasste Unterlagen zu lesen. Zudem war die Steuerungsgruppe durch einen Gast in der Bewertungsgruppe vertreten, um die Verfahrenskohärenz sicherzustellen. Begleitet und unterstützt wurden Bewertungs- und Steuerungsgruppe bei der Durchführung der Pilotstudie durch eine Projektgruppe in der Geschäftsstelle des Wissenschaftsrates.

Die folgende Darstellung der Pilotstudie Forschungsrating ist in vier Phasen unterteilt: 1. Fachspezifische Operationalisierung durch die Bewertungsgruppe, 2. Erhebung der Daten bei den Einrichtungen über Fachkoordinatoren, 3. Bewertung der Daten durch die Bewertungsgruppe, 4. Berichte und Empfehlungen zum Verfahren.

#### **A.II. Operationalisierung des Forschungsratings für das Fach Soziologie**

Das vom Wissenschaftsrat empfohlene Forschungsrating sieht eine Anpassung des Verfahrens an das zu bewertende Fach vor. Die drei Leistungsdimensionen Forschung, Nachwuchsförderung und Wissensvermittlung bilden dafür den fächerübergreifend konstant zu haltenden Rahmen.

## II.1. Definition des Fachs Soziologie

Die Erfassung der im Bereich der Soziologie forschungsaktiven Einrichtungen setzte zunächst eine Definition des Fachs voraus. Zu diesem Zweck bestimmte die Bewertungsgruppe 25 Teilgebiete, die zusammen mit ihren Unterbereichen das Gebiet der Soziologie definieren und gegen andere Fächer abgrenzen sollten. Die Teilgebiete sind:

- Allgemeine Soziologie und Soziologische Theorie,
- Arbeits- und Industriosozilogie,
- Bevölkerungs- und Migrationssoziologie (einschl. ethnische Minderheiten, Demographie),
- Bildungs- und Erziehungssoziologie,
- Entwicklungssoziologie und Sozialanthropologie,
- Soziologische Frauen- und Geschlechterforschung, Gender Studies,
- Kultursozilogie,
- Land- und Agrarsoziologie,
- Medien- und Kommunikationssoziologie,
- Medizin- und Gesundheitssoziologie,
- Methodologie und Methoden der empirischen Sozialforschung,
- Organisations- und Verwaltungssoziologie,
- Politische Soziologie (einschl. Parteien-, Wahl- und Militärsoziologie),
- Rechtssoziologie,
- Religionssoziologie,
- Soziologie der Familie, der Lebensalter und des Lebenslaufs,
- Soziologie des Wohlfahrtsstaats und der Sozialpolitik,
- Soziologie sozialer Probleme und sozialer Kontrolle,
- Soziologie sozialer Ungleichheit und Sozialstrukturanalyse,
- Sport- und Freizeitsoziologie,
- Stadt-, Regional-, und Ländersozilogie,
- Soziologie des Raumes,
- Technik- und Umweltsoziologie,
- Wirtschaftssoziologie,
- Wissenschafts- und Wissenssoziologie (einschl. Sprachsoziologie).<sup>4</sup>

<sup>4</sup> Ausgeschlossen waren hingegen: Publizistik und allgemeine Kommunikationsforschung, Sozialgeschichte, pädagogische Sozialkunde, Sozialpsychologie und Sozialpädagogik.

## II.2. Definition der Kriterien und Skalen

Die Definition der Bewertungskriterien beruht auf den Empfehlungen des Wissenschaftsrates von November 2004. In diesen Empfehlungen wurden drei Dimensionen der Bewertung, Forschung, Nachwuchsförderung und Wissenstransfer, vorgegeben und deren Bewertung nach neun Kriterien vorgeschlagen. Die genaue Ausgestaltung der Bewertungskriterien und Bewertungsaspekte sowie die Zuordnung von Indikatoren erfolgt jeweils fachspezifisch. Grundsätzlich müssen die Kriterien dazu geeignet sein, alle Einrichtungen in einem Fach nach einheitlichen Maßstäben bewerten zu können. Darüber hinaus ist eine nach Möglichkeit fächerübergreifende Standardisierung der Kriterien wünschenswert.

Die Bewertungsgruppe Soziologie hat in Abstimmung mit der Bewertungsgruppe Chemie und der Steuerungsgruppe die Bewertungskriterien von neun auf sechs Kriterien reduziert. Die Reduktion betraf die Dimensionen Wissenstransfer und Nachwuchsförderung, die nunmehr aus zwei bzw. einem Bewertungskriterium bestehen, während die Dimension Forschung weiterhin durch drei Kriterien definiert ist. Gründe für diese Reduktion waren vor allem pragmatische Erwägungen zur Aussagekraft und zur Erhebbarkeit von Indikatoren zu den einzelnen Kriterien.

**Tabelle 1: Dimensionen und Kriterien der Bewertung**

Dimension	Kriterium
Forschung	I. Forschungsqualität
	II. Impact/Effektivität
	III. Effizienz
Nachwuchsförderung	IV. Nachwuchsförderung
Wissenstransfer	V. Transfer in andere gesellschaftliche Bereiche
	VI. Wissensvermittlung und -verbreitung

Die Bewertungsgruppen haben in Abstimmung mit der Steuerungsgruppe auch die Definition der Bewertungsskala vorgenommen. Die generelle Skala ist fünfstufig. Zusätzlich ist die Klassifikation als „nicht bewertbar“ möglich:

**Tabelle 2: Bewertungsskala Kriterien I - IV**

5	exzellent
4	sehr gut
3	gut
2	befriedigend
1	nicht befriedigend
-	nicht bewertbar

Für die Bewertungskriterien „Transfer in andere gesellschaftliche Bereiche“ sowie „Wissensvermittlung und -verbreitung“ wurde die Skala im Fach Soziologie auf drei Stufen („unterdurchschnittlich“, „durchschnittlich“, „überdurchschnittlich“) reduziert. Dies war in der Heterogenität der schwer standardisierbaren Datengrundlage und dem Mangel an quantitativen Indikatoren zu diesen Kriterien begründet, die eine differenziertere Bewertung nach Auffassung der Bewertungsgruppe nicht möglich machte.

### **II.3. Zuordnung von Indikatoren**

Die sechs Bewertungskriterien wurden in einzelne, fachspezifische Bewertungsaspekte unterteilt. Diesen wiederum wurden jeweils bestimmte qualitative und quantitative Daten zugeordnet. Der Zusammenhang von Dimensionen, Kriterien, Aspekten und Indikatoren ist in der Bewertungsmatrix wiedergegeben.<sup>5</sup>

Die Auswahl der Indikatoren wurde durch die Bewertungsgruppe mit Zustimmung der Steuerungsgruppe fachspezifisch vorgenommen. Für die Auswahl der Indikatoren waren mehrere Faktoren ausschlaggebend: die Validität und die Erhebbarkeit der Daten und die möglichst unabhängige Aussagekraft der einzelnen Indikatoren zwecks Verminderung von Redundanz. Zu jedem Kriterium wurden möglichst sowohl quantitative als auch qualitative Indikatoren ausgewählt.

Eine erste Fassung der Bewertungsmatrix wurde in einem Pretest erprobt. An diesem Pretest nahmen zwei Universitäten und eine außeruniversitäre Einrichtung teil. Dadurch konnte überprüft werden, ob das Indikatorenset den verschiedenen Einrichtungstypen mit ihren spezifischen Aufgabenstellungen gerecht wird. Die teilnehmen-

---

<sup>5</sup> Die Bewertungsmatrix ist im Anhang 2 wiedergegeben.

den Einrichtungen wurden um Feedback gebeten, das in die Entwicklung der endgültigen Bewertungsmatrix und der Fragebögen sowie in die Gestaltung des Verfahrens allgemein einging. Der Pretest lieferte eine wichtige Grundlage für eine abschließende Anpassung der Indikatoren.

#### **II.4. Definition der Forschungseinheiten**

Im Zuge der Vorbereitung der Pilotstudie hat die Steuerungsgruppe als Vorgabe für die Bewertungsgruppe beschlossen, dass bei der Bewertung des wissenschaftlich zentralen Kriteriums I: „Forschungsqualität“ auch Unterschiede innerhalb einer Einrichtung wiedergegeben werden sollten. Dafür wurde die unterhalb der Ebene Einrichtung (Universität bzw. außeruniversitäre Forschungsinstitute) angesiedelte Erhebungs- und Bewertungsebene der „Forschungseinheit“ eingeführt.

Im Rahmen der Pilotstudie Soziologie waren die teilnehmenden Einrichtungen aufgefordert, Forschungseinheiten zu bilden, indem sie zunächst die einschlägigen Professuren mit den zugehörigen Mitarbeitern identifizieren und diese dann, soweit sie eine für ihre Forschungsaktivitäten relevante übergeordnete Forschungseinheit bilden oder über einen längeren Zeitraum ein zusammenhängendes Forschungsprogramm verfolgen, zu größeren Einheiten zusammenfassen. Alternativ konnten auch Gruppen von mindestens drei promovierten Wissenschaftlern ohne Professur angegeben werden, sofern sie gemeinsam ein selbständiges Forschungsprogramm bearbeiten. Um der Tatsache Rechnung zu tragen, dass Forschergruppen gelegentlich über mehrere Einrichtungen verbunden kooperativ forschen, konnten auch „institutionenübergreifende Forschungseinheiten“ benannt werden. Von dieser Möglichkeit wurde nur einmal Gebrauch gemacht.

Vor dem Hintergrund der Tatsache, dass das Forschungsrating nach Beschlusslage im Wissenschaftsrat nur Universitäten und diejenigen außeruniversitären Institute zum Gegenstand haben sollte, die nach Art. 91 b GG in von Bund und Ländern gemeinsam gefördert werden, wurde von Betroffenen mehrfach gefragt, warum universitätsnahe, landesfinanzierte Institute nicht berücksichtigt werden. Für Einrichtungen, die den Status eines An-Instituts haben, wurde deshalb im laufenden Verfahren eine Sonderregel aufgestellt: Sie konnten als Forschungseinheit einer Universität gemel-

det werden, wenn sie die für An-Institute üblichen Mindestanforderungen hinsichtlich der institutionalisierten Verbindung mit der Hochschule erfüllten (s. A. III.2, S. 17 f.).<sup>6</sup>

## **II.5. Entwicklung von Fragebögen**

Da nur ein kleiner Teil der nach der Bewertungsmatrix (s. II.3, S. 13) notwendigen Daten in vorhandenen Datenbanken erfasst ist und diese Datenbanken zudem untereinander inkompatible Definitionen von Fachgebieten und/oder Organisationseinheiten verwenden, mussten fast alle Daten für die Pilotstudie neu erhoben werden. Bei der Entwicklung der dafür notwendigen Fragebögen wurde die Bewertungsgruppe durch eine Unterarbeitsgruppe der Steuerungsgruppe unterstützt, die aus vorwiegend administrativen Vertretern von Universitäten und außeruniversitären Instituten bestand. Zusätzlich wurde die Geschäftsstelle durch das Institut für Forschungsinformation und Qualitätssicherung (IFQ) und das Zentrum für Umfragen, Methoden und Analysen (ZUMA) der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS) beraten. Die inhaltliche Basis der Fragebögen bildete die zuvor erstellte Bewertungsmatrix. Insgesamt wurden für die Erhebung drei Fragebögen verwendet: Fragebogen I diente der Erfassung der Struktur der Einrichtungen (Forschungseinheiten, Wissenschaftler), Fragebogen II erfasste die Daten auf der Ebene der Einrichtung, Fragebogen III erfasste die Daten auf der Ebene der Forschungseinheiten (zur Datenerhebung s. A.III, S. 16 ff.).

Jeder Fragebogen bestand aus einem Text- und einem Tabellenteil, die mit handelsüblicher Office-Software zu bearbeiten waren. Im Textteil wurden Rahmeninformationen zum Profil bzw. zur Aufgabenstellung der Einrichtungen bzw. Forschungseinheiten, Hintergrundinformationen und Selbstbeschreibungen abgefragt. Im Tabellenteil wurden Listen und quantitative Daten erhoben. Vorgaben hinsichtlich der Maximallänge oder -anzahl von Einträgen wurden durch Beschränkungen von Zeichen- bzw. Zeilenzahlen umgesetzt. Für eine Office-basierte Lösung sprach vor allem, dass eine aufwändige Zugangsdatenverwaltung innerhalb der einzelnen Einrichtungen unterbleiben konnte, da diese Fragebögen per E-Mail weitergeleitet werden konnten.

---

<sup>6</sup> Die Tätigkeit der Einrichtung muss sich im Rahmen des Auftrags der Hochschule vollziehen. Die Hochschule muss im Rahmen der Satzung den erforderlichen Einfluss haben, um den Anspruch auf Verbindung mit der Hochschule und Vereinbarkeit mit den Aufgaben der Hochschule zu sichern. Die leitenden Wissenschaftler sind nach wissenschaftlichen Kriterien und unter hochschulüblichen Verfahren zu berufen. Die für die Hochschule verbindliche Freiheit von Forschung und Lehre muss auch für die Wissenschaftler des An-Instituts gelten.

## **II.6. Rahmenbedingungen der Publikations- und Zitationsanalyse**

Für das Fach Soziologie wurde beschlossen, eine ausführliche Publikationsanalyse durchzuführen. Auf eine alle Forschungseinheiten umfassende Zitationsanalyse wurde hingegen nach eingehender, zum Teil kontroverser Beratung und nachdem die Bewertungsgruppe probeweise auch Zitationen auf die Publikationen ihrer Mitglieder hatte messen lassen, verzichtet (s. B.II.4, S. 45 f., sowie Anhang 1I).

Mit der Erstellung der Publikationsanalyse wurde das Informationszentrum Sozialwissenschaften (IZ) der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS) beauftragt. Datenbasis waren die IZ-Datenbank SOLIS und Datenbanken von Cambridge Scientific Abstracts (s. A.III.4, S. 20).

Für die von den Einrichtungen gemeldeten Wissenschaftler wurden in einem ersten Schritt individuelle Publikationslisten erhoben. Diese wurden an die Einrichtungen zurückgekoppelt und daraufhin ggf. korrigiert und ergänzt. Die korrigierten Publikationslisten wurden entsprechend der Zuordnung der Autoren zu Forschungseinheiten zusammengefasst und waren Grundlage der Publikationsanalyse.

Für die gemäß der Vorbewertung durch die Gutachter besten Forschungseinheiten wurde auf Basis der so gewonnenen Daten eine weitere Zitationsanalyse durchgeführt, um zu ermitteln, ob auf diese Weise die internationale Resonanz auf deren Publikationen entgegen den ersten Annahmen nicht doch in aussagekräftiger Weise ermittelt werden kann. Diese Zitationsanalyse wurde vom Institut für Wissenschafts- und Technikforschung (IWT) der Universität Bielefeld erstellt. Dabei wurden nur die Zitationen auf Aufsätze in Zeitschriften berücksichtigt, die im Web of Knowledge von Thompson Scientific (AHCI, SCI, SSCI) ausgewertet werden (s. A.III.4., S. 20)

## **A.III. Datenerhebung**

### **III.1. Organisation und Ablauf**

Vor Einleitung der Datenerhebung musste festgelegt werden, auf welche Untersuchungsfälle sich die Bewertungen beziehen sollten. Dazu wurden in einem ersten Schritt alle staatlichen und ausgewählte private Universitäten sowie die von Bund und Ländern gemeinsam geförderten außeruniversitären Einrichtungen, die soziologische Forschung durchführen, eingeladen, an der Pilotstudie Forschungsrating teil-



zunehmen. Insgesamt reagierten 64 Einrichtungen positiv auf diese Anfrage, sieben Einrichtungen nahmen ihre Teilnahme aus verschiedenen Gründen nachträglich zurück<sup>7</sup>, sodass die Pilotstudie sich schließlich auf 57 Einrichtungen bezog. 12 Einrichtungen, die sich an der Gesamterhebung der deutschen Soziologie durch die DGS seit 2004 beteiligt haben, nahmen nicht am Forschungsrating teil, drei hingegen beteiligten sich an der Pilotstudie, hatten aber an der DGS-Erhebung nicht teilgenommen. Etwa ein Viertel der teilnehmenden Einrichtungen verfügte über keinen einschlägigen grundständigen Studiengang.<sup>8</sup>

Mit der Einladung wurde auch darum gebeten, einen Fachkoordinator zu benennen, der in den Einrichtungen für die Koordinierung und Organisation der Datenerhebung zuständig sein sollte. Weil es neben der reinen Datenerhebung auch um fachlich-strategische Fragen ging, wurde empfohlen, dafür einen Wissenschaftler zu benennen, der durch die Verwaltung unterstützt werden sollte.

Die Erhebung erfolgte in zwei Phasen: In der ersten Erhebungsphase wurden die Struktur und das wissenschaftliche Personal der Soziologie an den teilnehmenden Einrichtungen erfasst. In der anschließenden zweiten Phase wurden die bewertungsrelevanten Daten erhoben. So konnte nach Abschluss der ersten Phase parallel zur zweiten Phase bereits mit der Publikationsrecherche begonnen werden, die auf den Angaben zum wissenschaftlichen Personal basierte.

### **III.2. Erfassung der Forschungseinheiten**

Die erste Erhebungsphase fand im Sommer 2006 statt. Die teilnehmenden Einrichtungen meldeten alle in der Soziologie beschäftigten Wissenschaftler und ordneten diese Forschungseinheiten zu. Als Wissenschaftler, die namentlich und unter Angabe des Beginns ihrer Beschäftigung benannt werden sollten, galten Professoren und wissenschaftliche Mitarbeiter<sup>9</sup>, sofern sie am Stichtag, dem 31.12.2005, angestellt waren und Forschung zu ihren Dienstaufgaben gehörte.

Zum Zuschnitt der Forschungseinheiten gab es Richtlinien (vgl. A.II.1, S. 11 f.), aber keine strikten Vorgaben. Dementsprechend wurde der Zuschnitt der Forschungsein-

---

<sup>7</sup> Technische Universität Braunschweig, Universität Flensburg, Technische Universität Ilmenau, Universität Erlangen-Nürnberg (Teil Nürnberg), Carl von Ossietzky Universität Oldenburg, Deutsche Hochschule für Verwaltungswissenschaften Speyer, Julius-Maximilians-Universität Würzburg.

<sup>8</sup> Alle Studiengangsinformationen nach dem Hochschulkompass der Hochschulrektorenkonferenz, Stand Oktober 2006.

<sup>9</sup> Emeriti, Vertretungs-, Gast- und Honorarprofessoren, Privatdozenten ohne Anstellungsvertrag sowie wissenschaftliche und studentische Hilfskräfte sind nicht berücksichtigt.

heiten von den Einrichtungen unterschiedlich gehandhabt, wobei häufig sehr kleine, nur aus einer Professur mit dem(n) zugehörigen Mitarbeiter(n) bestehende Forschungseinheiten benannt wurden. Mögliche Gründe für die heterogenen, überwiegend aber sehr kleinen Größenverhältnisse sind folgende Bedingungen:

- Die Hochschulen verfügen in der Regel über gewachsene (Instituts- bzw. Seminar-) Strukturen, die sich an der Lehre, nicht an der Forschung orientieren; eine gemeinsame Identität als Teil einer Einrichtung ist im Fach deshalb häufig nicht entwickelt.
- Soziologische Forschung ist nicht selten in mehreren Fakultäten oder Fachbereichen einer Universität in verschiedenen disziplinären Kontexten angesiedelt. An 20 der 54 teilnehmenden Universitäten wurden Einheiten aus zwei oder mehr Fakultäten/Fachbereichen gemeldet. Der Kontakt der soziologischen Professuren untereinander ist auch deshalb häufig locker, die Bereitschaft, gemeinsam eine Forschungseinheit zu bilden, entsprechend gering.
- Die Mehrzahl der Forschungseinheiten wendet nach eigenen Angaben nicht nur soziologische Fragestellungen und Methoden an, sondern forscht multidisziplinär. Ein Viertel der teilnehmenden Einheiten schätzt den Anteil der Soziologie an den eigenen Forschungen auf höchstens 50 %. Dies deutet darauf hin, dass die Soziologie an ihren breiten Rändern nur unscharf definiert ist, was die Bildung von größeren Forschungseinheiten erschwert haben dürfte.
- Außerdem werden bei der Bestimmung der Forschungseinheiten mit Blick auf die vorgestellten Evaluationsergebnisse auch taktische Überlegungen eine Rolle gespielt haben.

Insgesamt meldeten die 57 teilnehmenden Einrichtungen 254 Forschungseinheiten. Das heißt, im Mittel hat jede Einrichtung etwa 4,5 Forschungseinheiten benannt, wobei die Zahl der Forschungseinheiten pro Einrichtung von einer bis zehn variiert. Neun Universitäten haben jeweils acht bis zehn Forschungseinheiten gebildet, während sieben andere Universitäten ihr ganzes Soziologisches Institut zu einer Forschungseinheit zusammengefasst haben. Eine Forschungseinheit wurde zwei Universitäten zugeordnet („institutionenübergreifend“), bei sechs weiteren handelte es sich um An-Institute von Universitäten. Den 254 Forschungseinheiten wurden wiederum durchschnittlich etwa 5,7 Wissenschaftler (inklusive drittmittelfinanziertem Personal), darunter 1,5 Professoren zugeordnet, wobei drei Viertel aller Forschungs-

einheiten nur eine Professur umfassten und die Zahl der ausgewiesenen Mitarbeiterstellen von 0 bis 18 Vollzeitäquivalenten (VZÄ) pro Forschungseinheit sehr stark variiert. Mit 376 Professoren zum Stichtag (davon 78 Professorinnen, d.h. 21 %) haben etwa so viele Wissenschaftler an der Pilotstudie Soziologie teilgenommen wie die Deutsche Gesellschaft für Soziologie in ihrer Gesamterhebung seit 2004 ermittelt hat. Insgesamt wurde also eine sehr hohe Erfassungsquote erzielt.

**Tabelle 3: Struktur der Forschungseinheiten**

Universitäten und außeruniversitäre Einrichtungen	57
Forschungseinheiten	254 (4,5 je Einrichtung)
Wissenschaftler je Forschungseinheit (inkl. drittmittelfinanzierte), davon Professoren	5,7 1,5

### **III.3. Datenerhebung**

Die zweite Phase der Datenerhebung setzte im Oktober 2006 ein und war zunächst bis Mitte Dezember veranschlagt, lief nach Fristverlängerung aber bis Ende Januar 2007. In dieser Phase wurden mittels der zuvor entwickelten Fragebögen (vgl. II.5, S. 15 f.) bei den Einrichtungen und Forschungseinheiten die bewertungsrelevanten Daten erhoben. Die Fragebögen wurden in elektronischer Form versandt und konnten durch die Fachkoordinatoren auf Datenträgern oder per E-Mail an die Geschäftsstelle zurückgeschickt werden. Jede Einrichtung erhielt einen Fragebogen II, in dem die auf die Ebene der Einrichtung bezogenen Angaben abgefragt wurden, und je Forschungseinheit einen Fragebogen III für auf die Forschungseinheiten bezogene Angaben.

Neben den bei den Einrichtungen selbst erhobenen Daten wurden externe Daten für die Publikationsanalyse durch das Informationszentrum Sozialwissenschaften (IZ) in den Datenbanken SOLIS und mehreren Datenbanken von Cambridge Scientific Abstracts recherchiert und abschließend durch die Einrichtungen kontrolliert und ergänzt.

Die Erhebung der Daten erfolgte soweit möglich nach dem „Current Potential“-Prinzip, d.h. die Forschungsleistung eines Wissenschaftlers (insbesondere die Publikationsleistung) wird derjenigen Einrichtung zugerechnet, an welcher er am Stichtag 31.12.2005 beschäftigt war. Wechselte ein Wissenschaftler im Erhebungszeitraum 1.1.2001 - 31.12.2005 vor dem Stichtag an eine andere Einrichtung, wird seine ge-

samte Leistung der neuen Einrichtung zugerechnet. Ging ein Wissenschaftler vor dem 31.12.2005 in den Ruhestand, wird er nicht erfasst. Institutionelle Daten, etwa die verausgabten Drittmittel oder Promotionsprogramme, wurden jedoch nach dem hier ansonsten nicht angewandten „Work Done At“-Prinzip<sup>10</sup> erhoben (vgl. S. 42 f.).

Die Geschäftsstelle des Wissenschaftsrates hat die Fachkoordinatoren während der Datenerhebung telefonisch und postalisch betreut und auf ihrer Webseite eine ständig weiterentwickelte Liste häufig gestellter Fragen und Antworten eingerichtet (FAQs).<sup>11</sup> Über die Hälfte der teilnehmenden Einrichtungen wandten sich in der zweiten Erhebungsphase mit Fragen an die Geschäftsstelle. Trotz intensiver Betreuung und Begleitung der Erhebung waren bei 51 von 57 Einrichtungen (89%) Korrekturen und Nacherhebungen von Daten erforderlich. Zu den häufigsten Korrekturgründen zählten:

- Angaben außerhalb des Erhebungszeitraums;
- fehlende Angaben von Jahreszahlen / Zeiträumen;
- Unstimmigkeiten zwischen Fragebogen I (Erfassung der Wissenschaftler) und Fragebogen II (Personalzahlen);
- Unstimmigkeiten zwischen Fragebogen II (summarische Angaben der Einrichtung) und Fragebogen III (Detailinformationen der Forschungseinheiten);
- nicht gefragte Angaben zu Ämtern in der akademischen Selbstverwaltung oder zu von der Einrichtung selbst vergebenen Preisen;
- lückenhafte oder nicht sachdienliche Selbstbeschreibungen.

Die Geschäftsstelle teilte den jeweiligen Fachkoordinatoren alle Korrekturen mit einer kurzen Begründung mit und bat, soweit notwendig, um Ergänzungen.

#### **III.4. Publikationserhebung und bibliometrische Analyse**

Parallel zur Datenerhebung bei den Einrichtungen führte das IZ Sozialwissenschaften eine Publikationsrecherche durch. Auch hier war das leitende Prinzip die „Current-Potential“-Analyse. Die Publikationen der zum Stichtag (31.12.2005) in den Einrichtungen und Forschungseinheiten gemeldeten Wissenschaftler aus dem gesamten Erhebungszeitraum wurden also alle ihrer jeweils aktuellen Einrichtung zugerechnet. Dabei wurden alle Publikationen der gemeldeten Wissenschaftler erfasst,

<sup>10</sup> Die Methode wurde für das Forschungsrating im Fach Chemie angewandt.

<sup>11</sup> Die „Frequently Asked Questions“ sind einzusehen unter: [www.wissenschaftsrat.de/pilot\\_start.htm](http://www.wissenschaftsrat.de/pilot_start.htm).

das Fach Soziologie wurde also nicht etwa durch Festlegung einer bestimmten Zeitschriftenmenge (Positivliste) definiert, sondern ausschließlich über die Zuordnung der Wissenschaftler durch die Einrichtungen. Publikationen der folgenden Typen wurden berücksichtigt:

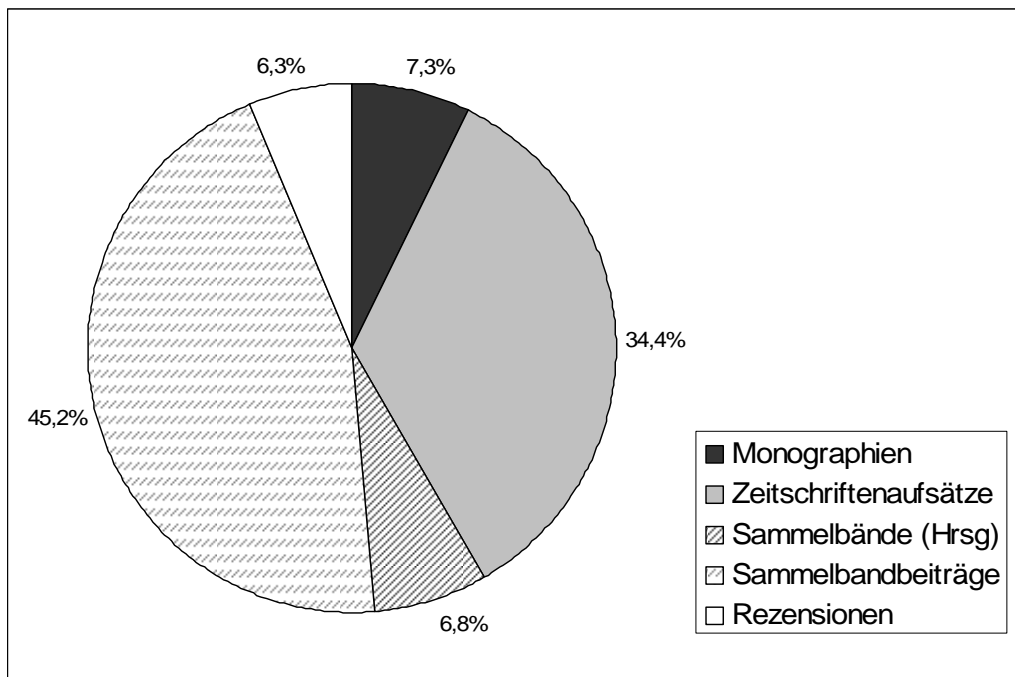
1. Monographien
2. Aufsätze in Zeitschriften
3. Herausgegebene Sammelwerke, Festschriften und Themenhefte
4. Beiträge zu Sammelwerken, Festschriften und Themenheften
5. Literaturberichte und Rezensionen (sofern Wissenschaftler letztere selbst nachtragen wollten).

Nicht erfasst wurde so genannte Graue Literatur, unter der nicht verlagsgebundene Publikationen verstanden wurden.

Die zunächst erstellten Publikationslisten wurden durch die Einrichtungen kontrolliert und konnten online über ein Selbstmeldesystem ggf. modifiziert werden. Dazu erhielt Anfang Januar 2007 jeder gemeldete Wissenschaftler über den Fachkoordinator einen eigenen passwortgeschützten Zugriff, mit dem Änderungen an den Publikationslisten bis zum 2. März 2007 möglich waren. Über eine Online-Maske konnten Korrekturen und Ergänzungen eingegeben werden, die anschließend von Assessoren des IZ verifiziert und auf Konformität mit den Erhebungsregeln der Pilotstudie hin überprüft wurden. Die Publikationsdaten gingen zusammen mit den detaillierten Publikationslisten in die Datenberichte der Einrichtungen ein.

Insgesamt wurden in der Publikationsanalyse für den Zeitraum 2001 bis 2005 10.622 Publikationen erfasst, die sich wie folgt auf die erfassten Publikationstypen verteilen:

**Abbildung 1: Verteilung der Publikationen auf Publikationstypen**



Den quantitativ größten Anteil an den Publikationen haben Beiträge zu Sammelbänden – einschließlich Festschriften, herausgegebenen Themenheften etc. – sowie die entsprechenden Herausgeberschaften. Etwa ein Drittel der Publikationen sind Artikel in Fachzeitschriften, die sich auf über 1.000 verschiedene Zeitschriftentitel verteilen. Monographien haben mit 7,3 % zahlenmäßig einen deutlich kleineren Anteil. In die Bewertung gingen in unterschiedlichen Kriterien nicht nur absolute Werte (Publikationszahlen), sondern auch relative (Publikationen pro gewichtetem Vollzeitäquivalent) ein. Auf eine Gewichtung der Publikationen, bei der eine Monographie beispielsweise mit einer bestimmten Zahl von Zeitschriftenaufsätzen gleichgesetzt werden könnte, wurde zugunsten einer differenzierten Aufschlüsselung der Publikationstypen verzichtet.

Um aus den Fachzeitschriften eine Unterauswahl besonders qualitätskontrollierter Organe zu treffen, wurden solche Zeitschriften als „Peer Reviewed“ angesehen, die als solche für bestimmte Zitationsindizes ausgewertet werden.<sup>12</sup> Die Bewertungsgruppe zog dieses Vorgehen der Definition einer eigenen Kernzeitschriftengruppe vor. Das berücksichtigte Spektrum an „Peer Reviewed“ Zeitschriften ist sehr weit de-

<sup>12</sup> Als referierte Zeitschriften („peer reviewed“) wurden von der Bewertungsgruppe solche verstanden, die für folgende Zitationsindizes ausgewertet werden: Arts and Humanities Citation Index (AHCI), Science Citation Index (SCI) und Social Sciences Citation Index (SSCI), CSA Sociological Abstracts Journals (SAJ). Für jeden dieser Indizes werden jeweils über 1.000 Zeitschriften ausgewertet. Dabei müssen die Zeitschriften bestimmte Anforderungen erfüllen, um für diese Indizes ausgewertet zu werden, vgl. <http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection/>

finiert. Insgesamt sind es 375 verschiedene Zeitschriften, in denen im Erhebungszeitraum Publikationen nachgewiesen werden. Außerdem wurden die Forschungseinheiten aufgefordert, eine von der Anzahl der Wissenschaftler je Forschungseinheit abhängige Zahl von Publikationen zur Begutachtung durch die Bewertungsgruppe einzureichen. Pro Forschungseinheit sollten mindestens zwei Publikationen eingereicht werden. Ab vier Wissenschaftlern konnte pro zusätzliche drei Wissenschaftler je eine weitere Publikation eingereicht werden (z.B. 4-6 Wissenschaftler durften maximal 3, 7-9 Wissenschaftler maximal 4 Publikationen einreichen usw.).

666 der 739 Publikationen (90%), die zur gutachterlichen Lektüre durch jeweils zwei Mitglieder der Bewertungsgruppe eingereicht wurden, entsprachen den zuvor angegebenen Auswahlkriterien der Bewertungsgruppe, nach denen Auszüge von Monographien (von nicht mehr als 50 Seiten), Aufsätze in Zeitschriften (auch E-Journals), Beiträge zu Sammelwerken sowie Literaturberichte und Rezensionen eingereicht werden konnten. Alle Publikationen mussten in den Jahren 2001 bis 2005 erschienen sein. Jeder Gutachter erhielt somit etwa 80 bis 90 Publikationen von maximal 50 Seiten Umfang zur Prüfung.

Insgesamt spielen Publikationsdaten bei mehreren Indikatoren in jeweils unterschiedlichen Aggregationen eine herausragende Rolle. Einzelheiten lassen sich der Bewertungsmatrix entnehmen (s. Anhang).

Für die versuchsweise durchgeführte Zitationsanalyse (vgl. II.6, S. 16 ff.) wurden 434 Publikationen in referierten, im Web of Science (AHCI, SCI, SSCI) erfassten Fachzeitschriften aus 52 verschiedenen Forschungseinheiten ausgewertet. Insgesamt wurden 1.629 Zitationen auf diese Artikel analysiert. Die Ergebnisse dieser eingeschränkten Zitationsanalyse gaben keinen Anlass zu Korrekturen der schon vorliegenden Bewertungen.

### **III.5. Aufbereitung der Daten**

Im Anschluss an die Datenerhebung bereitete die Geschäftsstelle des Wissenschaftsrates die Daten für die Bewertungsgruppe auf. Die bei den Einrichtungen erhobenen und die externen Daten wurden partiell zu abgeleiteten Indikatoren verrechnet, alle Angaben den Bewertungsebenen (Einrichtung oder Forschungseinheit) und Kriterien zugeordnet und zu so genannten „Datenberichten“ jeder einzelnen Ein-

richtung zusammengefasst. Zusätzlich enthielten die Datenberichte die Rahmeninformationen der Einrichtung und ihrer Forschungseinheiten sowie die Publikationslisten. Im Mittel waren die Datenberichte je Einrichtung etwa 45 Seiten lang; sie streuten zwischen 19 und 119 Seiten.

Nach Eingang und Kontrolle aller Daten wurden für die quantitativen Indikatoren statistische Lagemaße (Perzentile) ermittelt und in die Datenberichte eingefügt. Sie hatten die Funktion, den Gutachtern für ihre Qualitätseinschätzung von Forschungseinheiten und -einrichtungen die relative Lage der gemessenen Werte, bezogen auf die Gesamtverteilung der Indikatorenwerte, transparent zu machen. Weitere Verteilungsmaße (Median, 1. und 3. Quartil) wurden in einem „Leitfaden zu den Datenberichten“ ausgewiesen, sodass sie in der Bewertung berücksichtigt werden konnten.<sup>13</sup>

Die Datenberichte wurden den Einrichtungen zur abschließenden Kontrolle vorgelegt und erst nach erfolgter Rückkoppelung an die Bewertungsgruppe weitergeleitet. Da allerdings bei der Frist der Datenerhebung eine große Kulanz gewährt worden war, musste die Rückkopplungsschleife auf rund vier Wochen verkürzt werden, sodass die erneute Datenkontrolle seitens der Einrichtungen bei sehr spätem Einreichen der Fragebögen in kurzer Zeit erfolgen musste.

Zusammen mit den Datenberichten erhielten die Gutachter die ausgewählten Publikationen, die von den einzelnen Forschungseinheiten eingereicht worden waren.

#### **A.IV. Bewertungsphase**

##### **IV.1. Organisation und Ablauf**

Die Bewertung der Forschungsleistungen auf Basis der Datenberichte war Aufgabe der Bewertungsgruppe. Die Bewertungen wurden arbeitsteilig vorgenommen: Jeder Einrichtung und jeder Forschungseinheit wurden mindestens zwei Berichtersteller aus der Bewertungsgruppe zugeordnet, ausgehend von deren Expertise und orientiert an den Teilgebietszuordnungen der Forschungseinheiten. Dazu wurden Befangenheiten zuvor abgefragt und bei der Zuordnung berücksichtigt. Wenn es aufgrund von Befangenheiten oder fehlender Expertise nicht möglich war, Gutachter aus der Bewertungsgruppe zuzuordnen, wurde auf externe Sondergutachter zurückgegriffen.

---

<sup>13</sup> Der anonymisierte „Leitfaden zu den Datenberichten für die Bewertungsgruppe Soziologie“ ist einzusehen unter [http://www.wissenschaftsrat.de/pilot\\_start.htm](http://www.wissenschaftsrat.de/pilot_start.htm).



Sondergutachter wurden ferner auch dann hinzugezogen, wenn ein Dissens der Berichterstatter nicht aufzulösen war. Forschungseinheiten, in denen Mitglieder der Bewertungsgruppe selber vertreten waren, wurden vom Vorsitzenden der Bewertungsgruppe in Zusammenarbeit mit zwei weiteren Sondergutachtern mit Noten bewertet, die im Unterschied zu anderen gutachtlichen Bewertungsbefunden in der Bewertungsgruppe nicht diskutiert wurden. Insgesamt wurden für das Fach Soziologie sechs Sondergutachter hinzugezogen. Die Bewertung basierte auf den Datenberichten jeder Einrichtung und den im Leitfaden angegebenen Verteilungsmaßen sowie den Publikationslisten jeder Einrichtung bzw. Forschungseinheit. Zusätzlich waren die ausgewählten Publikationen direkt zu bewerten. Dieses besonders aufwändige Element der Bewertung bedeutete für jeden Gutachter die Lektüre von im Durchschnitt 80-90 eingereichten Publikationen mit insgesamt ca. 3000 Seiten.

Der Ablauf der Bewertung war in drei Phasen eingeteilt: In der ersten Phase, der individuellen Bewertungsphase, bewerteten die Berichterstatter unabhängig voneinander die ihnen zugeordneten Forschungseinheiten und Einrichtungen über ein durch das Zentrum für Evaluation und Methoden (ZEM) der Universität Bonn erstelltes Online-Eingabeformular bzw. über ein Papierformular. Das Formular sah die Möglichkeit vor, jede Bewertung einer Forschungseinheit bzw. einer Einrichtung in einem Anmerkungsfeld separat zu erläutern. Die Bewertung erfolgte auf der 5-stufigen Notenskala. In der individuellen Benotung konnte die Note mit einer Tendenz versehen werden (etwa 3 + = „3 gut“ mit einer Tendenz zu „4 sehr gut“), um die anschließende Konsensfindung zu orientieren. In der individuellen Bewertungsphase wurden auch Noten für jeden der in der Bewertungsmatrix verzeichneten Bewertungsaspekte (=Teilaspekte eines Kriteriums) vergeben (siehe Anhang 2, S. 75 ff.).

Für die Bewertung der quantitativen Daten wurden Richtwerte zur Interpretation der Skala aufgestellt, von denen aber in begründeten Fällen (etwa bei erkennbaren Ungereimtheiten der Datengrundlage) abgewichen werden konnte. Folgende Skalenswertbereiche der Perzentilwerte wurden festgelegt:

**Tabelle 4: Skalenwertbereiche der Perzentilwerte quantitativer Daten**

Perzentilwert	Skalenwert
90 – 100 %	5 exzellent
70 – 90 %	4 sehr gut
30 – 70 %	3 gut
10 – 30 %	2 befriedigend
0 - 10 %	1 nicht befriedigend

Allerdings gab es für die Gutachter keine strikte Vorgabe, mit welcher Gewichtung die Beurteilung eines Einzelaspekts in die Gesamtnote eines Kriteriums eingehen sollte. Deshalb war die empirische Verteilung der Benotungen durch diese Skalenverteilung nicht schon vorab festgelegt etwa in dem Sinne, dass 10 % der Einrichtungen „nicht befriedigend“ sein mussten.

Um auch ohne quantitative Gewichtungen eine möglichst konsistente Bewertung durch die Berichtersteller zu erreichen, wurden Bewertungsaspekte mit besonderer Relevanz innerhalb der Kriterien identifiziert, denen dann Vorrang gegeben werden sollte, wenn die Bewertung eines Kriteriums uneinheitlich ausfiel (siehe Qualitätsmodell). Auch innerhalb der Bewertungsaspekte wurden primär relevante Daten definiert, denen bei Abwägung disparater Informationen ein erhöhtes Gewicht zukommen sollte, ohne dies jedoch mit quantitativen Faktoren zu beziffern. Die Endbewertung erfolgte bei jedem Kriterium unter Berücksichtigung quantitativer Verteilungswerte im Sinne des „Informed Peer Review“, also durch qualitative Abwägungen der Gutachter.

Die Übereinstimmung der individuellen Gutachterurteile war in der ersten Bewertungsrunde mehrheitlich recht hoch, variierte allerdings deutlich zwischen den verschiedenen Kriterien: Bei den Kriterien I bis IV wichen die Vorschläge der Berichtersteller in 71 – 89 % der Fälle nicht bzw. um weniger als eine Notenstufe voneinander ab. Nur bei den beiden letzten Kriterien (Transfer in andere gesellschaftliche Bereiche und Wissensvermittlung und -verbreitung) kam es anfangs häufiger zu Abweichungen (je 41 % der Fälle) aus Gründen, die mit der wenig standardisierbaren und deshalb sehr heterogenen Datenlage speziell in diesen Fällen zusammenhängen.

Die Bewertungsgruppe hat dies zum Anlass genommen, bei diesen beiden Kriterien der Dimension „Wissenstransfer“ die Bewertungsskala auf drei Stufen (unterdurchschnittlich / durchschnittlich / überdurchschnittlich) zu reduzieren. In der zweiten Bewertungsphase waren die beiden Berichterstatter bei Nichtübereinstimmung ihrer ersten Notenvorstellungen zunächst dazu angehalten, bilateral einen Konsens zu erreichen, was in fast allen Fällen auch gelang, in wenigen Einzelfällen nach Einschaltung eines Sondergutachters mit Spezialkompetenz. Danach wurde in einer dritten Bewertungsphase über die individuellen Bewertungen in der gesamten Bewertungsgruppe plenar abgestimmt. Dafür wurden insgesamt drei zweitägige Sitzungen abgehalten. In der Abschlusssitzung wurde zusätzlich die Vergabe der Noten „5 exzellent“ und „1 nicht befriedigend“ noch einmal gesondert geprüft und darüber hinaus ein abschließender Abgleich aller vorgenommenen Bewertungen durchgeführt. In den Sitzungen wurden alle Bewertungen einzeln angesprochen, wobei 254 Bewertungen für Kriterium I (Ebene Forschungseinheit) und jeweils 57 Bewertungen für die Kriterien II, III, IV, V und VI festzulegen waren. In der Regel wurden die Noten kurz begründet, bevor das Plenum über die endgültige Note im Falle offenbleibender Fragen diskutierte und abstimmte. Befangene Gutachter nahmen bei den betroffenen Einrichtungen nicht an der Abstimmung teil. Erwartungsgemäß war der Diskussionsbedarf bei hohem Dissens in der Vorbenotung in der Regel höher als bei übereinstimmenden Bewertungen. Am Ende eines mehrstufigen Bewertungsprozesses gab es in jedem Fall eine einmütige Notenfestlegung. Minderheitenvoten wurden in keinem Fall zu Protokoll gegeben, obwohl dies in der Bewertungsgruppe nicht als problematisch gegolten hätte.

Statistisch lässt sich eine leichte Tendenz der Bewertungsgruppe aufzeigen, Dissens im Plenum positiv, d.h. in der Benotung etwas „nach oben“ aufzulösen; der Durchschnitt der endgültigen Bewertungen liegt je nach Kriterium um 0,1 bis 0,2 Notenstufen über dem Durchschnitt der Notenvorschläge der Berichterstatter. Welchem Berichterstatter eine Forschungseinheit zugeordnet worden ist, spielt für das plenar bestimmte Endergebnis der Bewertung statistisch keine Rolle.<sup>14</sup>

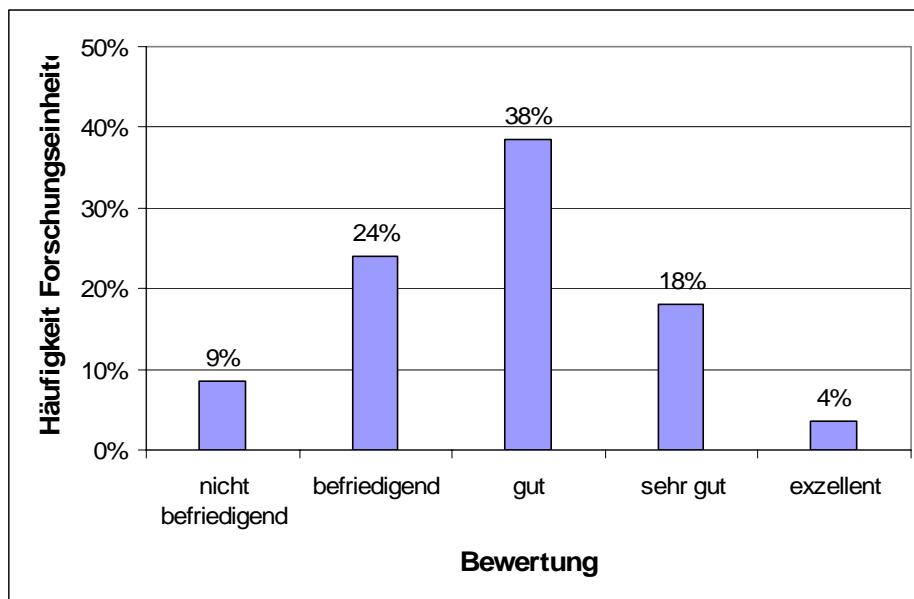
---

<sup>14</sup> Bei einer Erklärung der Bewertungsergebnisse durch multiple Regressionsanalyse mit der Berichterstatterzuordnung als Dummy-Variablen ist keine dieser Dummy-Variablen auf einem Niveau <0,05 signifikant.

## IV.2. Analyse der Bewertungsergebnisse

Die durch das Plenum abschließend abgestimmten Bewertungsergebnisse streuen je nach Bewertungskriterium unterschiedlich, wobei in allen Fällen die gesamte Notenskala ausgeschöpft wurde.

**Abbildung 2: Verteilung der Bewertungen zu Kriterium I Forschungsqualität**



An 100 % fehlende: „nicht bewertbar“ (Forschungsqualität: 7 %).

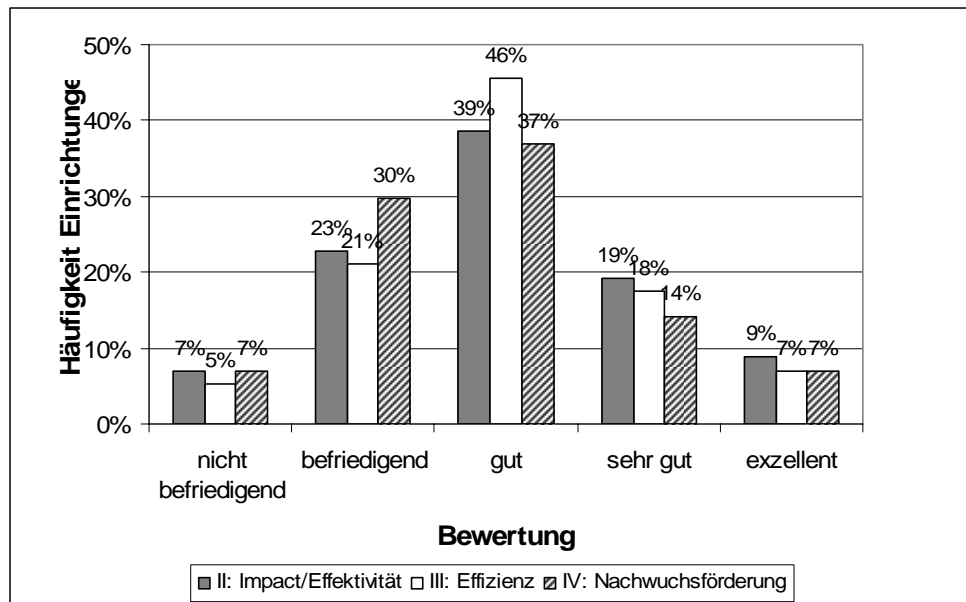
Die Bewertungen für Kriterium I Forschungsqualität streuen leicht linksschief um den Mittelwert, der deshalb etwas unterhalb von „gut“ liegt. Diese Asymmetrie spiegelt sich auch in den Extremwerten: Während die Forschungsqualität von 9 % der Forschungseinheiten mit „nicht befriedigend“ bewertet wurde, sind 4 % der Forschungseinheiten mit „exzellent“ bewertet worden. Mit 7 % ist der Anteil der „nicht bewertbaren“ Fälle auf Ebene der Forschungseinheiten am höchsten. Die häufigsten Gründe waren lückenhafte Daten oder Verletzungen der Mindestanforderungen an die Größe von Forschungseinheiten (eine Professur oder drei promovierte Mitarbeiter), etwa wegen Emeritierungen oder anders bedingten Abgängen vor dem Stichtag.

Die Spanne der Bewertungen von Forschungseinheiten innerhalb einer Einrichtung zeigt, dass eine differenzierte Bewertung auf dieser Ebene informativ ist. Die Leistungsträger der Forschung sind ganz überwiegend nicht die gesamten Einrichtungen, sondern deren kleinere Untereinheiten: So streuten die Bewertungen bei 24 der 57 Einrichtungen über drei Notenstufen, bei acht Einrichtungen über vier und bei einer

Einrichtung sogar über alle fünf Notenstufen. Von den sieben Einrichtungen, die ihre Soziologie als ganzes (also nur mit einer einzigen Forschungseinheit) zur Bewertungen angemeldet haben, haben sechs eine Bewertung mit „gut“ erhalten, was den nivellierenden Effekt einer undifferenzierten Bewertung mit entsprechenden Informationsverlusten belegt.

Die nachfolgende Übersicht zeigt für die bzgl. der Notenskala vergleichbaren Kriterien II, III und IV die jeweiligen Notenverteilungen auf Einrichtungsebene im Vergleich:

**Abbildung 3: Verteilung der Bewertungen zu Kriterien II, III und IV im Vergleich**



An 100 % fehlende: „nicht bewertbar“ (II Impact/Effektivität: 4 %; III Effizienz: 4 %; IV Nachwuchsförderung: 5 %).

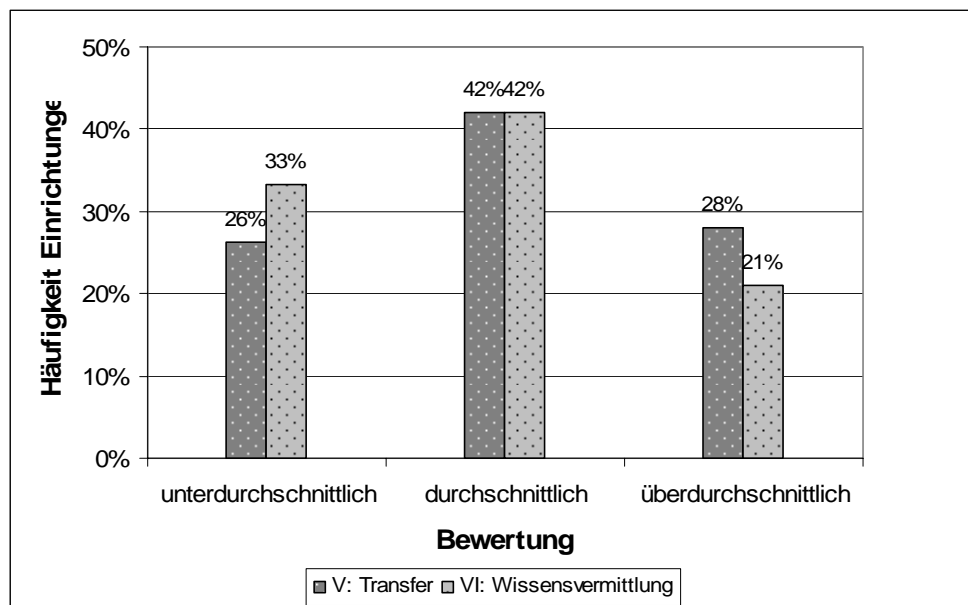
Impact/Effektivität und Effizienz streuen beide relativ symmetrisch um „gut“ (3,0), wobei die Effizienzbewertung etwas weniger stark streut – es sind weniger Extremnoten und mehr „gute“ Bewertungen vergeben worden. Dies könnte damit zu tun haben, dass die Pro-Kopf-Leistungen weniger stark divergieren als die Größen der Einrichtungen, was auch eine Analyse der Größenabhängigkeit der Ergebnisse belegt: Wie aufgrund der Definition der Kriterien zu erwarten, ist die Bewertung der Effizienz nicht von der Größe der Einrichtung abhängig (Korrelation gegen wissenschaftliches Personal:  $r^2=0,011$ ), während die Bewertung von Impact/Effektivität einen deutlich positiven Zusammenhang mit der Größe zeigt ( $r^2=0,366$ , Steigung 0,17 pro gewichte-

ten VZÄ). Auch die mittlere Bewertung der Forschungsqualität der Forschungseinheiten (nicht im Diagramm) korreliert nicht signifikant mit der Größe ( $r^2=0,003$ ); sowohl exzellente als auch nicht befriedigende Forschungseinheiten finden sich an Einrichtungen jeder Größe.

Die Bewertung der Nachwuchsförderung ist schwächer, aber wie die Impact/Effektivitätsbewertung ebenfalls positiv von der Größe abhängig ( $r^2=0,250$ , Steigung 0,14 pro gewichteten VZÄ). Vor allem die hier bewerteten Einrichtungen der Nachwuchsförderung (Promotionsprogramme etc.) setzen in vielen Fällen eine Mindestgröße der Einrichtung voraus. Auffallend ist, dass die Bewertung der Nachwuchsförderung etwas kritischer ausfällt als bei den Forschungskriterien (Mittelwert 2,83)<sup>15</sup>.

Die Bewertungen für Kriterien V und VI, Transfer in andere gesellschaftliche Bereiche sowie Wissensvermittlung und -verbreitung, die im laufenden Bewertungsverfahren auf drei Notenstufen reduziert wurden, verteilen sich folgendermaßen:

**Abbildung 4: Verteilung der Bewertungen zu Kriterien V und VI Wissenstransfer und Wissensverbreitung**



An 100 % fehlende: „nicht bewertbar“ (V Transfer: 4 %; VI Nachwuchsförderung: 4 %).

Keines der Kriterien korrelierte so hoch mit einem anderen Kriterium, dass es redundant wäre. Insgesamt führen die sechs Kriterien zu einer sinnvollen Differenzierung

<sup>15</sup> 2= befriedigend, 3=gut.

der Bewertung. Dies drückt sich auch darin aus, dass es keine dominanten „Cluster“ in der Bewertung der Einrichtungen gibt, sondern die Bewertungsergebnisse insgesamt stark heterogen sind und damit das jeweils individuelle Leistungsprofil der einzelnen Einrichtungen differenziert erfassen. Die in diesem Verfahren erprobte mehrdimensionale Bewertung hat sich somit bewährt und sollte auch für eine Weiterentwicklung ungeachtet der vorhandenen Möglichkeit leichter Zusammenfassungen (dazu B.III.1, S. 53ff.) im Prinzip beibehalten werden.

Folgerungen zu den Stärken und Schwächen der Soziologie in Deutschland, die aus den Ergebnissen gezogen werden können, finden sich in dem von der Steuerungsgruppe veröffentlichten Ergebnisbericht.<sup>16</sup>

## **A.V. Veröffentlichung und Reaktionen auf das Forschungsrating**

### **V.1. Veröffentlichung der Ergebnisse**

Die Veröffentlichung der Bewertungsergebnisse in einem Ergebnisbericht erfolgte durch die Steuerungsgruppe. Flankiert wurde die Veröffentlichung durch eine zusammenfassende Analyse der Bewertungsergebnisse im Sinne einer Stärken- und Schwächenanalyse der Soziologie in Deutschland durch die Bewertungsgruppe sowie eine kurze Erläuterung des in der Pilotstudie erstmals angewandten Verfahrens. Dieser Bericht wurde am 10. April 2008 durch die Steuerungsgruppe verabschiedet und am 18. April 2008 zusammen mit den Einzelergebnissen veröffentlicht.

In den Ergebnisdarstellungen sind neben den Bewertungen der einzelnen Einrichtungen nach den verschiedenen Kriterien auch die jeweiligen Mittelwerte über alle Einrichtungen angegeben. Dadurch ist ein Abgleich der Leistung einzelner Forschungseinrichtungen mit der Durchschnittsleistung ihrer Gesamtheit möglich.

Veröffentlicht wurden nur die Ergebnisse auf Ebene der Einrichtungen einschließlich eines anonymisierten Profils der Forschungseinheiten. Die Detailbewertungen der einzelnen Forschungseinheiten wurden hingegen aus Datenschutzgründen nur den Einrichtungen selbst, den Trägerorganisationen MPG und WGL sowie den jeweils zuständigen Ländern zum internen Gebrauch vorgelegt. Es wurde den Einrichtungen aber nahe gelegt, auch diese detaillierte Darstellung der Öffentlichkeit zugänglich zu

---

<sup>16</sup> Steuerungsgruppe im Auftrag des Wissenschaftsrats: Forschungsleistungen deutscher Universitäten und außeruniversitärer Forschungseinrichtungen in der Soziologie. Ergebnisse der Pilotstudie Forschungsrating des Wissenschaftsrates. Köln, 18.4.2008: Teil C.

machen. Die Geschäftsstelle des Wissenschaftsrates bot den Einrichtungen an, diese Veröffentlichung über eine zentrale Linkliste auf den Internetseiten der Geschäftsstelle des Wissenschaftsrates zu vernetzen.

Neben den Diagrammen enthält die Veröffentlichung der Bewertungsergebnisse in Einzelfällen Kommentare zur Bewertung. Diese liefern Erläuterungen zu eventuellen Besonderheiten der vorliegenden Bewertung der Einrichtung. Den Einrichtungen selbst wurden wiederum ggf. auch auf einzelne Forschungseinheiten bezogene Kommentare vorgelegt, die nicht durch die Steuerungsgruppe veröffentlicht wurden.

Die Einrichtungen, die Länder und die Trägerorganisationen erhielten die sie betreffenden Bewertungsergebnisse einschließlich der Detailergebnisse der einzelnen Forschungseinheiten rechtzeitig vor deren Veröffentlichung. Die Veröffentlichung wurde durch eine Pressekonferenz und ein Pressegespräch begleitet, an dem der Vorsitzende der Steuerungsgruppe, der Vorsitzende der Bewertungsgruppe Soziologie und der Generalsekretär des Wissenschaftsrates sowie der Vorsitzende der Deutschen Gesellschaft für Soziologie (DGS) teilnahmen.

## **V.2. Reaktionen auf das Forschungsrating**

Das Forschungsrating wurde vor und während der Pilotstudie in verschiedenen Foren vorgestellt, etwa im Rahmen der Mitgliederversammlung der Deutschen Gesellschaft für Soziologie (DGS) am Rande des Soziologiekongresses im Oktober 2006 in Kassel. Der Vorsitzende der Bewertungsgruppe informierte zudem in der Zeitschrift der DGS und in einem Newsletter des Berufsverbandes der Deutschen Soziologinnen und Soziologen (BDS) im Sommer 2006 über das laufende Projekt.<sup>17</sup> Auf dem Kongress der DGS 2008 wird das Forschungsrating Gegenstand einer speziellen Veranstaltung sein.

Die Fachkoordinatoren wurden gebeten, nach Abschluss der Datenerhebung der Bewertungsgruppe Feedback zum Verfahren zu geben. Einige Einrichtungen machten von dieser Option Gebrauch und äußerten sich nicht nur zu technischen Fragen bzw. Problemen im Umgang mit den Fragebögen, sondern setzten sich mit dem Verfahren in allgemeiner Hinsicht auseinander. Einige Fachkoordinatoren hatten sich

---

<sup>17</sup> Neidhardt, Friedhelm (2006): Forschungsrating der deutschen Soziologie durch den Wissenschaftsrat (in: Soziologie, 35 Jg., Heft 3, S. 303-308), auch unter: <http://www.wissenschaftsrat.de/texte/pilot-neidh.pdf>



bereits während der Erhebung zum Verfahren geäußert. Die Reaktionen der Fachkoordinatoren gehen in die Teile B und C dieses Berichts ein.



## **B. Empfehlungen**

### **B.I. Organisation und Ablauf**

Die Organisation des Forschungsratings war maßgeblich dadurch bestimmt, dass eine Pilotstudie durchzuführen war, die mit einer Massenerhebung auf die Ermittlung zahlreicher Einzelbewertungen abzielte. Es ging um ein Großprojekt, das einen erheblichen Aufwand mit sich brachte und eine differenzierte Organisation voraussetzte. Ohne die Ansiedelung der Pilotstudie beim Wissenschaftsrat hätte dieses Projekt nicht mit der ungewöhnlich hohen Beteiligung der betroffenen Forschungseinrichtungen stattfinden können.

Die Organisation, die für die Pilotstudie gewählt wurde, hat sich weitgehend bewährt. Größe und Zusammensetzung der **Bewertungsgruppe** selber erscheinen für ein Fach mit dem Ausdifferenzierungsgrad der Soziologie angemessen. Für die Akzeptanz eines Forschungsratings ist das Vertrauen in die Kompetenz und Integrität der Gutachter von großer Bedeutung. Deshalb ist es im Falle einer Institutionalisierung von Forschungsratings wichtig, bei der Zusammensetzung der Bewertungsgruppe nicht nur die großen Wissenschaftsgesellschaften, sondern auch die betroffenen Fachgesellschaften so zu beteiligen, wie das im vorliegenden Fall geschehen ist. Der Anteil ausländischer Kollegen sollte möglichst etwas größer sein, als in diesem Fall erreicht wurde. Auch die Beteiligung von emeritierten Wissenschaftlern erscheint als ein geeigneter Weg, um mögliche Befangenheiten in der Gruppe kontrollieren zu können. Es wird allerdings davon abgeraten, die Gruppe insgesamt zu vergrößern.

Die Anwesenheit eines Beobachters aus der Steuerungsgruppe war für die Einheitlichkeit des Verfahrens der nebeneinander laufenden Pilotprojekte in den Fächern Soziologie und Chemie nützlich. Für die Projektkoordination und zahlreiche Serviceleistungen erwies sich die Geschäftsstelle des Wissenschaftsrats nach Einschätzung der Bewertungsgruppe als sehr hilfreich. Die Einrichtung einer gesonderten Unterarbeitsgruppe der Steuerungsgruppe zur Entwicklung der Fragebögen war dem Pilotcharakter der Studie geschuldet und kann künftig entfallen.

Verbessert werden kann die **Organisation der Datenerhebung** in manchen der teilnehmenden Einrichtungen. In allen Universitäten und außeruniversitären Instituten waren für den Erhebungsprozess Fachkoordinatoren verantwortlich, die in der Regel

Wissenschaftler aus dem Fach Soziologie waren. Dies war deshalb nützlich, weil mit der Gliederung in Forschungseinheiten und der Abfassung von kurzen Selbstbeschreibungen Aufgaben erfüllt werden mussten, für die die fachliche Kompetenz von Wissenschaftlern unabdingbar war. Zudem sollten die Wissenschaftler die inhaltliche Letztkontrolle behalten. Gleichzeitig war es aber erkennbar ineffizient, wenn die eigentliche Datenerhebung einschließlich der Ermittlung von Finanzdaten und Vertragsbestimmungen von den Wissenschaftlern selbst durchgeführt wurde; sie sollte in der Hand der jeweiligen Verwaltung liegen. Unzureichende Unterstützung durch die Verwaltung und ein unterentwickeltes internes Controlling führten dazu, dass die Datenerhebung an einigen Universitäten nur unter hohem persönlichem Einsatz des Fachkoordinators möglich war. In einzelnen Fällen blieben die Daten lückenhaft und unsicher, sodass es häufiger, als wünschenswert ist, zu den Aufgaben nicht nur der Geschäftsstelle, sondern auch der Bewertungsgruppe gehörte, Zuverlässigkeitskontrollen durchzuführen und Unstimmigkeiten auszuräumen. Es wird daher für die Durchführung von Forschungsratings empfohlen, dass die Wahl der Forschungskoordinatoren sorgfältig erfolgt und sich an der Erhebung sowohl die Wissenschaftler selbst als auch die Verwaltung der Einrichtung beteiligen und miteinander effizient kooperieren.

Die Phasen der Pilotstudie waren von unterschiedlicher Dauer. Die erste **Phase der fachspezifischen Operationalisierung** der Kriterien und der Entwicklung von Fragebögen dauerte inklusive Pretest etwa 9 Monate. Dieser Zeitraum kann bei einer Wiederholung im selben Fach deutlich verkürzt werden. Wenn neue Indikatoren entwickelt werden, sollte auf einen Pretest aber nicht verzichtet werden.

Die zweite **Phase der Datenerhebung** wurde in der Pilotstudie in zwei Teilabschnitte aufgetrennt: Zunächst wurden die Forschungseinheiten und die ihnen zugeordneten Wissenschaftler erfasst. Dies erfolgte aus Zeitgründen bereits parallel zum Pretest der später zu verwendenden Fragebögen. Dadurch wurde zwar etwa ein Quartal eingespart, der Nachteil war aber, dass die teilnehmenden Einrichtungen noch nicht genau wussten, welche Daten sie zu den zu bildenden Forschungseinheiten würden bereitstellen müssen. Dies führte zu Irritationen und sollte deshalb künftig vermieden werden.

Für die eigentliche Datenerhebung wurde den Einrichtungen eine Erhebungsfrist von 15 Wochen gewährt, die in vielen Fällen noch ausgeweitet wurde, um eine möglichst hohe Erfassungsquote zu erhalten. Allerdings führte die großzügige Fristauslegung zu verkürzten Korrekturfristen und bedeutete eine gewisse Benachteiligung der fristgemäß einreichenden Einrichtungen. Problematisch waren auch verspätete Nachmeldungen von Personen oder Neuzuordnungen von einzelnen Wissenschaftlern zu anderen Forschungseinheiten innerhalb oder außerhalb einer Einrichtung. Für die Kontrolle und Ergänzung der Publikationslisten, die zeitlich versetzt erst zweieinhalb Monate nach dem Beginn der Datenerhebung beginnen konnten, galt eine Frist von rund sieben Wochen. Diese Frist war aus Sicht vieler beteiligter Wissenschaftler angesichts des Umfangs der notwendigen Korrekturen zu kurz, zumal sich die Freischaltung des Online-Systems aus technischen Gründen mehrmals verzögert hatte. Es hing mit dem Charakter einer Pilotstudie zusammen, dass sich das gesamte Verfahren über einen relativ großen Zeitraum hinzog und dabei längere Pausen in der Sitzungsfolge der Bewertungsgruppe entstanden. Weil eine Reihe von methodischen Einzelheiten erst während der Bewertungsphase am konkreten Beispiel ausdiskutiert werden konnten, war es nicht möglich, alle Bewertungen in der ursprünglich veranschlagten Zeit definitiv abzustimmen, sodass letzte Abstimmungen in eine weitere Sitzung verschoben werden mussten, die in größerem zeitlichem Abstand zu den ersten Bewertungssitzungen stattfand. Für solche Fälle sollte in der Bewertungsphase künftig ein ausreichender zeitlicher Puffer einkalkuliert werden.

## **B.II. Zu Einzelaspekten des Forschungsratings**

### **II.1. Zur fachspezifischen Operationalisierung des Verfahrens**

Der erste Schritt der fachspezifischen Operationalisierung war die **Definition des Faches Soziologie**. Die nicht strikt festlegende, für die Selbsteinschätzung der Forschungseinrichtungen aber orientierende Vorgabe von 25 Teilgebieten der Soziologie (vgl. A.II.1, S. 11 f.) sollte den Umfang des Fachs sowie vor allem seine Grenzen deutlich machen. Sie diene darüber hinaus der Zuordnung von Berichterstattern aus der Bewertungsgruppe zu einzelnen Forschungseinheiten.

Die Bewertungsgruppe entschied sich im Hinblick auf die besonderen Funktionen einer Pilotstudie dafür, eine relativ große Zahl von **Indikatoren** zu prüfen. Dies findet in einer Bewertungsmatrix („Qualitätsmodell“) seinen Ausdruck. Zwei Überlegungen

haben allerdings auch dazu geführt, bestimmte Indikatoren, die häufig in Evaluationen benutzt werden, im Forschungsrating nicht einzusetzen. Dies geschah erstens in der Folge theoretischer Überlegungen darüber, welche Deskriptoren tatsächlich geeignet sein können, nicht nur erhebbar, sondern vor allem auch valide Indikatoren zu sein. Zweitens waren auch die maßstabsetzenden Effekte eines vom Wissenschaftsrat durchgeführten Forschungsratings zu berücksichtigen: Die Wissenschaftler werden längerfristig jene Forschungsaktivitäten verstärken, an denen ihre Leistung gemessen wird. Eine hohe „Reagibilität“ des Feldes zwingt beim Forschungsrating nicht nur zu Reflexion methodischer, sondern auch normativer Gesichtspunkte.<sup>18</sup> Es geht also immer auch um die Frage, welche Leistungen der Forschungseinrichtungen als wünschenswert erscheinen. Mit Blick darauf wurden von der Bewertungsgruppe folgende Grundentscheidungen getroffen.

- Die Gutachter brauchen zur Einschätzung der Validität der Daten zahlreiche Kontextinformationen, die bei Evaluationen von Einzelfällen normalerweise durch so genannte Begehungen in Erfahrung gebracht werden: Aufgabenstellungen der Einrichtung; Bedingungen der Forschungsarbeit und deren Entwicklung; Einschätzungen der eigenen Forschungsarbeit und Planungsvorsätze; Informationen über die Praxis interner Kontrollen etc. Dafür war im Erhebungsbogen zum Teil offene Fragen gestellt, auch um den Betroffenen Gelegenheit zu geben, sich nach eigenen Maßstäben zu erklären.
- Auf eine umfassende Zitationsanalyse wurde verzichtet, da die in Zitationsdatenbanken erfassten referierten Zeitschriften ohne aufwändige Datenbereinigungen, die in einer Massenerhebung wie dem Forschungsrating im Einzelnen nicht durchführbar sind, nur einen kleinen Ausschnitt aus der Publikationspraxis der Soziologen abdecken (nur 8,8% der erfassten Publikationen waren Artikel in solchen Zeitschriften) (s. A.III.4, S. 20f.).
- Von jeder der 254 Forschungseinheiten, deren Forschungsqualität festzustellen war, wurde ein je nach Größe der Einheit mehr oder weniger großer Satz an Publikationen eigener Wahl angefordert. Sie sollten den Forschungsoutput der Forschungseinheiten unmittelbar einsehbar und für die Gutachter verlässlicher bewertbar machen. Die umfangreiche Lektüre gehörte zu den zentralen

---

<sup>18</sup> Dazu Espeland, Wendy Nelson / Sauder, Michael (2007): Ranking and Reactivity: How Public Measures Recreate Social Worlds. In: American Journal of Sociology, 113/1, S. 1-40.

Bestandteilen der Bewertungsarbeit. Sie bestimmte die Bewertung von Forschungsqualität in hohem Maße.

- Die Publikationslisten dienen dazu, den Gutachtern einen qualitativen Gesamteindruck von der Publikationstätigkeit einer Forschungseinheit zu vermitteln. In den quantitativen Auswertungen dieser Publikationslisten wurde dem Indikator „Aufsätze in peer-reviewed journals“ (s. A.III.4, S. 20) ein besonderes Gewicht zugewiesen, da diese Publikationen gutachterliche Bewertungen durchlaufen, an die sich das Peer Review der Bewertungsgruppe unmittelbar anschließen lässt. Die Veröffentlichung der in der Soziologie zahlreich produzierten Sammelbände wie auch von Monographien wird in Deutschland in der Regel nicht an unabhängige fachgutachterliche Kontrollen gebunden, sodass deren Zahl als Indikator fragwürdig ist. Vor allem die Monographien sollten aber bei der Gesamteinschätzung angemessen gewürdigt werden.
- Die Differenzierung nach dem Merkmal „Peer review“ ist auch für die Drittmittel geboten. Die Bewertungsgruppe hat Drittmitteldaten entsprechend differenziert, insgesamt aber zurückhaltender benutzt, als in vielen ministeriellen und universitären Entscheidungen über Mittelverteilungen üblich geworden ist; sie stellen für sich genommen Inputfaktoren dar, bei denen nicht sicher ist, in welchem Maße sie Aussagen über den Forschungsoutput erlauben.<sup>19</sup> Nur die Zahl der aufgrund von Peer-Review-Kontrollen vergebenen Drittmittelprojekte wurde bei der Bestimmung von Forschungsqualität benutzt. Dagegen wurden Drittmittelsummen bei der Bewertung der Kriterien II: Impact/Effektivität und III: Effizienz mit veranschlagt. Mitteleinwerbungen von Ministerien, Unternehmen und Verbänden wurden im Kriterium V: Transfer in gesellschaftliche Anwendungsbereiche als Indikatoren für Transferleistungen angerechnet.
- Der Bewertungsaspekt Reputation (Kriterium II: Impact) wurde nur auf Basis qualitativer Informationen bewertet. Es wäre wünschenswert, die Daten über Forschungspreise und Auszeichnungen sowie über Ämter in wissenschaftlichen Institutionen und Gremien quantifizieren und auf diese Weise systematischer berücksichtigen zu können. Da „Reputation“ auch frühere Leistungen fixiert hält, führt dieser Bewertungsaspekt dazu, den ansonsten auf wenige Jahre bezogenen Erhebungszeitraum an dieser Stelle wenigstens indirekt auszuweiten.

---

<sup>19</sup> Jansen, Dorothea, et al. (2007): Drittmittel als Performanzindikator der wissenschaftlichen Forschung. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 59, S.125-149.

- Nicht verwendet wurde eine Reihe von Vernetzungsdaten wie etwa die Zahlen von Gastaufenthalten, Konferenzteilnahmen, Vorträgen etc. Angesichts der in den Wissenschaften sowieso steigenden Kontakthäufigkeiten besteht die Gefahr, dass der Einschluss solcher Aktivitäten bei der Bildung von Leistungsindikatoren eher eine kollegiale Betriebsamkeit als wissenschaftliche Leistungen abbilden und befördern würde.
- Bei der Beurteilung von Kriterium IV: Nachwuchsförderung wurde eine relativ starke Gewichtung geförderter Promotionsprogramme (Graduiertenkollegs etc.) mit der Begründung vorgenommen, dass mit ihrer Institutionalisierung der Status der Doktoranden gesichert und der Einsatz der Betreuer stärker kontrollierbar wird. Solche Programme zeichnen sich für Ratingzwecke auch dadurch aus, dass ihrer Einrichtung in der Regel Begutachtungen durch Peers vorausgehen. Die Bewertungsgruppe hält eine Standardisierung dieses Indikators für wünschenswert und möglich, ohne zu verkennen, dass auch außerhalb institutionalisierter Programme die Doktorandenförderung gelingen kann; das lässt sich an anderen Daten (siehe Qualitätsmodell im Anhang) prüfen.
- Überwiegend ungelöst blieb die Indikatorenfrage für die Dimension Transfer (Kriterium V: Transfer in andere gesellschaftliche Bereiche und Kriterium VI: Wissensvermittlung und -verbreitung). Eine Reihe von Aktivitäten ließ sich zwar erheben, deren Standardisierbarkeit erschien allerdings problematisch. Die Bewertungsgruppe hat mit Blick darauf die vom Wissenschaftsrat ursprünglich vorgesehene Zahl diesbezüglicher Kriterien von vier auf zwei reduziert, empfiehlt aber für Fächer wie die Soziologie, diese weiter auf ein einziges Transferkriterium zu reduzieren, um die unterschiedlich starke Ausprägung der Praxisorientierung von Forschungseinrichtungen differenziert (fünfstufige Notenskala) bewerten zu können (vgl. A.II.2, S. 12 f.).

Die vollständige Systematisierung und Operationalisierung der kriterienbezogenen Indikatoren ist in der Bewertungsmatrix („Qualitätsmodell“) ausgedrückt, welche die Bewertungsgruppe Soziologie für ihre Zwecke konstruiert hat.

## **II.2. Zur Erfassung der Forschungseinheiten**

Die Erfassung der Forschungseinheiten, die für die Bestimmung des Zentralkriteriums „Forschungsqualität“ als Erhebungs- und Analyseeinheiten dienen, stellte eines



der Hauptprobleme sowohl in der Datenerhebung als auch in der Bewertung der Forschungsleistung dar. Soziologische Organisationseinheiten in den Universitäten sind in der Regel sowohl in Ausrichtung als auch Größe nach Lehrgesichtspunkten eingerichtet. „Forschungseinheiten“, mit jeweils gemeinsamer Forschungsplanung und abgestimmten Projektgestaltungen, sind oberhalb der Ebene der Professuren hingegen bislang selten institutionalisiert. Daraus ergaben sich für die Betroffenen in der Befragung trotz einiger Vorgaben bei der Erhebung unkontrollierbare Spielräume für ad-hoc-Entscheidungen über den **Zuschnitt** der Erhebungseinheit, sodass auch taktische Kalküle eine Rolle spielen konnten. Dies war aber nur in begrenztem Maße der Fall. Die Hälfte der Professoren hat sich nicht mit anderen zusammen zum Forschungsrating angemeldet, sodass etwa drei Viertel der Forschungseinheiten aus nur einer Professur und deren Mitarbeiter(n) bestand, was die überwiegend kleinteilige Infrastruktur der soziologischen Forschung an den deutschen Universitäten reflektiert. Andererseits standen diesen einige größer geschnittene Forschungseinheiten gegenüber (10 Forschungseinheiten mit fünf und mehr Professoren), ohne dass über die eingereichten Daten nachvollziehbar gewesen wäre, ob und in welchem Maße es sich tatsächlich um Aggregate mit signifikant ausgeprägter Kooperation und nicht um bloße Artefakte handelt. Um grobe Verzerrungen durch die **Größenunterschiede** von Forschungseinheiten bei der Datenanalyse möglichst zu vermeiden, wurde während der Pilotstudie beschlossen, für die Forschungseinheiten auch relative Drittmitteldaten und relative Publikationszahlen (Drittmittelprojekte und -summen sowie Publikationen je Wissenschaftler-Vollzeitäquivalente, letztere nach Publikationstypen differenziert) zu berechnen und in der Urteilsbildung zu berücksichtigen. Das wird in Folgeerhebungen von Anfang an und systematischer geschehen können.

Angesichts der erkennbar gewordenen Probleme bei der Bestimmung von Forschungseinheiten hat die Bewertungsgruppe erwogen, ob auf deren Erhebung zugunsten größerer Aggregate verzichtet werden sollte. Mehrere Gesichtspunkte lassen sich dafür anführen. (1) Der Erhebungs- und Bewertungsaufwand ließe sich vermindern, wenn als Erhebungs- und Bewertungseinheiten in der Regel die Seminare und Institute dienen würden, die in der Pilotstudie meist mit der „Einrichtung“ identisch waren. (2) Die auch für Manipulationen offenen Spielräume bei der Definition von Forschungseinheiten würden verschwinden. (3) Ebenso ließe sich das datenschutzrechtliche Problem reduzieren, dass die Bewertung kleinster Forschungseinheiten der namentlich identifizierbaren Evaluation von Einzelpersonen nahekommen

kann. (4) Überdies ließe sich spekulieren, dass die Zusammenfassung von Forschungseinheiten zu einer Forschungseinrichtung der kollektiven Identität der Soziologie am jeweiligen Ort dienlich wäre und dann auch die Selbstkontrollprozesse innerhalb der Einrichtung verstärken würde.

In den Abwägungen der Bewertungsgruppe haben sich diese Gesichtspunkte nicht durchgesetzt. Mehrheitlich war die Bewertungsgruppe der Ansicht, dass bei der Bestimmung von Forschungsqualität und darüber hinaus künftig auch bei der Einschätzung der Transferleistungen auf Forschungseinheiten rekuriert werden solle. Der entscheidende Gesichtspunkt wurde durch die empirischen Befunde des Forschungsratings verstärkt: Forschungseinheiten und nicht Forschungseinrichtungen sind in deutschen Universitäten die primären kollektiven Leistungsträger für Forschung und Forschungsvermittlung. Die festgestellten Heterogenitäten der Forschungseinheiten innerhalb der jeweiligen Einrichtung würden auf der Aggregatebene der Einrichtung künstlich nivelliert erscheinen. Das ist für die Dimension Nachwuchsförderung nicht in gleichem Maße der Fall (siehe Bewertungsmatrix, Anhang 2). Da die Leistungen von Forschungseinheiten in den anderen Dimensionen innerhalb der Forschungseinrichtungen stark variieren, müssten die durch Aggregation von Einheiten entstehenden „Durchschnittswerte“ sehr invalide sein. Die überwiegend kleinteilige Produktionsstruktur der deutschen Soziologie sollte im Verfahren des Forschungsratings nicht wegdefiniert werden. Stellt man die Erhebungen und Bewertungen also, zwar nicht vollständig, wohl aber teilweise auf Forschungseinheiten ab, so muss dann allerdings effektiver dafür gesorgt werden, dass ihre Festlegungen in den Forschungseinrichtungen durch Vorgaben stärker strukturiert, die jeweiligen Entscheidungen begründet und die Begründungen durch die Bewertungsgruppe geprüft werden.

### **II.3. Zur Datenerhebung**

Die Erhebung wurde im Fach Soziologie nach dem Prinzip „Current Potential“ durchgeführt. Damit wurde das Ziel verfolgt, die aktuelle Leistungsfähigkeit einer Einrichtung am Stichtag abzubilden. Das „**Current Potential**“-Prinzip bedeutet, dass die Forschungsleistung eines Wissenschaftlers immer der Forschungseinheit und Einrichtung zugerechnet wird, an dem er am Stichtag 31.12.2005 tätig war. Wechselt ein Wissenschaftler innerhalb des Erhebungszeitraumes, aber vor dem Stichtag an eine

andere Einrichtung, werden die zuvor erbrachten Leistungen der neuen Einrichtung zugerechnet. Geht ein Wissenschaftler vor dem 31.12.2005 in den Ruhestand, wird er nicht erfasst. Dieser Punkt war aus der Sicht jener Einrichtungen besonders kritisch, die kurz vor dem Stichtag Wissenschaftler emeritiert haben.

Das „Current Potential“-Prinzip erscheint bei einer Analyse von Institutionen weniger robust gegenüber einigen Zufallseffekten, bspw. durch temporär unbesetzte Stellen, als die Alternative der von der Bewertungsgruppe Chemie praktizierten „Work Done At“-Erhebung, da ein Wissenschaftler seine Leistungen bei einem Stellenwechsel gleichsam an eine neue Arbeitsstelle „mitnimmt“. Es ist dafür aber besser geeignet, den aktuellen Stand der Leistungsfähigkeit einer Einrichtung widerzuspiegeln und ihr für strategische Entscheidungen relevantes Entwicklungspotential zu erfassen, da es im Unterschied zum „Work Done At“-Prinzip nur die am Stichtag tatsächlich vorhandenen Wissenschaftler berücksichtigt. Die Bewertungsgruppe spricht sich daher mehrheitlich für eine Beibehaltung des „Current Potential“-Prinzips aus, um die Prognosefähigkeit des Ratings zu unterstreichen.

Unabhängig vom Erhebungsprinzip ist es wichtig, durch eine gute Vorbereitung der Datenerhebung und -auswertung die Zeit zwischen dem für die Erhebung gewählten Stichtag und der Bewertung der Daten bzw. der Veröffentlichung der Ergebnisse zu verkürzen. Dieser Zeitraum war in der Pilotstudie mit mehr als zwei Jahren zu lang. Die Erhebungsdauer könnte etwa dadurch verkürzt werden, dass den Einrichtungen eine **längere Vorlaufzeit** eingeräumt würde. Das Erhebungsprogramm mit Angabe der Regeln, nach denen die dokumentierten Leistungen bewertet werden, möglichst auch die Fragebogenformate, sollten den potentiell teilnehmenden Einrichtungen mit einem ausreichenden Vorlauf vor dem Stichtag bekannt gegeben werden, sodass die Daten im laufenden Betrieb gesammelt werden könnten und nicht vollständig im Nachhinein rekonstruiert werden müssten. Optimal wäre eine Vorlaufzeit entsprechend dem Erhebungszeitraum, sie sollte aber mindestens mehrere Monate betragen, besser 2-3 Jahre. Es entstände in den Einrichtungen eine größere Erwartungssicherheit, und es wäre damit zu rechnen, dass sich die Erhebungsdauer verkürzen und der Erhebungsaufwand durch Vorsorge verringern ließe.

Der für die Datenerhebung gewählte **Erhebungszeitraum** betrug fünf Jahre. Dieser Zeitraum sollte nach Einschätzung der Bewertungsgruppe Soziologie nicht unter-

schritten werden. Eine Verkürzung des Erhebungszeitraums würde zwar die Datenmenge sowie die Notwendigkeit, Daten über mehrere Jahre vorzuhalten, und damit den Erhebungsaufwand etwas verringern. Zugleich würde dies jedoch die Gefahr mit sich bringen, dass zyklische Schwankungen in der Forschungsleistung nicht ausgeglichen und die Resultate mehrjähriger Projekte nicht angemessen gewürdigt werden können. Auch sollte die mit dem Forschungsrating verbundene Kontrolle der Forschungseinheiten zeitlich nicht zu stark verdichtet werden. Bei einer in den Sozial- und auch den Geisteswissenschaften möglicherweise sinnvollen Verlängerung des Bewertungsturnus auf sechs bis sieben Jahre ist natürlich zu bedenken, dass der Erhebungszeitraum und der Turnus der Erhebungen aufeinander abgestimmt sein sollten. Andernfalls würde es zu Ungerechtigkeiten dadurch kommen, dass der Abschluss längerer Projekte oder die Publikation einer Monographie zufällig in den nicht erfassten Zeitraum fallen könnte.

Deutlichen Optimierungsbedarf gibt es bei der Gestaltung und technischen Handhabung der **Fragebögen**. Häufig hatten Fachkoordinatoren Schwierigkeiten bei der Übertragung bereits vorhandener Daten in die teilweise geschützten Formulare, die Trennung in tabellarische und textförmige Antwortmöglichkeiten sowie inhaltlich begründete Beschränkungen etwa von Listen auf eine bestimmte Anzahl von Einträgen bemängelt. Gerade diese technischen Umstände haben teilweise zu Verstimmungen und einer negativen Grundhaltung gegenüber der durch das Rating verursachten zusätzlichen Arbeitsbelastung geführt. Bei einer Wiederholung oder Verstetigung des Verfahrens sollte die technische Realisierung des Erhebungsinstruments und die Datensammlung einem spezialisierten Anbieter übertragen werden. Die Vorbereitung der Erhebung sowie die Kontrolle und Aufbereitung der Daten sollten indes weiterhin der Geschäftsstelle des Wissenschaftsrats obliegen, die das Verfahren organisiert und einerseits mit den Gutachtern und andererseits mit den Einrichtungen ständig kooperiert. Eine denkbare Vereinfachung der Erhebung wäre ein **Online-Erhebungsinstrument**. Dies war im Rahmen der Pilotstudie aus zeitlichen Gründen nicht möglich, wäre aber ggf. bei einer Wiederholung oder Verstetigung des Verfahrens erneut auf seine Machbarkeit zu überprüfen.

Die Erfahrungen aus der Pilotstudie haben außerdem gezeigt, dass die möglichst präzise, aber zugleich nicht zu einengende **Formulierung der Fragen** für die Homogenität und Vergleichbarkeit der Daten einerseits und die individuellen Ausdrucks-

möglichkeiten der Einrichtungen andererseits entscheidend ist. Zu offene Fragen führen zu sehr heterogenen und teilweise abschweifenden Angaben, zu eng definierte Fragen provozieren Kritik, wenn Einrichtungen glauben, sich nicht adäquat darstellen zu können. Hier wird es jeweils Aufgabe der Bewertungsgruppe sein, in der fachspezifischen Entwicklung der Fragebögen das richtige Maß zu finden und damit zur Sicherung einer hohen Qualität der Datenbasis beizutragen (zu einzelnen Indikatoren s. Anhang 1, S. 63 ff.).

#### **II.4. Zur Publikationserhebung und bibliometrischen Analyse**

Die Einrichtungen wurden gebeten, die in der Datenbank SOLIS und sechs Datenbanken von CSA<sup>20</sup> recherchierten Publikationen und deren Zuordnung zu einzelnen Wissenschaftlern zu überprüfen und ggf. zu korrigieren. Die **Publikationslisten** wurden dazu vom Informationszentrum für Sozialwissenschaften (IZ) in einem Online-System für die einzelnen Wissenschaftler zugänglich gemacht. Jeder Wissenschaftler erhielt über seinen Fachkoordinator ein Passwort, mit dem er seine Publikationsliste korrigieren konnte. Begleitet wurde diese Überarbeitungsphase durch eine ausführliche Bedienungsanleitung und die Möglichkeit von E-Mail-Anfragen, von der 16% aller Wissenschaftler Gebrauch machten.

Nur rund die Hälfte aller Wissenschaftler (52%) nutzte allerdings die Möglichkeit, die ihnen online zugänglich gemachten Publikationslisten um weitere Publikationen zu ergänzen. Unter den Professoren lag der Anteil bei etwas über zwei Dritteln. An einigen Einrichtungen wurden die Zugangsdaten durch den Fachkoordinator, wie sich später herausstellte, nicht weitergeleitet. Den Berichterstattern der Bewertungsgruppe wurde in den Datenberichten zur Kenntnis gegeben, wie viel Prozent der Wissenschaftler einer Forschungseinheit bzw. Einrichtung vom Online-Zugang zu ihren Publikationslisten Gebrauch gemacht hatten, um die Validität der Publikationslisten und der daraus gewonnenen quantitativen Indikatoren einschätzen zu können. Bei Wiederholungen des Forschungsratings sollte aber versucht werden, die Kontrolle der Publikationslisten durch die betroffenen Wissenschaftler zu verbreitern.

---

<sup>20</sup> Diese waren: Sociological Abstracts (Schwerpunkt: Soziologie und benachbarte Disziplinen), Social Services Abstracts (Schwerpunkt: Sozialarbeit und angrenzende Gebiete), PAIS International (Schwerpunkt: Politik, Verwaltung, Wirtschafts- und Sozialwissenschaften), Worldwide Political Science Abstracts (Schwerpunkt: Politikwissenschaften und verwandte Gebiete), ASSIA: Applied Social Sciences Index and Abstracts (Schwerpunkt: Gesundheits- und Sozialwesen, Soziologie, Psychologie, Wirtschaft, Politikwissenschaft und Erziehung), sowie Physical Education Index (Schwerpunkt: Sportsoziologie, Sportpsychologie, Sport und Gesundheit). Für jede dieser Datenbanken werden jährlich hunderte von Zeitschriften, Reihen, Rezensionen, Konferenzbände u. a. m. ausgewertet.

Die **Korrekturphase** zeigte sehr deutlich, dass die verwendeten Datenbanken keineswegs den gesamten Output der deutschen soziologischen Forschung abdecken. Vielmehr wurde der auf Basis der Datenbankrecherche eingefüllte Bestand von 5.332 Publikationen durch die den Kriterien entsprechenden Nachmeldungen verdoppelt auf 10.622 Titel (s. A.III.4, S. 20). Insgesamt wurden von den 52% der betroffenen Wissenschaftler, die sich beteiligten, 9.123 Titel über die Online-Eingabemaske des IZ nachgemeldet, von denen allerdings nur 5.290 (58%) den aufgestellten Kriterien entsprachen und auch in die Publikationslisten aufgenommen wurden. 3.833 nachgemeldete Publikationen (42 %) entsprachen nicht den zuvor bekanntgegebenen Kriterien, da sie entweder nicht aus dem relevanten Zeitraum 2001-2005 stammten (39%), bereits auf der Publikationsliste enthalten waren (33%) oder zur so genannten Grauen Literatur gehörten (28%).

**Tabelle 5: Publikationsdaten**

Herkunft der Publikationsdaten	N	%
Voreingefüllt in IZ Online-System aus SOLIS und CSA-Datenbanken	5.332	50,2%
Von Wissenschaftlern ergänzte Publikationen	9.123	-
davon regelkonform	5.290	49,8 %
Gesamtbestand	10.622	100%

Die nachträglich in die Publikationslisten aufgenommenen Titel bestanden zu einem Viertel aus internationalen Veröffentlichungen, die in der SOLIS-Datenbank des IZ generell nicht erfasst sind und auch über die CSA-Datenbanken (vgl. A.III.4, S. 20) nicht vollständig ermittelt werden konnten. Da keine andere einzelne umfassendere Datenbank mit deutscher soziologischer oder geistes- und sozialwissenschaftlicher Forschungsliteratur vorhanden ist, bleibt auch künftig nur die **Kombination unterschiedlicher Datenbanken**, um den Publikationsoutput eines Faches wie der Soziologie zu erfassen. Umstritten war unter den Einrichtungen und auch innerhalb der Bewertungsgruppe der Verzicht auf so genannte Graue Literatur, die nicht verlagsgebunden erscheint. Obwohl diese Literatursorte in der Pilotstudie ausdrücklich ausgeschlossen und auf eine Regeländerung im laufenden Verfahren verzichtet worden war, wurden 1.062 solcher Titel über das Online-System manuell nachgemeldet, aber nicht in die Publikationslisten aufgenommen.

Das **Online-System** des IZ, das zur Kontrolle und Ergänzung der Publikationslisten zur Verfügung stand, sollte nach Ansicht vieler Nutzer künftig in der Handhabung weiter vereinfacht werden und künftig stärker selbsterklärend funktionieren. Ein Desiderat bleibt die Entwicklung einer einfach zu bedienenden Schnittstelle, die den Import eigener Literaturangaben in solch ein Online-System erleichtert. Die Alternative zur Verwendung einer einheitlichen Datenbank, nämlich die Einreichung beliebig formatierter Publikationslisten durch die einzelnen Forschungseinheiten, erscheint nicht praktikabel, da diese weder systematisch kontrollierbar noch nach unterschiedlichen Kriterien auszähl- und sortierbar wären, um quantitative Indikatoren für das „Informed Peer Review“-Verfahren zu gewinnen. Ebenfalls zu verbessern ist die Zuordnung von Publikationen zu Publikationstypen, die Voraussetzung einer möglichst fehlerfreien Bildung quantitativer Indikatoren ist. Hier waren häufig aufwändige Überprüfungen durch die einzelnen Berichtersteller nötig, die den Arbeitsaufwand der Bewertungen wesentlich gesteigert haben.

Die Bereitstellung möglichst vollständiger und fehlerfreier Publikationslisten ist nach Ansicht der Bewertungsgruppe auch für Forschungsratings anderer Fächer, die sich kaum auf Zitationsdaten verlassen können, sondern stärker auf die qualitative Prüfung von Publikationslisten mit unterschiedlichen Publikationstypen stützen müssen, von zentraler Bedeutung. Der Weiterentwicklung entsprechender Instrumente in Zusammenarbeit mit Fachinformationszentren und anderen Sachverständigen sollte große Bedeutung zugemessen werden.

Zusätzlich zur qualitativen Begutachtung der Publikationslisten und deren Auswertung für quantitative Indikatoren hat die Bewertungsgruppe Soziologie beschlossen, in größerem Umfang Gebrauch von „**lektürebasiertem Bewertungswissen**“ zu machen. Die Lektüre von meist mehr als 3.000 Textseiten je Gutachter (vgl. A.III.4, S. 20 ff.) ermöglichte ein unmittelbar qualitatives Urteil über die Forschungsleistung einer gegebenen Einheit, erhöhte aber wesentlich den Bewertungsaufwand für die einzelnen Gutachter. Die Einschätzungen der eingereichten Texte erwiesen sich für die Benotung von Forschungsqualität als wichtigster Einzelfaktor, und die Bewertungsgruppe ist auch im Hinblick auf Wiederholungen des Verfahrens der Auffassung, dass in Fächern wie der Soziologie die Textlektüre unter dem Gesichtspunkt einer validen Bewertung der Forschungsqualität unverzichtbar ist. Die Entlastung der Gutachter sollte vor allem durch eine höhere Verlässlichkeit der Bewertungsgrundla-

ge und durch eine verbesserte Datenqualität insbesondere bei den Publikationslisten und -analysen erreicht werden, damit sie für deren Nutzung nicht auf eigene zusätzliche Recherchen angewiesen sind.

Die Bewertungsgruppe hat sich nach ausführlicher Beratung und Nutzung externer Expertise entschlossen, bei ihrem Forschungsrating auf eine umfassende **Zitationsanalyse** zu verzichten und sich auf die Erhebung der Publikationen zu konzentrieren (vgl. Anhang 11). Nur für die potentiell besten Forschungseinheiten, die von den jeweiligen Berichterstatlern mit den Noten „4 sehr gut“ oder „5 exzellent“ vorbewertet worden waren, wurde eine Zitationsanalyse versuchsweise beauftragt (vgl. A.III.4, S. 20 f.). Die Bewertungsgruppe sah sich bei der Einschätzung ihrer Ergebnisse mehrheitlich in ihrer Skepsis gegenüber dem umfassenden Einbezug bibliometrischer Daten in das Indikatorenset des Forschungsratings Soziologie bestätigt, da unter den gegebenen Umständen und mit den vorhandenen Instrumenten nur für die bestbewerteten Forschungseinheiten überhaupt eine analysierbare Zahl an Zitationen aufgewiesen werden kann. Für alle übrigen Forschungseinheiten diskriminieren die Zitationsdaten zu wenig, um belastbare Aussagen auf dieser Datenbasis zu ermöglichen. Zudem ist die extrem hohe Diskrepanz zwischen den durchschnittlichen Zitationsraten der unterschiedlichen Teilgebiete der Soziologie mit den verfügbaren Methoden nicht adäquat auszugleichen, sodass eine Verwendung dieser Daten zu Verzerrungen innerhalb der Soziologie führen würde.

## **II.5. Zur Datenaufbereitung**

Die Geschäftsstelle hat alle erhobenen Daten darauf überprüft, ob die Erhebungsregeln eingehalten wurden und die Angaben in sich konsistent waren. Die häufigsten Gründe für Rückfragen an die Fachkoordinatoren waren lückenhafte Angaben sowie offenkundige Verletzungen teils trivialer Regeln, die in den Fragebögen explizit formuliert waren – der häufigste Fehler war, dass Angaben zu Aktivitäten gemacht wurden, die nicht in den Erhebungszeitraum fielen (vgl. A.III.3, S. 19 f.). Auch wenn diese Probleme mit einer Routinisierung und steigenden Akzeptanz eines solchen Verfahrens geringer werden, zeigt dies, wie wichtig auf der einen Seite ein komfortables und übersichtliches Erhebungsinstrument, auf der anderen Seite aber auch eine sorgfältige Datenkontrolle ist.



Bei einzelnen Fragen gab es in den Einrichtungen fachintern offenkundig Auffassungsunterschiede, die seitens der Ausfüllenden bei offenen Fragen zu einem bisweilen ausufernden Meldeverhalten führten. Soweit möglich, wurde versucht, dies durch Präzisierungen der Regeln im laufenden Verfahren (bspw. durch die Beschränkung auf exemplarische Angaben) zu korrigieren. Die abschließende **Datenkontrolle** durch die Geschäftsstelle konnte solche inhaltlichen Missverständnisse nur begrenzt bereinigen, da es dazu einer innerfachlichen Entscheidung bedurft hätte. Die Gutachter der Bewertungsgruppe sollten künftig frühzeitiger in die Datenkontrolle eingebunden werden. Nach der Kontrolle der Daten und den fälligen Nacherhebungen wurden die Ergebnisse für jede Einrichtung in so genannten Datenberichten zusammengefasst, in denen die Rahmeninformationen und die Indikatoren in einen Sinnzusammenhang gemäß der definierten sechs Kriterien gestellt wurden. Diese Datenberichte wurden den Einrichtungen zur Abschlusskontrolle vorgelegt. Allerdings hatte eine große Kulanz in der Erhebungsfrist dazu geführt, dass diese Rückkopplungsschleife sich in einigen Fällen auf eine Woche verkürzte. Dieser Rückkopplung muss künftig mehr Zeit eingeräumt werden.

Erst im Anschluss an die Rückkopplung und nach Einarbeitung evtl. notwendiger Änderungen konnten die quantitativen Daten für die Grundgesamtheit ausgewertet werden und Lagemaße (genauer: Perzentilwerte) ebenfalls in die **Datenberichte** eingefügt werden. Weitere Verteilungsmaße (1. und 3. Quartil, Median) wurden in dem Leitfaden zu den Datenberichten mitgeteilt und durch statistische Analysen ergänzt (s. A.III.5, S. 23). Nach Ansicht der Bewertungsgruppe Soziologie hat sich die Umrechnung der quantitativen Daten in skaleninvariante Lagemaße grundsätzlich bewährt. Allerdings ist darauf hinzuweisen, dass solche Werte im Bewertungsprozess eine Komplexitätsreduktion darstellen, die in den Einzelfällen reflektiert werden muss. Beispielsweise kann es vorkommen, dass die Bedeutsamkeit der Perzentildifferenzen überschätzt wird, wenn die im Leitfaden enthaltenen **statistischen Analysen**, die auch die Streuung der zugrunde liegenden Messwerte wiedergeben, nicht mit berücksichtigt werden. Auch aus diesem Grunde warnt die Bewertungsgruppe Soziologie vor der Berechnung hoch aggregierter Zahlen, etwa gewichteter Gesamtindikatoren für einzelne Kriterien, wenn nicht gleichzeitig Zusatzinformationen zur Verfügung stehen, mit denen diese fallspezifisch validiert werden können.

## II.6. Zur Bewertungsphase

Die Bewertung wurde durch eine 16-köpfige Gruppe von **Fachvertretern** vorgenommen, die die meisten Teilgebiete der Soziologie mit ihrer Expertise abdeckte. Allerdings war in Einzelfällen, etwa bei kleinen Spezialgebieten, oder bei Befangenheiten, der Rückgriff auf externe Sondergutachter notwendig. Es erwies sich als empfehlenswert, auf wenige Sondergutachter zurückzugreifen, die mit entsprechenden Vergleichsmöglichkeiten jeweils für mehrere Bewertungen zuständig sein sollten.

Für die Bewertungen wurden zwei maßgebliche **Regeln** aufgestellt: Einzelne Daten und Bewertungsaspekte wurden als primär relevant ausgezeichnet, sodass diesen bei der Urteilsbildung im Zweifelsfall Vorrang zu geben war.<sup>21</sup> Zudem wurden Perzentilbereiche den einzelnen Skalenstufen zugeordnet. Eine gewisse Spezifizierung der Bewertungsmaßstäbe wurde dann in der plenaren Bewertung vorgenommen. Die fälligen Abwägungen hat die Bewertungsgruppe zuerst über die jeweiligen Gutachter, sodann in plenaren Diskussionen mit relativ großem Aufwand betrieben. Die Notengebung ist deshalb Resultat eines mehrstufigen Interaktionsprozesses. Dies führte dazu, dass die Benotungen der Einzelfälle kollektiv abgesichert und weitgehend konsensual waren. Es entspricht dem von der Bewertungsgruppe praktizierten Prinzip des „Informed Peer Review“, dass qualitative Abwägungen aller vorliegenden Informationen am Ende entscheidend waren. Dies spricht nicht gegen eine fortschreitende Standardisierung von Datennachfrage und Indikatorenbildung. Diese erscheint insofern wünschenswert, als sie die Zuverlässigkeit gutachterlicher Entscheidungen erhöhen und zur Ökonomie des Verfahrens beitragen kann. Die Frage ist, wie weit Standardisierungen durchgeführt werden können. Im Hinblick darauf wurde auch eine quantitative Festlegung von Gewichtungen der einzelnen Indikatoren angeregt und an bestimmten Beispielen erprobt. Für die Mehrheit der Bewertungsgruppe waren die dabei erkennbaren Validitätsprobleme aber ausschlaggebend, um quantitative Gewichtungen nicht in das Verfahren einzuführen.

Die fünfstufige **Bewertungsskala** entspricht internationalen Beispielen und hat sich im Wesentlichen bewährt. Dass die Notenstufen mit Ausnahme der Spitzennote weder semantisch noch durch eine vorgegebene Verteilung näher zu bestimmen waren, war im Bewertungsprozess unproblematisch. Dass die Skala bei den Kriterien Trans-

---

<sup>21</sup> Diese waren: „Qualität des Outputs“ für das Kriterium Forschungsqualität, sowie „Publikationsaktivität“ im Kriterium Impact/Effektivität, vgl. Bewertungsmatrix unter: [http://www.wissenschaftsrat.de/texte/Bewertungsmatrix\\_Soz.pdf](http://www.wissenschaftsrat.de/texte/Bewertungsmatrix_Soz.pdf).

fer in gesellschaftliche Anwendungsbereiche sowie Wissensvermittlung und -verbreitung auf drei Stufen reduziert wurde, schränkt die Vergleichbarkeit der Kriterien ein, erschien aber wegen des heterogenen und lückenhaften Datenmaterials in diesen Fällen unumgänglich. Die Standardisierung einiger Messgrößen in dieser Dimension sowie eine Zusammenlegung der beiden Transferkriterien würde auch hier eine differenziertere Skalenbildung erlauben (vgl. Anhang 1).

Die Grundlage der Bewertung waren die **Datenberichte**; alle Gutachter erhielten alle Datenberichte. Dies stellte sicher, dass in der plenaren Bewertung alle Mitglieder der Bewertungsgruppe auf der gleichen Informationsbasis diskutieren konnten. Die von der Geschäftsstelle erstellten Datenberichte wurden in der vorgelegten Form von den Gutachtern als übersichtlich und gut strukturiert empfunden. In den Bewertungssitzungen wurden nach Bedarf zusätzliche Auswertungen vorgelegt.

Kurze **Kommentare** zu den Bewertungen für die Öffentlichkeit oder für die betroffene Einrichtung wurden nur in Einzelfällen festgehalten, etwa dann, wenn die Bewertung durch außergewöhnliche Bedingungen stark beeinflusst wurde. Die Klassifizierung als „nicht bewertbar“ wurde zumindest für die Einrichtung selbst immer kommentiert. Übereinstimmung bestand dahingehend, Kommentare ausschließlich als Hinweis auf Besonderheiten zu benutzen. Weitergehende Analysen, insbesondere Vermutungen darüber, was die Leistungsfähigkeit der bewerteten Einrichtung erklärt, sind in Forschungsratings ohne eine umfassendere Einzelevaluation nicht zu substantiieren.

Bei der Bewertung zeigte sich eine unterschiedlich hohe **Belastbarkeit** der einzelnen Kriterien. Die Bewertung der Kriterien der Dimension Forschung (Kriterium I: Forschungsqualität, Kriterium II: Impact/Effektivität, Kriterium III: Effizienz) erscheint der Bewertungsgruppe fundierter als die Bewertung der übrigen Kriterien, für die weniger geeignete Indikatoren zur Verfügung standen. Am wenigsten differenziert ist die Bewertung der Kriterien in der Dimension Transfer, vor allem aufgrund einer schmalen und heterogenen Datengrundlage, die nur einen quantitativen Indikator umfasste, sodass es der Bewertungsgruppe folgerichtig erschien, die Notenskala von 5 auf 3 Stufen zu reduzieren (vgl. Anhang 1, Abschnitte V und VI, S. 71 ff.).

## **II.7. Zur Veröffentlichung**

Es entspricht ihrer allgemeinen Einschätzung der Funktionsvoraussetzungen des Forschungsratings, dass die Bewertungsgruppe nicht nur eine Veröffentlichung ihrer Befunde, sondern auch der Verfahren, mit denen sie erreicht wurden, durch den Wissenschaftsrat befürwortet.

Die vorgesehene Veröffentlichungsform in einer Pressekonferenz, verbunden mit einem Pressegespräch, an dem auch ein Vertreter der Deutschen Gesellschaft für Soziologie teilnehmen wird, sowie ein begleitender Internet-Auftritt werden es ermöglichen, eine breite Öffentlichkeit zügig zu erreichen. Für die Akzeptanz des Verfahrens in der Wissenschaft wird es im Übrigen hilfreich sein, dass die Einrichtungen ihre eigenen Ergebnisse und den endgültigen Datenbericht mit statistischen Informationen zu den sie betreffenden Daten etwa eine Woche vor der Veröffentlichung erhalten werden.

Es sollte künftig möglich sein, auch die auf der Ebene der Forschungseinheiten vorgenommenen Bewertungen der Forschungsqualität der Öffentlichkeit namentlich zugänglich zu machen. Im Rahmen der Pilotstudie werden diese differenzierten Befunde nur den Einrichtungen intern zur Verfügung gestellt, da die Veröffentlichung personenbeziehbarer Bewertungen ohne Zustimmung der Betroffenen datenschutzrechtlich bedenklich ist. Die Detailbewertungen auf der Ebene der Forschungseinheiten machen jedoch einen wesentlichen Teil des Nutzens des Forschungsratings aus; sie könnten auch für externe Adressaten von Interesse sein, etwa als Orientierungshilfe für zukünftige Doktoranden oder für Institutionen, die Ansprechpartner zu einzelnen Problemfeldern suchen. Aufgrund der überwiegend kleinteiligen Kooperationsstrukturen in der soziologischen Forschung wird die Bildung personenbezogener Forschungseinheiten auch künftig nicht zu vermeiden sein. Deshalb sollten die Einrichtungen von vorneherein darauf aufmerksam gemacht werden, dass die Bewertungen der Forschungseinheiten veröffentlicht werden. Unter diesen Voraussetzungen kann den Einrichtungen zugemutet werden, die betroffenen Wissenschaftler anzuhalten, dazu ihr Einverständnis zu erteilen.

Die Bewertungsgruppe spricht sich auch dafür aus, dass der Wissenschaftsrat die Basisdaten aus der Pilotstudie für wissenschaftliche Zwecke, insbesondere für Vorhaben der Evaluationsforschung zur Verfügung stellt, wobei die Anonymisierung des

Materials gewährleistet sein muss. Sekundäranalysen des für die Pilotstudie erhobenen Materials könnten neue, Erkenntnisse zur Weiterentwicklung der Leistungsindikatoren in den Sozial-, aber auch den Geisteswissenschaften erbringen und so zur Verbesserung der Bewertungsverfahren beitragen.

### **B.III. Zum weiteren Vorgehen**

#### **III.1. Aufwand und Ertragserwägungen**

Die Bewertungsgruppe Soziologie hält es grundsätzlich für wünschenswert, ein Forschungsrating unter der Kontrolle des Wissenschaftsrats zu routinisieren. Voraussetzung ist, dass bestimmte Bedingungen zur Qualitätssicherung von Forschungsratings erfüllt werden (darüber ausführlicher Abschnitt III.2, S. 57 ff.). Entscheidend wird sein, bei der Optimierung gleichzeitig auch die Machbarkeit des Verfahrens im Auge zu behalten, wobei in jedem Fall davon ausgegangen werden muss, dass akzeptable Forschungsratings ohne einen hohen Aufwand nicht durchführbar sind. Gleichwohl hält es die Bewertungsgruppe für nützlich, die Erfahrungen aus der Pilotstudie zu benutzen, um für künftige Forschungsratings auch einige Vorschläge zur Ökonomie des Verfahrens zu skizzieren.

Bei der Bewertung des Aufwandes des in der Pilotstudie erprobten Verfahrens ist zu beachten, dass der Pilotcharakter der Studie eine Reihe von experimentellen Komponenten und Absicherungsmaßnahmen notwendig gemacht hat, die bei einer Wiederholung entfallen und somit den Aufwand reduzieren würden. Die wesentlichen Träger des Aufwands sind die Gutachter, die bewerteten Einrichtungen im Hinblick auf die bei ihnen erfolgende Datenerhebung und -analyse, die Geschäftsstelle und die für die Publikationsanalyse wichtigen Datenbanken. Im Folgenden wird kurz angegeben, welche Faktoren den Aufwand an diesen Stellen hervorrufen und möglicherweise verändert werden können.

#### **a) Gutachter**

Die 16 Gutachter der Bewertungsgruppe Soziologie sind während der Pilotstudie im Zeitraum Februar 2006 bis März 2008 zu insgesamt elf Sitzungen (15 Sitzungstage) zusammengetroffen. Zu den Sitzungstagen kommt die individuelle Vorbereitung hinzu, die vor allem in der Bewertungsphase einen erhöhten Aufwand bedeutete. Die Gutachter schätzen ihren Einsatz für das Forschungsrating im Durchschnitt insge-

samt auf etwa zwei bis drei Arbeitsmonate pro Person. Ihnen wurde von ihren Heimatuniversitäten mehrheitlich eine Ermäßigung des Lehrdeputats angeboten. Dies sollte zumindest in Fächern, in denen eine Lektüre von Publikationen unverzichtbar ist, auch in Zukunft so gehandhabt werden.

Der hohe Arbeitsaufwand der Gutachter ist zum Teil dem Pilotcharakter der Studie geschuldet. So mussten zunächst grundlegende Operationalisierungsfragen geklärt, Fragebögen entwickelt und in einem Pretest erprobt sowie die Bewertungsskalen definiert werden. Bei einer Wiederholung des Verfahrens könnten mit Standardisierungen, die sich aus den Erfahrungen der Bewertungsgruppen ableiten lassen, der Aufwand der Operationalisierungsphase reduziert werden.

Der **Aufwand der Bewertungsphase** hängt vor allem von der Zahl der Forschungseinheiten, der Zahl der Kriterien und insbesondere der Qualität der Daten ab. In der Pilotstudie (254 Forschungseinheiten, sechs Kriterien) hatten 14 der 16 der Gutachter als Berichterstatter die Verantwortung durchschnittlich für etwa 33 Forschungseinheiten und etwa neun Einrichtungen. Insgesamt mussten knapp 80 Einzelbewertungen mit einer mehrfachen Zahl an Einschätzungen von Einzelinformationen von jedem Berichterstatter vorgenommen werden, bevor diese in Plenarveranstaltungen abschließend behandelt wurden. Um diesen Aufwand zu verringern, erscheint der Bewertungsgruppe folgendes naheliegend:

1. Es ist wünschenswert und möglich, die Zahl der Forschungseinheiten durch eine bessere Kontrolle ihrer Festlegungen durch die Forschungseinrichtungen etwas zu reduzieren (s. II.2, S. 40 f.). Dies wird besonders dringlich, wenn sie nach den Vorschlägen der Bewertungsgruppe als Erhebungs- und Bewertungseinheiten auch im Transferbereich berücksichtigt werden.
2. Der Wissenschaftsrat sollte prüfen, ob sich die von der Bewertungsgruppe jetzt schon von 9 auf 6 gekürzte Zahl der Kriterien weiter reduzieren lässt (s. II.1, S. 37 f.) Die Bewertungsgruppe spricht sich mehrheitlich für eine Zusammenlegung der beiden Transferkriterien V: Transfer in gesellschaftliche Bereiche und VI: Wissensvermittlung und -verbreitung aus, da beide Kriterien Leistungen der Wissenschaft an außerwissenschaftliche Akteure zum Gegenstand haben und – obwohl analytisch unterscheidbar – von den wissenschaftlichen Einrichtungen kaum differenziert werden.

3. Die Bewertungsgruppe schlägt im Übrigen vor, bei künftigen Forschungsratings zu prüfen, ob in der Dimension Forschung bei Beibehaltung der Zahl von drei Kriterien der Aufwand ihrer Messung dadurch reduziert werden kann, dass sowohl Impact/Effektivität als auch Effizienz für die Forschungseinrichtungen in stärkerem Maße rechnerisch, nämlich durch Aggregation aus den Daten abgeleitet werden, die auf der Ebene der Forschungseinheiten auch für das Kriterium Forschungsqualität erhoben werden.
4. Die Publikationsdaten und in einigen Punkten auch die Antworten der Einrichtungen waren von so heterogener und dabei teilweise unzulänglicher Qualität, dass die Gutachter sich zu einer unabhängigen Überprüfung der Angaben gezwungen sahen. Die gebotene Verbesserung der Datenqualität, insbesondere der Publikationslisten, könnte diesen Aufwand ganz erheblich reduzieren.

Maßgeblich auf den Aufwand wirkte sich vor allem der Entschluss der Bewertungsgruppe aus, von den Forschungseinheiten ausgewählte Publikationen anzufordern, um die Bewertung des Kriterium I: Forschungsqualität mit eigenen Lektüreeindrücken abzusichern. Dieser Aufwand ist zwar beträchtlich, lässt sich nach Auffassung der Bewertungsgruppe aber nicht ohne erhebliche Qualitätsverluste bei der Urteilsbildung wesentlich reduzieren.

#### **b) Bewertete Einrichtungen / Datenerhebung**

Die Datengrundlage, die für das „Informed Peer Review“-Verfahren benötigt wurde, musste überwiegend bei den betroffenen Einrichtungen erhoben werden. Die zuständigen Fachkoordinatoren haben den Aufwand der Datenerhebung unterschiedlich hoch eingeschätzt. Großen Einfluss auf den Aufwand hatten neben der Größe einer Einrichtung vor allem die Qualität des internen Controllings und die Unterstützung der Wissenschaftler seitens der Verwaltung.

Zu den wichtigsten Faktoren, die den **Aufwand bei den Einrichtungen** determinierten, gehört, dass der Inhalt der Befragung nicht vorab bekannt war, sodass ein Teil der Daten ad hoc erhoben und ex post rekonstruiert werden musste. Hinzu kamen technische Schwierigkeiten bei der Bedienung des Erhebungsinstrumentes. Alle diese Aspekte sind vor allem dem Pilotcharakter der Studie geschuldet. Eine **längere Vorlaufzeit** sowie die Entwicklung eines komfortableren Erhebungsinstrumentes sollten bei einer etwaigen Wiederholung des Verfahrens gewährleistet werden.

Zudem haben sich einige Daten in der Pilotstudie als wenig belastbar erwiesen, weshalb in Wiederholungsfällen von einer etwas **reduzierten Datennachfrage** der Bewertungsgruppe ausgegangen werden kann.

Ein kritisch gegen die Pilotstudie insgesamt vorgebrachtes Argument der Fachkoordinatoren war die von ihnen berichtete „Ratingmüdigkeit“ der Wissenschaftler. In Anbetracht zahlreicher Evaluationen sei es fraglich, ob sich der Aufwand für ein weiteres Verfahren überhaupt rechtfertigen lasse. Die Bewertungsgruppe geht allerdings davon aus, dass Forschungsratings des Wissenschaftsrats mittelfristig zu einer gewissen Standardisierung der Datennachfrage bei Evaluationen führen wird, sodass es für die Einrichtungen geraten erscheint, einen Mindestsatz an zentralen, kontinuierlich aktualisierten Datenbeständen anzulegen, die ohne großen Verwaltungsaufwand bedient werden können und für mehrere Zwecke einsetzbar sind.

### **c) Geschäftsstelle**

Die Pilotstudie wurde in der Geschäftsstelle des Wissenschaftsrats unterstützt von einer Projektgruppe mit insgesamt fünf Mitarbeitern (vier VZÄ), die parallel auch die Pilotstudie Chemie betreute. Neben der Organisation des gesamten Verfahrens wirkte diese Gruppe auch an der Entwicklung von Indikatoren mit und führte die Datenerhebung und -kontrolle durch. Eine ausreichende **Geschäftsstellenkapazität** ist für die Qualität der Verfahrensentwicklung und -organisation und der Datenerhebung essentiell und trägt auch zur Entlastung der Gutachter bei; sie sollte in vergleichbaren Fällen nach Auffassung der Bewertungsgruppe eher größer als kleiner festgelegt sein. Im Rahmen der Pilotstudie war es hilfreich, dass für jedes Fach mindestens ein wissenschaftlicher Mitarbeiter als Ansprechpartner zur Verfügung stand, der die Gutachter bei der Lösung fachspezifischer Probleme der Operationalisierung und die Fachkoordinatoren während der Datenerhebung unterstützen konnte.

### **d) Publikationsdatenbanken**

Publikationsdaten waren für die Pilotstudie Soziologie von zentraler Bedeutung. Es zeigte sich, dass zur Erhebung, Korrektur und Vervollständigung der Publikationslisten große Anstrengungen nötig waren, da die vorhandenen Datenbanken den Publikationsoutput der deutschen Soziologie nur sehr unvollständig abdecken und im Hinblick auf einen entsprechenden Datenservice bislang auch noch ungeübt sind. In er-



heblichem Umfang war die Mithilfe der bewerteten Wissenschaftler erforderlich, um die Publikationslisten zu vervollständigen. Ob und in welchem Maße in anderen Fächern Publikationsdatenbanken zur Verfügung stehen, deren Material hinreichend reichhaltig und zuverlässig ist, kann die Bewertungsgruppe Soziologie nicht ermes- sen. Sie weist aber darauf hin, dass deren Existenz für das Gelingen des For- schungsratings eine wichtige Voraussetzung ist.

### III.2. Qualitätsbedingungen des Forschungsratings

Die Bewertungsgruppe spricht sich dafür aus, dass in der Regie des Wissenschafts- rats künftig Forschungsratings durchgeführt werden, wenn dabei bestimmte Quali- tätsbedingungen erfüllt werden. Forschungsratings erzeugen eine **Transparenz**, die das eigenverantwortliche Verhalten der Forschungseinrichtungen auf eine verbindli- che Weise kontrollierbar machen und dadurch verbessern kann. Sie explizieren mit den Indikatoren, die sie für ihre Bewertungen einsetzen, diskutierbare und korrigier- bare Standards der Forschung, die das Verhalten der Forschungseinrichtungen ori- entieren können. Gegenwärtig finden, durch Ministerien, Universitätsleitungen und Medien veranlasst, Rankings wissenschaftlicher Leistungen zahlreich und nicht durchweg mit wünschenswerter Qualität statt; „Peer Review“ ist kein gängiger Be- standteil dieser Verfahren. Forschungsratings, deren Akteure und Abläufe von Wis- senschaftsrat kontrolliert werden, können demgegenüber unter Bedingungen stattfin- den, die geeignet sind, die Durchsichtigkeit des Verfahrens sowie die Qualität der Ergebnisse zu erhöhen. Folgendes ist dabei zu beachten.

- Forschungsratings sollten als Teil eines fortschreitenden Lernprozesses ver- standen werden, an denen die Einrichtungen der Wissenschaft sowie die Fachgesellschaften beteiligt werden. Ihre Beteiligung erscheint für die Akzep- tanz und Wirksamkeit der Ergebnisse sehr wichtig. Maßgeblich für wün- schenswerte Effekte ist das in den Disziplinen vorhandene Ausmaß von **Ver- trauen in die Kompetenz und Integrität** des Wissenschaftsrats und speziell der jeweils fungierenden Bewertungsgruppe. Das Rating selber bedarf inso- weit regelmäßig vertrauensbildender Maßnahmen, zu denen die Offenlegung des Verfahrens und Möglichkeiten der Datenkontrolle bei deren Erhebung durch die Betroffenen (Rückkopplungen) gehören, nicht zuletzt auch die Betei- ligung der Wissenschaftsorganisationen und Fachgesellschaften an der Be-

setzung der Bewertungsgruppe sowie Vorkehrungen zur Neutralisierung von Befangenheiten.

- Als unabdingbar für das Gelingen von Forschungsratings ist festzuhalten, dass das **Prinzip des „Informed Peer Review“** mit seiner doppelten Bindung für alle Verfahrensschritte maßgeblich ist. Einerseits braucht der Bewertungsprozess eine relativ große Menge von Informationen, und von diesen ist im Sinne der in B.II formulierten Standards zu fordern, dass die benötigten Daten zuverlässig erhoben und sorgfältig kontrolliert worden sind und sich zu möglichst instruktiven Indikatoren verdichten lassen. Dabei ist nicht zu vernachlässigen, dass auch die für die Bewertung relevanten Kontextbedingungen, unter denen die Einrichtungen arbeiten, transparent werden. Deren Berücksichtigung erscheint für die Validierung der Indikatoren wichtig, damit im „Peer Review“ unter Abwägung aller entscheidungsrelevanten Gesichtspunkte belastbare Einzelbewertungen möglich werden. Die Indikatoren orientieren diese Abwägung, können sie aber nicht vollständig determinieren. Dieses Prinzip erscheint im Falle der Soziologie noch grundlegender als für die im Pilotverfahren des Wissenschaftsrats parallel untersuchte Chemie, da die Aussagekraft der in der Soziologie verfügbaren quantitativen Indikatoren überwiegend schwächer ist.
- In den Sozial- und Geisteswissenschaften stellen **Publikationen** die wichtigsten Daten für die Bestimmung wissenschaftlicher Leistungen dar. Für solide Forschungsratings ist deshalb bei einer insgesamt umfangreich anzulegenden Materialerhebung von zentraler Bedeutung, dass die Publikationslisten korrekt und vollständig sind. Dabei ist darauf zu achten, dass die unterschiedlichen Textsorten, die für eine Disziplin zentral sind, erfasst werden. Für die Soziologie sind dies vor allem Zeitschriftenartikel (differenziert nach referierten und nicht referierten Zeitschriften) und Monographien.
- Die mit Hilfe von Publikationen und anderen Leistungsgrößen gebildeten Indikatoren ersetzen nicht die **Lektüre einer Mindestmenge von Texten** durch die Gutachter. Sie erscheint der Bewertungsgruppe in Fächern wie der Soziologie als eine notwendige Voraussetzung valider Urteilsbildung.
- Die Bewertungsgruppe Soziologie spricht sich dafür aus, dass im Forschungsrating die unterhalb der Forschungseinrichtungen zu bildenden **Forschungseinheiten** trotz des damit verbundenen Mehraufwandes der Bewertung als

Erhebungs- und Bewertungseinheiten auch künftig eine wichtige Rolle spielen. Größere Aggregate wie zum Beispiel die Einrichtungen (Universitäten bzw. ihre Seminare/Institute, außeruniversitäre Institute), die für pauschale Orientierungen durchaus einen Informationswert besitzen, stellen in aller Regel nicht die eigentlichen Leistungsträger dar. Das gilt sowohl für die Forschung als auch für den Transfer von Forschungsbefunden. Forschungseinheiten sollten mit ihren spezifischen Leistungen also sowohl bei der Forschungs- als auch bei der Transferdimension in jeweils einem Kriterium ausgewiesen werden.

- Sowohl auf der Ebene der Forschungseinheiten als auch bei den Forschungseinrichtungen stellt sich das Problem stark differierender Bezugsgrößen (Zahl Wissenschaftler). Die hohe Größenvarianz macht sich bei den quantitativ erfassbaren Indikatoren dadurch bemerkbar, dass Leistungs- und Größeneffekte konfundiert werden. Die Bewertungsgruppe Soziologie plädiert dafür, dass in den verschiedenen Kriterienbereichen auf der Ebene der vergleichsweise personennahen Forschungseinheiten generell **relative**, also im Hinblick **auf Personenzahl gewichtete Größen** in die Bewertung eingehen. Auf der Ebene der Forschungseinrichtungen dagegen sollten in allen Kriterienbereichen, wie schon jetzt bei Effektivität vs. Effizienz, sowohl absolute als auch relative Bewertungen vorgenommen werden.
- Es erscheint der Bewertungsgruppe wünschenswert, dass bei fortgesetztem Einsatz von Forschungsratings sowohl die Datennachfrage als auch die Indikatorenbildung noch stärker standardisiert wird. Besonders dringlich erscheint dieses in bei den Transferkriterien (dazu Anhang 1). **Standardisierungen** würden im „Informed Peer Review“ der Bewertungsgruppe qualitative Einschätzungen und Abwägungen durch die Gutachter nicht obsolet machen, sondern diese zusätzlich informieren und die Zuverlässigkeit ihrer Entscheidungen erhöhen. Von einigen Mitgliedern der Bewertungsgruppe wurden präzise Gewichtungen der einzelnen quantitativen Indikatoren, die eine rechnerische Ermittlung von Bewertungen erlauben würden, für nützlich und statthaft gehalten. Dieser Verfahrensschritt wird von der Bewertungsgruppe insgesamt aber nicht befürwortet, da eine Mehrheit ihrer Mitglieder die Möglichkeit valider Gewichtungen (z.B. bei Monographien im Unterschied zu Aufsätzen, bei Drittmitteln der DFG im Vergleich mit solchen des BMBF) nicht für gegeben hält.

Die damit formulierten Präferenzen und Vorschläge sind aus den Erfahrungen abgeleitet, die bei der Durchführung des Pilotprojekts entstanden sind. Sie bedürfen der Kommentierung und Bewertung durch Betroffene und Experten, und sie sind bei Wiederholungen des Forschungsratings selber wieder zu prüfen. Alle Erfahrungen sprechen dafür, dass die Institutionalisierung von Ratings oder Rankings, die Qualitätsmaßstäbe explizit machen, neben den intendierten auch unerwünschte Folgen hat, die einer kritischen Beobachtung bedürfen. Die Bewertungsgruppe Soziologie plädiert für eine verantwortliche Rolle des Wissenschaftsrats bei künftigen Forschungsratings auch deshalb, weil sie davon ausgeht, dass sich der Wissenschaftsrat in einem besonderen Maße verpflichtet weiß, das Problem unerwünschter Folgen von Evaluationen aufmerksam zu kontrollieren.

## Anhang



## **Anhang 1 Empfehlungen zu den einzelnen Kriterien und zur Datengrundlage der Pilotstudie Forschungsrating Soziologie**

Im Rahmen der Pilotstudie ist bewusst eine relativ umfangreiche Datenbasis erhoben worden, um sicherzustellen, dass auf jeden Fall ausreichend Informationen für eine zuverlässige und differenzierte Bewertung vorliegen (vgl. Anhang 2: Qualitätsmodell). Die breite Anlage der Datennachfrage erfolgte aber auch mit dem Ziel, aufgrund der Erfahrungen aus der Pilotstudie Folgerungen über die Bewährung bzw. Nichtbewährung erhobener Daten zu ziehen und darüber hinaus Möglichkeiten weitergehender Datenanalysen zu prüfen. Eine Regressionsanalyse der Bewertungsergebnisse wurde durchgeführt, um die Signifikanz einzelner Daten und Indikatoren für die Notengebung der Bewertungsgruppe zu ermitteln. Einige von ihnen erscheinen nach dieser Analyse zwar insgesamt als nicht bedeutsam, sind aber im Hinblick auf bestimmte Fälle von Forschungseinheiten und Einrichtungen für die Gutachter doch instruktiv und sollten deshalb nicht allein aus statistischen Gründen ausgeschieden werden. Es ist insofern davon auszugehen, dass ein gewisser Überschuss an Daten auch bei nachfolgenden Forschungsratings unvermeidbar bleibt. Anzumerken ist, dass im Rahmen der Pilotstudie bereits im Anschluss an den Pretest des Erhebungsinstruments der Umfang der Datenerhebung deutlich reduziert wurde.

### **I. Forschungsqualität**

Das Kriterium I: Forschungsqualität wurde als das für die Bewertung der Forschungsleistung zentrale Kriterium angesehen. Daher hat die Steuerungsgruppe beschlossen, durch die Einführung der untersten Aggregationsebene „Forschungseinheit“ eine Differenzierung innerhalb einer Einrichtung zu ermöglichen. Zwar war der Aufwand für die Erhebung auf dieser Aggregationsebene teils recht hoch, die erhobenen Daten lieferten aber eine nach Auffassung der Bewertungsgruppe solide Bewertungsgrundlage.

Eine Regressionsanalyse der Bewertungen und der Vorbewertungen durch die Berichterstatter zeigt, dass die Einschätzung der eingereichten Publikationen und die qualitative Begutachtung der Publikationslisten mindestens das gleiche Gewicht wie die Bewertung der quantitativen Indikatoren hatte. Eine Prognose der Bewertung durch ein optimiertes Regressionsmodell nur auf Basis der quantitativen Indikato-

ren<sup>22</sup> würde im Vergleich zu der tatsächlich vorgenommenen Bewertung in 36 % der Fälle zu einer anderen Einschätzung führen.<sup>23</sup> Auf die Bewertung exemplarischer Publikationen sollte demnach nicht verzichtet werden. Hinsichtlich der Datenerhebung sind folgende Empfehlungen zu berücksichtigen:

- **Eingereichte Publikationen:** Künftig sollte die Vorgabe gemacht werden, dass nur Kapitel aus Monographien, Beiträge zu Sammelbänden und Zeitschriftenaufsätze eingereicht werden können. Forschungsberichte und Rezensionen sind als eigenständige Publikationstypen nicht sinnvoll zu bewerten. Bei größeren Forschungseinheiten sollte vermieden werden, dass alle eingereichten Publikationen nur von einem einzigen Autor stammen.
- **Publikationslisten:** Für die Urteilsbildung der Gutachter ist der Zugriff auf die standardisierten Publikationslisten, die gemeinsam mit den bewerteten Einrichtungen bereinigt wurden, auch künftig unverzichtbar. Ihre Vollständigkeit und die korrekte Zuordnung der einzelnen Publikationstypen sind deutlich zu verbessern.
- **Artikel in Fachzeitschriften mit Peer Review:** Diesem Indikator wurde von vornherein eine besondere Bedeutung zugemessen, weil die entsprechenden Publikationen zuverlässiger als andere im Wissenschaftsprozess qualitätskontrolliert sind. Diese Bedeutung ist im Bewertungsprozess allerdings im Hinblick darauf zu relativieren, dass die Rolle von Fachzeitschriften mit Peer Review im Publikationsspektrum der Soziologie begrenzt ist und die von diesem Indikator generierten Fallzahlen entsprechend relativ niedrig ausfallen. Der Indikator kann also nur im Zusammenhang mit anderen Indikatoren seine besondere Rolle spielen. Dieser Umstand entwertet nicht das Prinzip, dass für Evaluationen, wo immer möglich, der Anschluss an die im Fach schon institutionalisierten Peer-Review-Prozesse gesucht wird. Im Falle der Monographien und Sammelbände ist dies zumindest in der Soziologie nicht umfassend möglich, da die deutschen Verlage ihre Veröffentlichungen nur sehr begrenzt mit fachlichen Gutachten absichern. Im vorliegenden Fall wurden Daten durch den Rückgriff auf die Datenbanken von Thompson Scientific und CSA Sociological

---

<sup>22</sup>  $r^2=0,669$ , Indikatoren: Artikel in referierten Zeitschriften, Monographien, Drittmittelprojekte.

<sup>23</sup> Anteil der mit dem Regressionsmodell prognostizierten Werte, die mehr als eine halbe Notenstufe vom Gutachterurteil abweichen und somit durch Rundung eine andere Bewertung ergeben würden. In zwei Fällen weicht der prognostizierte Wert um mehr als eine Notenstufe vom Gutachterurteil ab. Die Voraussetzungen für eine lineare Regressionsanalyse hinsichtlich des Skalenniveaus sind hier verletzt; sie simuliert jedoch das Vorgehen bei herkömmlichen Rankings, die mit linearen Gewichtungsfaktoren für die einzelnen Indikatoren operieren und eine Alternative zum Forschungsrating darstellen könnten.



Abstracts Journals (s. A.III.3, S.19 f.) gewonnen. Diese Operationalisierung ist nicht völlig trennscharf; einerseits finden sich in den Indizes Zeitschriften, deren System der Qualitätskontrolle nicht höchsten Standards genügt, andererseits sind eine Reihe bekannter, referierter Zeitschriften nicht in den Indizes enthalten. Dieses Problem könnte nur durch eine eigens definierte Positivliste von referierten Zeitschriften gelöst werden. Akzeptanz würde eine solche Positivliste aber nur finden können, wenn sie in einem umfangreichen Konsultationsprozess mit der Fachgemeinschaft erarbeitet würde. Erst dann wäre es auch möglich, die unterschiedliche Bedeutung diverser Fachzeitschriften im Hinblick auf Forschungsqualität differenzierter zu gewichten.

- **Wissenschaftsgesteuert zugewiesene Drittmittelprojekte:** Um Verzerrungen zugunsten mittelintensiver Forschungszweige bei der Bestimmung von Forschungsqualität zu vermeiden, verzichtete die Bewertungsgruppe auf die Analyse von *Drittmittelsummen* auf der Ebene der Forschungseinheiten und beschränkte sich auf die Anzahl von *Drittmittelprojekten* (einschließlich Fortsetzungsbewilligungen), die von DFG, BMBF, EU oder Stiftungen nach Peer Review bewilligt wurden.
- **Liste der Drittmittelprojekte:** Obwohl nur solche Drittmittelprojekte angegeben werden sollten, die im Erhebungszeitraum bewilligt wurden, gaben die betroffenen Einrichtungen zahlreiche Projekte an, die im Erhebungszeitraum noch andauerten, aber früher bewilligt, oder Projekte, die sogar erst nach dem Erhebungszeitraum bewilligt worden waren. Die Häufigkeit des erstgenannten Fehlers zeigt, dass den Fachkoordinatoren offenbar nicht plausibel zu machen war, dass an den Drittmittelprojekten in erster Linie das Faktum der erfolgreichen Begutachtung interessierte. Dies machte aufwändige Prüfungen der Angaben erforderlich. Die bereinigten Listen der Drittmittelprojekte wurden ausgezählt, um den Indikator Anzahl wissenschaftsgesteuert zugewiesener Drittmittelprojekte zu bilden.
- **Zitationszahlen:** Eine für die potentiell besten Forschungseinheiten probeweise durchgeführte Zitationsanalyse, die auch der Überprüfung der internationalen Wirksamkeit ihrer Forschungsleistung diene, bestätigte die Skepsis der Mehrheit der Bewertungsgruppe gegenüber der Anwendung dieses Instruments im laufenden Verfahren. Ein erhebliches, nicht durch Normierungsfaktoren adäquat zu korrigierendes Bias zugunsten bestimmter Teilgebiete begrenzt

das Anwendungsgebiet der Zitationsanalyse in einem Forschungsrating der Soziologie stark. Da eine Verbesserung der bibliometrischen Verfahren auch für ein Fach wie die Soziologie aber nicht grundsätzlich auszuschließen ist, sollte der Einsatz einer Zitationsanalyse bei einer Wiederholung des Ratings erneut geprüft werden. Sinnvoll erscheint auch die Erprobung neuer Indikatoren, um die Rezeption weiterer Publikationssorten neben den referierten Zeitschriften zu messen, etwa durch eine „cited reference“-Analyse, die auch Zitationen auf Monographien und Sammelwerke erfasst, oder durch eine Analyse der Rezeption von Monographien anhand von Rezensionenbanken.

- **Selbstdarstellung der eigenen Stärken und Schwächen:** Die verbalen Ausführungen zu diesem Punkt besaßen nicht durchweg den Informationsgehalt, den sich die Bewertungsgruppe erhofft hatte, da viele der befragten Forschungseinheiten es wohl nicht angebracht fanden, selbstkritische Reflexionen, wenn überhaupt gepflegt, offenzulegen. Die Texte lesen sich zu großen Teilen so, als ginge es bei der Beantwortung von Fragen zum Forschungsrating um Öffentlichkeitsarbeit. Da auch dies im Einzelnen nicht uninformativ ist und da sich die defensive Attitüde auch zur Frage nach den eigenen Schwächen bei Wiederholungen verändern kann, sollte sie nicht aufgegeben werden.

## II. Impact/Effektivität

Eine Regressionsanalyse der Bewertungen nach diesem Kriterium zeigt, dass nur die Publikationsindikatoren Gesamtpublikationszahl und Anzahl Artikel in referierten Zeitschriften signifikant für das Ergebnis sind. Weder das Drittmittelvolumen noch die Anzahl drittmittelfinanzierter Mitarbeiter noch die Anzahlen von nicht-deutschen Publikationen oder von Publikationen in Fachzeitschriften außerhalb der Soziologie i. e. S. sind demgegenüber für die Bewertungsergebnisse statistisch ausschlaggebend. Auf ihre Erfassung sollte aber allein aus diesem Grund nicht vollständig verzichtet werden, da ihre Nennung in den zuletzt genannten Fällen nicht nur dem Informationsbedarf der Bewertungsgruppe dient, sondern auch als Relevanzhinweis für die befragten Wissenschaftler nützlich sein kann. Im Hinblick auf den Drittmittelfaktor demonstrieren die Befunde der Bewertungsgruppe allerdings, dass er in vielen der gegenwärtig praktizierten Evaluationen als Qualitätsausweis überschätzt wird; er ist, wenn pauschal und ohne Berücksichtigung der Finanzierungsquellen behandelt, eher ein Aktivitäts- als ein Qualitätsindikator.

- **Gesamtvolumen wissenschaftsgesteuert zugewiesener Drittmittel:** Da auf Einrichtungsebene verausgabte Drittmittel reliabler zu erfassen waren als eingenommene, bezogen sich die Drittmittelsummen nicht ausschließlich auf die im Erhebungszeitraum bewilligten Drittmittelprojekte, die von den einzelnen Forschungseinheiten aufgelistet wurden, sondern schlossen sowohl verausgabte Mittel von vor dem Erhebungszeitraum bewilligten Projekten ein als auch von solchen, deren Antragsteller die Einrichtung noch vor dem Stichtag verlassen hatten. Ein konsequentes Umsetzen des „Current Potential“-Prinzips, nachdem hier nur verausgabte Drittmittel von zum Stichtag beschäftigten Wissenschaftlern hätten angegeben werden dürfen, war nicht möglich, da die Einrichtungen dazu die Drittmittel auf Rohdatenebene hätten reanalysieren müssen.
- **Anzahl drittmittelfinanzierter wissenschaftlicher Mitarbeiter:** Die hierzu von den Einrichtungen gemachten Angaben wichen häufig von den aggregierten Angaben der Forschungseinheiten ab, sodass dieser Indikator von den Gutachtern als wenig reliabel eingeschätzt wurde. Lässt sich die Datenqualität verbessern, so erscheint es aus inhaltlichen Gründen dennoch wünschenswert, diesen Indikator auch künftig zu verwenden, da er weniger anfällig gegenüber hohen Sachmittelaufwendungen bestimmter Forschungsrichtungen ist als das bloße Drittmittelvolumen.
- **Ausstrahlung der Forschungsaktivität:** Beide zugeordneten Indikatoren, die Zahl der nicht-deutschsprachigen Publikationen und die Zahl der Veröffentlichungen in Fachzeitschriften außerhalb der Soziologie, konnten aus den Literaturdatenbanken abgeleitet werden. Angesichts der unscharfen Grenzen der Soziologie und der damit verbundenen definitorischen Uneindeutigkeit, aber auch wegen bestehender Zweifel an der Reliabilität der Befunde wurde auf den letztgenannten Interdisziplinaritätsindikator im Bewertungsprozess kaum Bezug genommen. Auf weitere Indikatoren, wie zu Gastaufenthalten einkommende und abgehende Wissenschaftler, internationale Kooperationen und veranstaltete Tagungen, wurde nach Erprobung im Pretest verzichtet, da sich nicht nur deren Erhebung als sehr aufwändig erwies, sondern auch weil sich bei den Daten zeigte, dass es sich hierbei eher um einen Aktivitäts- als einen Qualitätsausweis handelt. Es lässt sich in der Wissenschaft nicht unterstellen, dass bloße Geschäftigkeit leistungsfördernd ist.

- **Reputation:** Neben Forschungspreisen und Auszeichnungen wurden exemplarische Angaben zu ausgewählten, als besonders bedeutsam eingeschätzten Ämtern in Gremien und Institutionen außerhalb der eigenen Einrichtung erbeten. Diese Aufforderung wurde sehr unterschiedlich gehandhabt. Manche Wissenschaftler schienen dieses Datum weniger als einen beispielhaften Ausweis von Reputation denn als eine Dokumentation ihrer hohen Arbeitsbelastung zu sehen, die es deshalb durch möglichst umfangreiche Listen eindrucksvoll zu machen gelte. Häufig wurde die Mitgliedschaft in Gremien der eigenen Universität angeführt, wozu Angaben nicht erbeten waren. Angesichts dieser Unschärfen mussten sich die Gutachter auf qualitative Einschätzungen verlassen. Da Reputation aber als ein sehr instruktiver Indikator gelten kann, sollten die auf ihn bezogenen Fragestellungen bei künftigen Erhebungen spezifischer dimensioniert und eindeutiger definiert werden. Dann erscheint es auch möglich, im Hinblick auf eine Reihe von Daten (Herausgeber-, Beirats-, Fachgutachter-tätigkeiten, ebenso bestimmte Arten von Ehrungen) Standardisierungen und Quantifizierungen vorzunehmen. Deren Aggregation setzt allerdings Gewichtungen voraus, über deren Festlegungen sich lange streiten lässt.

### III. Effizienz

Die Effizienz hinsichtlich der Forschungsleistungen wurde durch Relativierung der für die Impactmessung verwendeten Indikatoren auf den Personaleinsatz bestimmt. Wie die Impactbewertung ist auch die Effizienzbewertung nach den Ergebnissen einer Regressionsanalyse primär durch die – relativierten – Publikationsindikatoren determiniert.

- **Personalkategorien:** Für die Relativierung auf den Personaleinsatz ist die Gewichtung der verschiedenen Personalkategorien entscheidend. Eine strittige Frage ist insbesondere, ob und in welchem Umfang drittmittelfinanziertes Personal eingehen soll. Auf der einen Seite stellt es eine Leistung der jeweiligen Einrichtung dar, auf Basis einer begrenzten Grundausstattung durch Drittmiteleinwerbung eine höhere Gesamtkapazität zu erzielen; auf der anderen Seite ist der Einsatz von Drittmitteln nur dort volkswirtschaftlich effizient, wo er zu einem höheren Pro-Kopf-Ertrag führt. Die Bewertungsgruppe hat (grundfinanzierte) Professoren mit dem Faktor 1, grundfinanzierte wissenschaftliche Mitar-

beiter mit dem Faktor 0,5 und drittmittelfinanzierte wissenschaftliche Mitarbeiter mit dem Faktor 0,33 gewichtet. Mit diesem Kompromiss ist die normative Grundsatzfrage zwar nicht beantwortet, allerdings erwiesen sich die berechneten Indikatoren bei entsprechenden statistischen Analysen gegenüber Änderungen der Gewichtungsfaktoren als relativ robust.

- **Lehrbelastung:** Die Lehrbelastung durch Erfragen der geleisteten Semesterwochenstunden empirisch zu bestimmen erwies sich im Pretest als nicht praktikabel. Um dennoch die Unterschiede zwischen Universitäten und außeruniversitären Forschungseinrichtungen zu berücksichtigen, wurden in der Haupterhebung die angegebenen Vollzeitäquivalente an Universitäten mit dem Faktor 0,5 und an außeruniversitären Einrichtungen mit dem Faktor 0,9 multipliziert. Dies erschien der Bewertungsgruppe angemessen, blieb aber eine nicht unumstrittene Schätzung. Eine differenzierte Erhebung der Lehrbelastung wäre gerade angesichts des teils großen Ausmaßes an Lehrexport der Soziologie wünschenswert, doch ist dabei darauf zu achten, dass der Erhebungsaufwand nicht unverhältnismäßig wird.
- **Anteil des drittmittelfinanzierten Personals:** Dieser Indikator wurde wegen häufig fehlerhafter Angaben (vgl. zu Indikatoren unter Kriterium II: Impact/Effektivität) kaum genutzt und ist auch statistisch nicht für das Bewertungsergebnis relevant.

#### IV. Nachwuchsförderung

Die Indikatoren zur Nachwuchsförderung erfassen fast ausschließlich die akademische Nachwuchsförderung. Eine Bewertung der Nachwuchsförderung für den außerakademischen Bereich muss bedauerlicherweise solange Desiderat bleiben, wie aussagekräftige Absolventenstudien an den bewerteten Einrichtungen nicht regelmäßig durchgeführt werden.

- **Promotionszahlen:** Zusätzlich zu den absoluten Zahlen wurden relative Werte (Promotionen pro Professor) errechnet. Dies diene vor allem der Kontrolle der Angaben, da sehr hohe absolute Zahlen bei schlechter Betreuungsrelation ein Indiz für schlechte Qualität sein können. Auch unabhängig davon hat die Bewertungsgruppe Promotionszahlen als Qualitätsindikatoren der Nachwuchsförderung nur zurückhaltend eingeschätzt; ihre Validität ist problematisch, da

Promotionszahlen in erster Linie ein Produktivitäts- und kein Qualitätsausweis sind.

- **Promotionsstipendien und -stellen:** Die Angaben der Einrichtungen zu Promotionsstipendien und -stellen aus Promotionsprogrammen und die Angaben der Forschungseinheiten zu Stipendien aus diesen und anderen Quellen erschienen häufig nicht kongruent. Eine bessere Datenvorhaltung ist hier dringend geboten.
- **Promotionen außeruniversitärer Einrichtungen:** Die Promotionszahlen von Universitäten sind Zahlen abgeschlossener Promotionsverfahren, diejenigen von außeruniversitären Einrichtungen Zahlen betreuter Promotionen. Dadurch kommt es zu Doppelzählungen. Eine Bereinigung um solche Promotionen ist allerdings methodisch schwierig, weil außeruniversitäre Betreuungsverhältnisse nur erfasst werden, wenn die jeweilige Einrichtung am Forschungsrating teilnimmt. Zudem könnte eine Bereinigung einen unerwünschten Anreiz gegen Kooperationen von Universitäten mit außeruniversitären Einrichtungen darstellen.
- **Liste der Promotionen mit Verlagsangaben:** Die Liste der Promotionen erwies sich sinnvoll zur Qualitätseinschätzung der abgeschlossenen Promotionen, zu der auch die Angabe der Publikationsorte (Verlage) zählte - obwohl Verlagsangaben nicht unbedingt als valide Qualitätsausweise gelten können. Häufig überstiegen die angegebenen Promovendenzahlen die Anzahl der von den einzelnen Forschungseinheiten angegebenen Arbeiten. Die Bewertungsgruppe ist davon ausgegangen, dass dies damit zu tun hat, dass die Zahlen der Prüfungsämter implizit eine „Work Done At“-Perspektive widerspiegeln, und hat die Zahlen deshalb im Sinne der „Current Potential“-Perspektive entsprechend korrigiert.
- **Strukturierte Promotionsprogramme:** Ihrer Existenz rechnete die Bewertungsgruppe besondere Bedeutung zu, da eine Institutionalisierung der Doktorandenförderung sowohl die Rolle der Betreuer als auch der Doktoranden zuverlässiger qualifiziert und damit kontrollierbarer macht, und überdies die Existenz eines Mindestmaßes an Ausbildungsprogrammen sichert. Sie haben für ein Forschungsrating darüber hinaus den Vorteil, dass ihre Etablierung einer mehr oder weniger formalisierten Kontrolle durch Peer Review unterliegt. Wo die Soziologie nur über wenige Stellen verfügt, ist es nicht möglich, rein sozio-

logische Graduiertenkollegs einzuwerben; allenfalls kann sich die Soziologie an interdisziplinären Kollegs beteiligen. Auch an größeren Einrichtungen werden Promotionsprogramme nur selten von der Soziologie alleine getragen werden. Für einen fairen Vergleich ist es deshalb notwendig, den Anteil der Soziologie an einem bestimmten Programm zu quantifizieren. Zu diesem Zweck wurde die Anzahl der an Soziologen vergebenen Promotionsstipendien erfragt. Angesichts der Diskrepanzen zwischen diesen Zahlen und den Angaben der einzelnen Forschungseinheiten zu Stipendien erscheinen diese Zahlen jedoch nur bedingt verlässlich. Eine bessere Datenhaltung der Hochschulen ist dringend erforderlich. Wäre diese gegeben, ließe sich auch eine stärkere Standardisierung der im Bereich der Nachwuchsförderung relevanten Daten erreichen.

- **Nachwuchsgruppenleiter:** Die Kategorie des selbständigen Nachwuchsgruppenleiters ist in der Soziologie praktisch unbekannt und kann künftig entfallen.
- **Erstrufe:** Für akademisch orientierte Wissenschaftler ist dies der zentrale Indikator des Erfolgs der Förderung des wissenschaftlichen Nachwuchses, und seine Erfassung gestaltete sich vergleichsweise problemlos. Als Indikator erscheint die Zahl von Erstrufen also vorzüglich geeignet; allerdings hilft dieser Indikator wegen des hohen Konzentrationsgrades und der geringen Fallzahlen nicht, die Leistungen der Einrichtungen im mittleren und unteren Leistungsbe- reich verlässlich zu diskriminieren.

## V. Transfer in andere gesellschaftliche Bereiche

Der Transfer drückt sich in einem Fach wie der Soziologie nicht primär in ökonomischen Produkten wie Patenten und Unternehmensbeteiligungen aus. Vielmehr stehen Beratung und andere Dienstleistungen im Vordergrund, die nur schwer standardisiert zu erfassen sind. Die Datengrundlage besteht deshalb vor allem aus qualitativen Angaben (Listen und Selbstbeschreibungen) der einzelnen Wissenschaftler. Die Bewertungsgruppe sieht in den erhobenen Daten ein wertvolles Material, auf dessen Basis Kategorien für eine künftig stärkere Standardisierung der Datenerhebung entwickelt werden sollten. Zieht man diese Daten mit den Daten zu Kriterium VI: Wissensvermittlung zusammen, wie die Bewertungsgruppe Soziologie vorschlägt, könnte auch im Hinblick auf die Transferleistungen mit einer differenzierteren, nämlich fünfstufigen Notenskala bewertet werden.

- **Drittmittel von Unternehmen, Verbänden und Ministerien mit Ausnahme des BMBF:** Transferdrittmittel stellen den einzigen quantitativen Indikator der Dimension Transfer dar. Die erbetenen Angaben zu „sonstigen EU-Mitteln“ (ohne EU-Rahmenprogramm) konnten dabei nicht einbezogen werden, da diese Differenzierung in den Universitätsverwaltungen offenbar nicht nachvollzogen werden konnte. Da Kommunen und Kreise für die Soziologie, wie sich aus den Angaben der Dienstleistungen, Beratungsämtern und den Selbstbeschreibungen der Transferaktivitäten ergab, offenbar wichtige Transferadressaten darstellen, wäre es wünschenswert, künftig solche Mittel gesondert erfassen zu können.
- **Dienstleistungen, Auftragsforschung und Gutachten:** Die Angaben waren äußerst heterogen und häufig unvollständig. Dennoch handelt es sich um eine unverzichtbare Basis für die Bewertung, die künftig auch die unter „weitere Forschungsprodukte“ erfragten Daten umfassen soll, da sich diese beiden Indikatoren als wenig trennscharf erwiesen.
- **Spinoffs und Unternehmensbeteiligungen:** Auf die gesonderte Erhebung dieses Indikators kann für die Soziologie künftig verzichtet werden, da kaum relevante Angaben gemacht wurden. Wo sie in Einzelfällen bestehen, lassen sie sich in den Selbstbeschreibungen der Einrichtungen benennen.

## **VI. Wissensvermittlung und -verbreitung**

Für dieses Kriterium konnte kein quantitativer Indikator gebildet werden. Die qualitativen Angaben waren sehr heterogen und unterschiedlich stark detailliert. Eine Zusammenlegung der beiden Kriterien in der Dimension Transfer erscheint in der Soziologie daher künftig sinnvoll. Da entsprechende Daten von den Einrichtungen noch seltener vorgehalten werden als von den jeweiligen Akteuren selbst, wurden sie überwiegend auf Ebene der Forschungseinheiten erhoben. Dieser Umstand erlaubt die Feststellung, dass die Aktivitäten hier innerhalb der Einrichtungen ebenso stark divergieren wie die Forschungsqualität. Dies spricht dafür, sie auch auf dieser Ebene zu bewerten. Problematisch erwies sich die Aufforderung zur exemplarischen Nennung herausgehobener Produkte der Wissensverbreitung, da die Befragten angesichts der Ungewissheit, was die Gutachter positiv werten würden, auf eine vollständige Auflistung nur ungern verzichten wollten. Eine stärkere Standardisierung der Indikatoren sollte versucht werden.



- **Weiterbildungsangebote und weitere Maßnahmen zur Wissensverbreitung:** Häufig unvollständige Angaben zu diesen Indikatoren spiegeln die Tatsache wider, dass Weiterbildungs- und Wissensvermittlungsaktivitäten an den Einrichtungen bislang nur geringe Aufmerksamkeit erlangen.
- **Aufsätze und Beiträge außerhalb wissenschaftlicher Fachzeitschriften:** Die Ausstrahlung soziologischer Forschung in die Öffentlichkeit wird häufig nicht nur durch die unter diesem Kriterium erfassten Medienbeiträge erzielt, sondern beruht ebenso oder sogar stärker auf Publikationen, die im Rahmen der Analyse für die Kriterien der Dimension Forschung ausgewertet wurden. Das Feld der „Public Sociology“, die sich häufig in Monographien („Zeit- / Gesellschaftsdiagnosen“) abspielt, ist schwierig zu operationalisieren und von der wissenschaftlichen Kommunikation im engeren Sinne zu trennen. Eine Bewertung von Wissensvermittlung und -verbreitung muss deshalb auch auf die für die einzelnen Forschungseinheiten erstellten Publikationslisten zurückgreifen. Auch dies spricht aus Sicht der Bewertungsgruppe für eine Erhebung und Bewertung des Transfers auf Ebene der Forschungseinheiten, bevor sie für Bewertungen auf der Ebene von Forschungseinrichtungen aggregiert werden. Auf jeden Fall sollten die Schwierigkeiten der Ermittlung aber nicht den Verzicht auf sie veranlassen, da es nach Auffassung der Bewertungsgruppe zu den Aufgaben der Soziologie zählt, sich mit ihren Forschungsbefunden auch außerhalb des Faches mitzuteilen.



## Anhang 2 Bewertungsmatrix/Qualitätsmodell

Qualitätsmodell für die Bewertung von Universitäten und außeruniversitären Forschungseinrichtungen durch Gutachter im Forschungsrating („Informed Peer Review“): Matrix der Kriterien, Bewertungsaspekte und Daten für das Fachgebiet Soziologie

Kriterium	Bewertungsaspekte (mit Gewichtung)	Daten (mit Gewichtung)
<b>Dimension: Forschung</b>		
<b>I. Forschungsqualität<sup>24</sup></b> (= Originalität und wissenschaftliche Bedeutung der Forschungsleistungen sowie Eignung der Methoden)  (Bezugseinheit: Forschungseinheit)	1. Qualität des Outputs ( <b>Y</b> ) <sup>25</sup>	<i>qualitativ:</i> <ul style="list-style-type: none"> <li>– eingereichte Publikationen (<b>X</b>)</li> <li>– Publikationsliste IZ</li> <li>– Selbstdarstellung der eigenen Stärken und Schwächen</li> <li>– Ergänzende Angaben zur Forschung</li> </ul>
	2. Qualitätseinschätzungen von anderen Peers	<i>quantitativ:</i> <ul style="list-style-type: none"> <li>– Zahl der Artikel in Fachzeitschriften mit Peer Review (SSCI, SCI, AHCI, CSA) - absolut (<b>X</b>) und relativ</li> <li>– <i>Hintergrundinformation: Zahl Artikel in Fachzeitschriften mit Peer Review (SSCI, SCI, AHCI ohne CSA (absolut und relativ)</i></li> <li>– Zahl wissenschaftsgesteuert zugewiesener Drittmittelprojekte - absolut und relativ</li> </ul> <i>qualitativ:</i> <ul style="list-style-type: none"> <li>– Liste der Drittmittelprojekte</li> </ul> Hintergrundinformation: <ul style="list-style-type: none"> <li>– Zahl aller Publikationstypen absolut und relativ (Monographien / Hg. Sammelwerke / Zeitschriftenaufsätze / Beiträge zu Sammelwerken / Literaturberichte und Rezensionen)</li> </ul>

<sup>24</sup> Das Kriterium "Forschungsqualität" wird für die verschiedenen "Forschungseinheiten" einer Forschungseinrichtung gesondert bestimmt. Für die nachfolgenden Kriterien II-VI sind die "Forschungseinrichtungen" selber die Bewertungseinheit.

<sup>25</sup> Das Symbol (**X**) bezeichnet die für die Kriterien jeweils primär relevanten Daten. Die sonstigen Daten sollten aber bei der gutachterlichen Einschätzung angemessen, d.h. in Abhängigkeit von variablen Kontextbedingungen, berücksichtigt werden. Das Symbol (**Y**) hingegen bezeichnet die primär relevanten Bewertungsaspekte innerhalb eines Kriteriums. Weitere Erläuterungen zu den erhobenen Daten enthält der Leitfaden zu den Datenberichten. Der Leitfaden ist zu finden unter: [http://www.wissenschaftsrat.de/pilot\\_start.htm](http://www.wissenschaftsrat.de/pilot_start.htm).

<p><b>II. Impact/Effektivität</b></p> <p>(= Beitrag zur Entwicklung der Wissenschaft im Fachgebiet und darüber hinaus)</p> <p>(Bezugseinheit: Gesamteinrichtung - ohne Berücksichtigung ihrer Größe)</p>	<p>1. Publikationsaktivität (<b>Y</b>)</p>	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"> <li>- Zahl wissenschaftlicher Publikationen insgesamt (<b>X</b>)</li> <li>- Zahl der Artikel in Fachzeitschriften mit Peer Review (SSCI, SCI, AHCI, CSA) (<b>X</b>)</li> </ul> <p>Hintergrundinformation:</p> <ul style="list-style-type: none"> <li>- Zahl aller Publikationstypen (Monographien / Hg. Sammelwerke / Zeitschriftenaufsätze / Beiträge zu Sammelwerken / Literaturberichte und Rezensionen)</li> </ul>
	<p>2. Einwerbung qualifizierter Drittmittel</p>	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"> <li>- Gesamtvolumen wissenschaftsgesteuert zugewiesener Drittmittel</li> <li>- Anzahl der drittmittelfinanzierten wissenschaftlichen Mitarbeiter/-innen (VZÄ) (<b>X</b>)</li> </ul>
	<p>3. Ausstrahlung der Forschungsaktivität (<b>Y</b>)</p>	<p><i>quantitativ</i></p> <ul style="list-style-type: none"> <li>- Zahl nicht-deutschsprachiger Publikationen (<b>X</b>)</li> <li>- Zahl der Veröffentlichungen in Fachzeitschriften außerhalb der Soziologie (SCI, AHCI, SSCI, CSA minus Soziologie)</li> </ul>
	<p>4. Reputation</p>	<p><i>qualitativ:</i></p> <ul style="list-style-type: none"> <li>- Forschungspreise und Auszeichnungen</li> <li>- Ämter wissenschaftlichen Institutionen und Gremien außerhalb der eigenen Einrichtung</li> </ul>

<p><b>III. Forschungsleistung (Effizienz)</b></p> <p>(= Impact in Relation zum Aufwand)</p> <p>(Bezugseinheit: Gesamteinrichtung unter Berücksichtigung ihrer Größe)</p>	<p>Erzielter Impact (Krit. II) im Verhältnis zum Personaleinsatz</p>	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"><li>- Absolute Zahl an Publikationen / VZÄ gewichtet nach Statusgruppen und Lehrbelastung <b>(X)</b></li><li>- Absolute Zahl Artikel in Fachzeitschriften mit Peer Review (SSCI, SCI, AHCI, CSA) / VZÄ gewichtet nach Statusgr. u. Lehrbelastung <b>(X)</b></li><li>- Gesamtvolumen wissenschaftsgesteuert zugewiesener Drittmittel / VZÄ gewichtet nach Statusgruppen und Lehrbelastung</li><li>- Verhältnis des drittmittelfinanzierten Personals zum grundfinanzierten Personal (VZÄ) <b>(X)</b></li><li>- Zahl nicht-deutschsprachiger Publikationen/ VZÄ gewichtet nach Statusgruppen und Lehrbelastung <b>(X)</b></li><li>- Zahl der Veröffentlichungen in Fachzeitschriften außerhalb der Soziologie / VZÄ gewichtet nach Statusgruppen und Lehrbelastung</li></ul> <p>Hintergrundinformation:</p> <ul style="list-style-type: none"><li>- Für Spiegelstriche 1-3 und 4-5 jeweils VZÄ-Gewichtung unter Ausschluss drittmittelfinanzierten Personals</li></ul>
--	--	---

<b>Dimension: Nachwuchsförderung</b>		
<p><b>IV. Nachwuchsförderung</b></p> <p>(= Maßnahmen und Erfolge der Förderung des graduierten wissenschaftlichen Nachwuchses im Fachgebiet)</p> <p>(Bezugseinheit: Gesamteinrichtung - ohne Berücksichtigung ihrer Größe)</p>	1. Doktoranden	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"> <li>- Zahl der Stipendien in Promotionsprogrammen und Doktorandenstellen <b>(X)</b></li> <li>- Anzahl der Promotionen</li> <li>- <i>Hintergrundinformation: Zahl von Stipendien und von Promotionen pro Professor.</i></li> </ul> <p><i>qualitativ:</i></p> <ul style="list-style-type: none"> <li>- Liste strukturierter Promotionsprogramme <b>(X)</b></li> <li>- Liste der abgeschlossenen Dissertationen mit Titeln und Verlagsangaben</li> </ul>
	2. Nachwuchswissenschaftler	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"> <li>- Zahl der Postdocstipendien und Nachwuchsgruppenleiter</li> <li>- Rufe an Nachwuchswissenschaftler <b>(X)</b></li> </ul> <p><i>qualitativ:</i></p> <ul style="list-style-type: none"> <li>- Selbstbeschreibung von "Maßnahmen und Erfolgen der Nachwuchsförderung" <b>(X)</b></li> <li>- Liste der Rufe an Nachwuchswissenschaftler</li> </ul>

<b>Dimension: Wissenstransfer</b>		
<p><b>V. Transfer gesellschaftliche Anwendungsbereiche</b></p> <p>(= Beiträge zur Umsetzung von Forschungsergebnissen in Wirtschaft, Politik, Verwaltung, Verbänden etc. durch Anwendung und Beratung)</p> <p>(Bezugseinheit: Gesamteinrichtung - ohne Berücksichtigung ihrer Größe)</p>	<p>1. Umsetzung von Forschungsleistungen durch Dienstleistungen, Unternehmerische Aktivität und Beratungen</p>	<p><i>quantitativ:</i></p> <ul style="list-style-type: none"> <li>- Gesamtvolumen der Drittmittel von Ministerien, Unternehmen und Verbänden <b>(X)</b></li> <li>- <i>Hintergrundinformation: Datum auch auch pro VZÄ.</i></li> </ul> <p><i>qualitativ:</i></p> <ul style="list-style-type: none"> <li>- Liste von Ämtern außerhalb der Wissenschaft <b>(X)</b></li> <li>- Dienstleistungen, Auftragsforschung und Gutachten</li> <li>- Liste weiterer Forschungsprodukte (ohne Publikationen) wie Datenbanken, Software etc.</li> <li>- Spinoffs und Unternehmensbeteiligungen</li> <li>- Selbstbeschreibung zum Transfer in andere gesellschaftliche Anwendungsbereiche</li> </ul>
<p><b>VI. Wissensvermittlung und -verbreitung</b></p> <p>(= Orientierungs- und Bildungsangebote an Nichtfachleute und das allgemeine Publikum)</p> <p>(Bezugseinheit: Gesamteinrichtung - ohne Berücksichtigung ihrer Größe)</p>	<p>1. Fort- und Weiterbildungsangebot und Wissensverbreitung</p>	<p><i>qualitativ:</i></p> <ul style="list-style-type: none"> <li>- Liste eigener Weiterbildungsangebote</li> <li>- Beschreibung sonstiger Wissensverbreitungsmaßnahmen</li> <li>- Aufsätze außerhalb wissenschaftlicher Fachzeitschriften (Zeitungen, Zeitschriften, sonstige Medienbeiträge) <b>(X)</b></li> </ul>





### **Anhang 3 Mitglieder der Bewertungsgruppe Soziologie**

#### **Sachverständige:**

Herr Professor Dr. Friedhelm Neidhardt (Vorsitz), ehemals Wissenschaftszentrum Berlin

Herr Professor Dr. Andreas Diekmann, ETH Zürich

Frau Professor Dr. Birgit Geissler, Universität Bielefeld

Herr Professor Dr. Jürgen Gerhards, Freie Universität Berlin

Herr Professor Dr. Ronald Hitzler, Technische Universität Dortmund

Herr Professor Dr. Hans Joas, Max-Weber-Kolleg, Universität Erfurt

Herr Professor Dr. Karl-Dieter Keim, ehemals Leibniz-Institut für Regionalentwicklung und Strukturplanung

Herr Professor Dr. Siegfried Lamnek, Katholische Universität Eichstätt-Ingolstadt

Herr Professor Dr. Richard Münch, Universität Bamberg

Herr Professor Dr. Werner Rammert, Technische Universität Berlin

Herr Bernhard von Rosenblatt, TNS Infratest, München

Herr Professor Dr. Uwe Schimank, Fernuniversität Hagen

Frau Professor Dr. Heike Solga, WZB Berlin

Herr Professor Dr. Peter Weingart, Universität Bielefeld

Herr Professor Dr. Helmut Wiesenhal, ehemals Humboldt-Universität zu Berlin

Herr Professor Dr. Reinhard Wippler, ehemals Universität Utrecht

**Vertreter der Steuerungsgruppe:**

Herr Professor Dr. Peter Lichter

Deutsches Krebsforschungszentrum, Heidelberg