

Empfehlungen zu Rankings im Wissenschaftssystem
Teil 1: Forschung

<u>Inhalt</u>	<u>Seite</u>
Vorbemerkung	ii
Kurzfassung	iii
A. Ausgangslage	1
I. Typen und Funktionen vergleichender Bewertungsverfahren	1
II. Internationale Beispiele	4
III. Rankings im deutschen Wissenschaftssystem	17
IV. Vergleich bestehender Verfahren vergleichender Bewertung	29
B. Empfehlungen	33
I. Vorbemerkungen	33
II. Empfehlungen zu vergleichenden Bewertungsverfahren im Wissenschaftssystem	35
III. Empfehlungen zu einem Forschungsrating	42
1. Ziele, Adressaten, Objekte	43
2. Allgemeine Leistungsdimensionen und Rahmenkriterien	45
3. Forschungsprofile der Einrichtungen	48
4. Quantitative Indikatoren	49
5. Fachgebietsspezifische Bewertung durch Bewertungsgruppen	50
6. Ergebnisdarstellung	52
7. Trägerschaft, Organisation und Finanzierung	52
8. Pilotstudie	55
9. Internationales Benchmarking	56
Anhang	58
Literaturverzeichnis	58
Schema der Forschungsprofile	62

Vorbemerkung

Bund und Länder haben den Wissenschaftsrat im Mai 2003 gebeten, ein Konzept für ein Ranking zu entwickeln. Der Wissenschaftsrat hat daraufhin im Juli 2003 eine Arbeitsgruppe eingerichtet, die sich mit Zielen und Methodiken von Rankings befasst und dazu nationale und internationale Experten angehört hat. Die Arbeitsgruppe ist zu dem Schluss gekommen, dass ein Vorschlag zur Methodik einer vergleichenden Leistungsbewertung eine klare Zielsetzung haben muss und nur dann auf Akzeptanz stoßen wird, wenn er methodologisch gut begründet ist. Sie hat deshalb zunächst Empfehlungen zu Rankings im Wissenschaftssystem entwickelt. Aus wissenschaftspolitischer Sicht werden sowohl für den Bereich der Lehre als auch für den Bereich der Forschung Verfahren benötigt, die diese Kriterien erfüllen. Für den Bereich der Lehre sind dafür noch Vorarbeiten zu einem für Vergleichszwecke verwendbaren Qualitätsbegriff in der Lehre notwendig. Zudem ist der Bereich der Hochschullehre derzeit wegen des Bologna-Prozesses großen Veränderungen unterworfen. In einem ersten Schritt der Konkretisierung hat die Arbeitsgruppe deshalb Empfehlungen zu einem Forschungsrating vorbereitet, das sich in erster Linie an Entscheidungsträger in den Universitäten und außeruniversitären Forschungseinrichtungen sowie an ihre Partner in den Ministerien richtet. Daran soll sich in einem zweiten Schritt ein Vorschlag für ein Verfahren der vergleichenden Bewertung der Lehre anschließen.

Der Arbeitsgruppe haben auch Sachverständige angehört, die nicht Mitglieder des Wissenschaftsrates sind. Ihnen ist der Wissenschaftsrat zu besonderem Dank verpflichtet.

Der Wissenschaftsrat hat die vorliegenden Empfehlungen am 12. November 2004 verabschiedet.

Kurzfassung

1. Vergleichende Leistungsbewertungen durch Rankings und verwandte Verfahren können die Leistungstransparenz im Wissenschaftssystem erhöhen, wissenschaftliche Einrichtungen in ihren strategischen Entscheidungen unterstützen und wesentlich zu einem effektiven und effizienten Wettbewerb beitragen.

Rankings dokumentieren erbrachte Leistungen und aktuelle Leistungspotentiale. In Verbindung mit anderen Instrumenten der strategischen Planung und Qualitätssicherung können sie die Akteure im Wissenschaftssystem bei rationalen strategischen Entscheidungen unterstützen. Eine vollständig oder überwiegend auf Rankingergebnissen basierende Steuerung des Wissenschaftssystems ist dagegen weder zu empfehlen, noch ist das vom Wissenschaftsrat empfohlene Verfahren darauf angelegt.

Rankings im engeren Sinne, die zu ordinalen Ranglisten führen, sind nur unter spezifischen Voraussetzungen sinnvoll. In der Regel genügt eine Differenzierung durch Ranggruppenbildung oder entlang einer vorgegebenen Notenskala (Rating). Unterschiedlichen Nutzerinteressen, die sich in abweichenden Gewichtungen verschiedener Kriterien äußern, wird durch eine multidimensionale Bewertung, die zu differenzierten Profilen der bewerteten Einrichtungen führt, Rechnung getragen.

Anerkannter Standard für vergleichende Bewertungsverfahren ist eine Verschränkung von Peer-Review und quantitativen Indikatoren, wobei die Gewichtung von Fall zu Fall unterschiedlich sein kann.

2. Wegen der Bedeutung der Forschung für Erfolg und internationale Sichtbarkeit der wissenschaftlichen Einrichtungen empfiehlt der Wissenschaftsrat ein Forschungsrating von Universitäten und außeruniversitären Forschungseinrichtungen, das diese bei strategischen Entscheidungen unterstützen soll.

Das Forschungsrating soll fachspezifisch erfolgen. Dafür ist eine Taxonomie notwendig, die in Anlehnung an heute schon national oder international gebräuchliche Taxonomien gewählt werden sollte.

Die Bewertung soll multidimensional nach neun Rahmenkriterien in den drei Dimensionen Forschung, Nachwuchsförderung und Wissenstransfer erfolgen. Es ist weder

eine Gewichtung der Fachgebiete und der Kriterien noch eine Verrechnung zu einer Gesamtnote vorgesehen.

Die Bewertung soll anhand einer siebenstufigen Skala nach internationalen Maßstäben erfolgen und durch Bewertungsgruppen für jedes Fachgebiet vorgenommen werden. Den Bewertungsgruppen obliegt auch die Operationalisierung der Rahmenkriterien für das jeweilige Fachgebiet.

Grundlage der Bewertung sind fachgebietsspezifische Forschungsprofile, die durch die bewerteten Einrichtungen einzureichen sind, sowie bibliometrische Indikatoren. Die Bewertungsgruppen können für ihr Fachgebiet spezifische Anforderungen an die Bewertungsgrundlagen formulieren.

Ergebnisse des Verfahrens sind fachgebietsspezifische Ratings nach den unterschiedlichen Kriterien, die es erlauben, die Universitäten oder die außeruniversitären Forschungseinrichtungen jeweils untereinander zu vergleichen. Zugleich sollen die Ergebnisse auch zur Bildung von Leistungsprofilen der einzelnen Einrichtungen herangezogen werden können.

Die Verfahrenshoheit soll bei einer Steuerungsgruppe liegen, die aus renommierten Wissenschaftlern besteht und in der die großen Wissenschaftsorganisationen in angemessener Weise vertreten sind. Als Träger ist eine wissenschaftsnahe Organisation vorgesehen, die über organisatorische und methodische Kompetenz in der Forschungsbewertung verfügt und von den bewerteten Einrichtungen unabhängig ist.

Um die Methodik des Forschungsratings zu erproben und zu konkretisieren, empfiehlt der Wissenschaftsrat eine Pilotstudie in zwei Fachgebieten. Nach erfolgreichem Abschluss der Pilotstudie könnte das Forschungsrating in einem rollierenden, auf fünf bis sechs Jahre angelegten Turnus alle Fachgebiete abdecken.

Nach Abschluss der Pilotstudie sollte außerdem gemeinsam mit anderen Ländern, die vergleichbare Verfahren bereits etabliert haben, geprüft werden, ob eine gemeinsame oder gekoppelte Bewertung von Forschungsleistungen im Sinne eines Internationalen Benchmarkings realisiert werden kann. Ein solches Benchmarking kann den beteiligten Ländern helfen, den Rang ihrer Universitäten und außeruniversitären Forschungseinrichtungen noch überzeugender zu bestimmen, ihre Maßstäbe wechselseitig zu kontrollieren und in methodischer Hinsicht voneinander zu lernen.

A. Ausgangslage

A.I. Typen und Funktionen vergleichender Bewertungsverfahren

In den letzten Jahrzehnten ist im deutschen Wissenschaftssystem ein umfangreiches Instrumentarium der Leistungsbewertung und Qualitätssicherung in Gebrauch gekommen. Dazu zählen in erster Linie die verschiedenen Arten von Evaluationen:

- Die Max-Planck-Gesellschaft verfügt über ein mehrstufiges System der Qualitätssicherung, in dessen Mittelpunkt die regelmäßige, externe Bewertung der Institute durch ihre Fachbeiräte steht. Hinzu kommt u. a. in größeren Abständen die so genannte Erweiterte Evaluation mehrerer thematisch ähnlicher Institute.¹⁾
- Nachdem der Wissenschaftsrat schon seit den 1980er Jahren die Institute der Blauen Liste systematisch und regelmäßig evaluiert hat, um die Voraussetzungen ihrer gemeinsamen Förderung durch Bund und Länder zu überprüfen, ist die Verantwortung für die in der Förderung befindlichen Einrichtungen 2003 in die Hände des Senats der Leibniz-Gemeinschaft übergegangen. Jedes Institut der Leibniz-Gemeinschaft soll künftig mindestens alle sieben Jahre durch eine vom Senatsausschuss Evaluierung eingesetzte Arbeitsgruppe evaluiert werden.²⁾
- Die Helmholtz-Gemeinschaft hat mit der Umstellung auf eine programmorientierte Förderung im Jahr 2001 begonnen, ihre Forschungsaktivitäten in Forschungsprogrammen zu reorganisieren, die alle fünf Jahre einer strategischen Evaluation unterzogen werden. Daneben bestehen die institutionellen Evaluationsverfahren der einzelnen Forschungszentren fort.

¹⁾ MPG 2002.

²⁾ Vgl. www.wgl.de/evaluation.

- In den Jahren 1998 bis 2001 wurden die Deutsche Forschungsgemeinschaft und die Trägerorganisationen der außeruniversitären Forschungseinrichtungen mit Blick auf ihre Organisation und Leistungsfähigkeit Systemevaluations unterzogen.³⁾

Auch zahlreiche Universitäten verfügen, zum Teil aufgrund entsprechender Vereinbarungen mit dem jeweiligen Sitzland, heute über etablierte Evaluationsverfahren. In einigen Bundesländern werden Evaluationen im Universitätsbereich zentral organisiert.⁴⁾

Weniger etabliert sind hingegen in Deutschland explizit vergleichende Bewertungen. Sie können die auf einzelne Einrichtungen bezogenen Verfahren ergänzen, indem sie Hinweise auf Stärken und Schwächen geben und helfen, Ergebnisse interner Bewertungen einzuordnen. Bekannt sind Rankings, die in die Aufstellung von Ranglisten münden, bisher vor allem im Bereich der universitären Lehre. Die größte Aufmerksamkeit erzielen derzeit Rankings von Studiengängen der Hochschulen, die in Publikumszeitschriften veröffentlicht werden und sich an Studienanfänger und ihre Eltern richten.

Rankings unterscheiden sich von Evaluationen durch die Fokussierung auf die Messung und Bewertung des Outputs – sie beinhalten keine Handlungsempfehlungen und sind nicht prozessorientiert – sowie die Vergleichsabsicht, es werden also mehrere Institutionen oder Förderprogramme nach den gleichen Maßstäben analysiert und bewertet. Unter den verschiedenen Verfahren des Leistungsvergleichs zeichnet Rankings aus, dass die Objekte einer bestimmten Menge (z. B. „Universitäten in Deutschland“) möglichst vollständig erfasst, dass die Leistungskriterien durch ein Kennzahlensystem operationalisiert und dass die Ergebnisse der Leistungsmessung durch die Bildung von Ranglisten verdichtet werden.⁵⁾ Die vollständige Erfassung unterscheidet Rankings von Benchmarkings, bei denen entscheidungsstützende In-

³⁾ Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission, 1998; Forschungsförderung in Deutschland. Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft, Hannover 1999; Wissenschaftsrat: Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Helmholtz-Gemeinschaft Deutscher Forschungszentren, Köln 2001; Wissenschaftsrat: Systemevaluation der Blauen Liste – Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste, Köln 2001.

⁴⁾ Exemplarisch zum Verfahren der Wissenschaftlichen Kommission Niedersachsen s. S. 26 f.

⁵⁾ Bayer (1999).

formationen durch den Vergleich mit gezielt ausgewählten, in der Regel besonders leistungsfähigen Referenzeinrichtungen gewonnen werden sollen; Operationalisierung durch Kennzahlen und Verdichtung in Ranglisten wiederum unterscheidet Rankings von Ratings, also der Bewertung von Institutionen nach einer vorgegebenen Notenskala, die in der Regel durch Expertengruppen vorgenommen wird. Auch Ratings können allerdings zur Bildung von Ranggruppen führen oder in Rankings eingehen.

„Informationsverdichtung“ ist eine zentrale Funktion von Rankings: Rankings dienen dazu, Informationen über eine größere Zahl heterogener Institutionen und Bewertungsdimensionen zugänglich zu machen und sie so aufzubereiten, dass sie handlungsleitend werden können. In der Regel geht es dabei um Institutionen, die miteinander im Wettbewerb stehen, und das Handeln der Adressaten von Rankings ist für diesen Wettbewerb konstitutiv. Rankings sind also ein Instrument, um wettbewerblich organisierte Systeme durch Erhöhung der Transparenz effektiver zu machen.

Transparenz ist nicht nur für die Akteure innerhalb eines Systems, sondern gerade auch für Nutzer und Förderer von großer Bedeutung. Wichtige Nutzer der Hochschulen sind Studienanfänger, die mit Hilfe von Rankings ihren Studienort, also den optimalen Punkt zum Eintritt in den tertiären Bildungssektor und damit zumindest temporär in das Wissenschaftssystem suchen. Sie sind deshalb die typischen Adressaten der meisten bestehenden Rankings im In- und Ausland und zugleich eine „Ressource“ für die Hochschulen. Auch für Studierende oder Wissenschaftler aus dem Ausland, die ebenfalls „Außenstehende“ des deutschen Wissenschaftssystems sind, enthalten Rankings wertvolle Informationen, und es ist denkbar, dass sich die Verfügbarkeit solcher Auskünfte nicht nur auf die Ortswahl innerhalb eines Landes auswirkt, sondern die Interessenten sich überhaupt eher zu einem Gastaufenthalt in einem Land entschließen, wenn sie sich über die relative Qualität der zur Auswahl stehenden Institutionen gut informiert fühlen. Insofern können Rankings als Instrument dazu angesehen werden, die Attraktivität eines Wissenschaftssystems für internationale Interessenten zu erhöhen.

In der Regel wird bei Rankings nur ein Ausschnitt aus den Gesamtleistungen der Objekte erfasst. So ist bei den internationalen Beispielen für Rankings und verwandte

Verfahren im Wissenschaftsbereich zwischen Lehrrankings, die sich in der Regel an Studierende wenden, und Forschungsrankings zu unterscheiden.

A.II. Internationale Beispiele

Das Interesse an Rankings wissenschaftlicher Einrichtungen ist in Deutschland nicht zuletzt dadurch gewachsen, dass Rankings und verwandte Formen des Leistungsvergleichs im angloamerikanischen Sprachraum seit langem etabliert sind und dort eine große Bedeutung haben. Einige besonders bekannte und typische Rankings und verwandte Verfahren werden im Folgenden kurz dargestellt.

„America’s Best Colleges” – U.S. News & World Report

Die Zeitschrift *U.S. News & World Report* veröffentlicht seit 1983 jährlich Rankings US-amerikanischer Universitäten und Colleges, die sie gemeinsam mit der Firma Common Data Set Initiative erstellt. Die Ergebnisse werden als Buch oder als kostenpflichtiges Internetangebot vertrieben. Die Rankings richten sich an Studienanfänger und Studienwechsler und sollen ihnen helfen, die richtige Bildungseinrichtung zu finden, um ihre Karrierechancen und ihre Lebensqualität zu verbessern.⁶⁾ Sie stärken damit den Wettbewerb zwischen den Bildungseinrichtungen um Studierende.

Gegenstand der U.S. News-Rankings sind Universitäten und Colleges, wobei keine fachliche Differenzierung vorgenommen wird, sondern ganze Institutionen gerankt werden. Die Institutionen werden jedoch in Anlehnung an die Carnegie-Klassifikation⁷⁾ in vier Kategorien eingeteilt, für die getrennte Ranglisten aufgestellt werden: (1.) nationale Universitäten mit vollem Angebot an Abschlüssen und hoher Forschungsaktivität („national universities – doctoral“), (2.) nationale Liberal Arts Colleges mit Schwerpunkt auf Geistes- und Sozialwissenschaften (mindestens 40 % der Abschlüsse), deren Studenten vorwiegend den Bachelor anstreben, (3.) regionale Universitäten mit einem breiten Fachspektrum und Abschlüssen vor allem im Bachelor- und zu einem geringeren Anteil im Master-Bereich sowie nur seltenen Promotionen, und schließlich (4.) allgemeine regionale Colleges, die ausschließlich grund-

⁶⁾ Angaben nach dem Internetangebot von U.S. News & World Report unter www.usnews.com.

⁷⁾ Die erstmals 1973 veröffentlichte und seitdem mehrfach revidierte Carnegie-Klassifikation (Carnegie Foundation 2001) soll die höheren Bildungseinrichtungen in den USA in relativ homogene Typen einteilen. Sie wurde ursprünglich vor allem für Zwecke der Bildungsforschung geschaffen, ist heute aber auch im hochschulpolitischen Diskurs in den USA gebräuchlich.

ständige Studiengänge bis zum Bachelor in unterschiedlichen Fächern, aber ohne den für Liberal Arts Colleges typischen Fokus auf Geistes- und Sozialwissenschaften anbieten. Für die beiden regionalen Kategorien werden getrennte Ranglisten für den Norden, den Süden, den Mittleren Westen sowie den Westen der USA aufgestellt, so dass insgesamt zehn separate Ranglisten vorliegen. Außerhalb der Ranglisten bereitet *U.S. News & World Report* Daten für Kunst- und Musikhochschulen und andere Spezialschulen auf; außerdem gibt es fachspezifische Rankings von Studienprogrammen in den Bereichen Ingenieurwissenschaften und Betriebswirtschaft.

Die Rankings von *U.S. News & World Report* fußen auf derzeit 16 quantitativen Indikatoren, die zu sieben Faktoren zusammengefasst, gewichtet und dann zur Bildung der Ranglisten addiert werden.⁸⁾ Es handelt sich um folgende Faktoren (in Klammern die Gewichtung in Prozent):

- Akademische Reputation gemäß einer Umfrage unter Präsidenten, Provosts und Dekanen von Hochschulen (25 %);⁹⁾
- Ausstattung für die Lehre, unter anderem auf Basis des durchschnittlichen Gehalts der Lehrkräfte und der Betreuungsrelationen sowie von Angaben zur Größe von Lehrveranstaltungen (20 %);
- Verbleibquote der Studienanfänger im 3. Semester (20 %);
- Selektivität der Studierendenauswahl, unter anderem auf Basis der Eingangstests und der Annahmehquoten (15 %);
- Finanzielle Ausstattung (10 %);
- Erfolgsquote (5 %); und
- Spenden der Alumni (5 %).

Die Rankings sind also auf die Reputation, die infrastrukturellen Voraussetzungen der Lehre und die Selektivität der Studierendenauswahl fokussiert. Neben Rankings nach „Leistung“, die durch die bereits beschriebene Gewichtung erzielt werden, bietet *U.S. News & World Report* auch Rankings nach dem Preis-Leistungs-Verhältnis an, die wiederum durch gewichtete Addition von Punkten berechnet werden, und

⁸⁾ Eine Normierung von Punktzahlen erfolgt, indem die jeweils beste Institution den Wert 100 zugeordnet bekommt.

⁹⁾ Es handelt sich also um eine Umfrage unter Administratoren. Die von der Zeitschrift verwendete Bezeichnung „peer assessment“ soll deutlich machen, dass die Institutionen von ihren „peer institutions“, d. h. von den Administratoren der Institutionen in der gleichen institutionellen Kategorie, bewertet werden. Dies ist nicht mit der Erhebung von Urteilen von Fachkollegen i. S. von Peer-Review zu verwechseln, die bei einem nicht fachspezifischen Ranking wie dem von *U.S. News* ohnehin der Grundlage entbehren würde.

zwar für das Verhältnis von Leistungspunktwertung und Nettokosten für einen durchschnittlichen Studenten (unter Einbeziehung von Studiengebühren, Unterbringungs- und Lebenshaltungskosten und Stipendien), für den Anteil von Studierenden mit Stipendien sowie für den Anteil von Studierenden, denen eine Gebührenermäßigung eingeräumt wird.

„Research Doctorate Programs in the U.S.“ – National Research Council

In den USA sind seit 1925 wiederholt Wissenschaftler nach der Qualität der Graduiertenstudien an den verschiedenen Universitäten und Colleges befragt worden. Diese Befragungen wurden zunächst von einzelnen Wissenschaftlern und Administratoren initiiert, später von den Fachgesellschaften übernommen. 1982 führte erstmals der National Research Council (von den Nationalen Akademien getragenes Beratungsgremium) auf Initiative des Rats der Fachgesellschaften eine Studie über die Bewertung forschungsbezogener Promotionsprogramme in den USA durch, die 1993 mit veränderter Methodik wiederholt wurde.¹⁰⁾ Gegenwärtig befindet sich eine erneute Studie in Vorbereitung. Eine Vorstudie über die Methodologie der Bewertung forschungsbezogener Promotionsprogramme wurde bereits abgeschlossen.¹¹⁾

Die Studien von 1982 und 1993 hatten drei Ziele und Adressatengruppen im Auge:

- Studenten und ihre Berater bei der Suche nach einem geeigneten Promotionsprogramm zu unterstützen;
- Universitätsadministratoren, Politikern auf nationaler und einzelstaatlicher Ebene sowie Managern in Fördereinrichtungen entscheidungsunterstützende Informationen bereitzustellen sowie
- eine aktuelle Datenbasis für Wissenschaftler zu schaffen, die das Bildungssystem der USA und das mit ihm verbundene Forschungssystem erforschen.

Im Mittelpunkt beider Studien stehen Reputationsdaten über die erfassten Promotionsprogramme. In einem Punkt allerdings unterscheiden sie sich: während die Daten 1982 noch in fachspezifischen Listen alphabetisch nach den Namen der Universitäten und Colleges aufgelistet wurden, entschied man sich 1993 erstmals, auf Basis der Reputationswerte Ranglisten aufzustellen. Zur Begründung für diese Wahl der

¹⁰⁾ Jones et al. (Hrsg. 1982); Goldberger et al. (Hrsg. 1995).

¹¹⁾ Ostriker & Kuh (2003).

Darstellungsform wird angeführt, die alphabetische Darstellung sei nicht nutzerfreundlich („a source of frustration for many users“).¹²⁾

Die Studie von 1993 umfasst insgesamt 41 Fächer in den Natur-, Ingenieur-, Geistes- und Sozialwissenschaften. Die Fächer der „professional schools“, für die der Ph.D. nicht der reguläre Doktorgrad ist, also Jura, Betriebswirtschaft und Medizin, wurden nicht einbezogen, ebenso Universitäten, die in dem jeweiligen Fach weniger als fünf Promotionen in fünf Jahren zu verzeichnen hatten. Insgesamt 3.634 Promotionsprogramme von 274 Universitäten wurden erfasst, die zusammen 90 % der in den abgedeckten Fächern erteilten Promotionen auf sich vereinten.

Für die Reputationsmessung wurden fast 17.000 Wissenschaftler in den USA befragt. Jeder von ihnen wurde gebeten, eine Zufallsauswahl von 50 Promotionsprogrammen in seinem Fach zu bewerten, und zwar nach drei Kriterien:

- wissenschaftliche Qualität des Personals (6er-Skala);
- Effektivität in der Ausbildung von Wissenschaftlern (4er-Skala) und
- Veränderung der Qualität des Programms über die letzten fünf Jahre (besser/unverändert/schlechter).

Zusätzlich wurden die Befragten gebeten, den Grad ihrer Bekanntheit mit dem Programm einzuschätzen. Eine Namensliste der an dem Programm beteiligten Wissenschaftler sowie eine Angabe über die Zahl der Promotionen in den letzten fünf Jahren lagen dem Fragebogen bei. Die Autoren der Studie haben sich zum Ziel gesetzt, für jedes der bewerteten Promotionsprogramme mindestens 100 Bewertungen einzuholen. Aufgrund der langjährigen Erfahrungen mit Reputationsmaßen beinhaltet die Studie eine ausführliche, zur Vorsicht im Umgang mit den Ergebnissen mahnende Diskussion, in der unter anderem auf die Beeinflussung der Reputation durch die bloße Sichtbarkeit (Größe) eines Programms sowie auf die so genannten Halo- und Star-Effekte hingewiesen wird.¹³⁾

¹²⁾ Goldberger et al. (Hrsg. 1995), S. 13.

¹³⁾ Ein Halo-Effekt liegt vor, wenn die gute Reputation einer übergeordneten Einrichtung auf ihre Teilorganisationen abstrahlt. Ein Star-Effekt liegt vor, wenn die gute Reputation eines einzelnen Wissenschaftlers oder einer Gruppe von Wissenschaftlern auf ihre Trägerorganisation abstrahlt.

Die Fokussierung auf Reputationsmaße ist in den USA nicht unkritisiert geblieben. Die Studie von 1993 enthält ergänzend zu den Reputationsmaßen eine Reihe von quantitativen Daten zu jedem Promotionsprogramm:

Kategorie	Inhalt
Personal	Zahl (abs.), Anteil Vollzeitprofessoren, Anteil Wissenschaftler mit Drittmitteln, Anteil Preisträger, Publikationen/Wissenschaftler, Gini Pub, Zitationen/Wissenschaftler, Gini Zit ¹⁴⁾
Studenten	Zahl (abs.), Frauenanteil, Promotionen (abs.)
Absolventen	Frauenanteil, Minderheitenanteil, US-Bürger-Anteil, Anteil der Research Assistants, Anteil der Teaching Assistants, Median der Promotionsdauer

Weitere Erhebungen, unter anderem zum Karriereerfolg von Absolventen, wurden in Erwägung gezogen, aber nicht umgesetzt.

Die Autoren der methodologischen Untersuchung zur Vorbereitung einer erneuten Studie empfehlen unter anderem eine regelmäßige (jährliche) Aktualisierung einer ausgeweiteten Menge statistischer Informationen, einen Verzicht auf eine gesonderte Frage zur „Effektivität“ der Promotionsprogramme bei der Reputationserhebung,¹⁵⁾ größere Anstrengungen, die Bildungsleistung von Promotionsprogrammen zu messen, sowie die Wahl einer anderen Darstellungsform als der einfachen Rangliste, um Fehlinterpretationen aufgrund des fälschlichen Eindrucks übergroßer Genauigkeit (der durch die Angabe von zwei Dezimalstellen bei den Mittelwerten der Reputation noch forciert wurde) zu vermeiden. Um die Trägheit von Reputationsmaßen und die langen Erhebungsintervalle zu kompensieren, analysieren sie zudem die Abhängigkeit der gemessenen Reputation von den leichter zu erhebenden quantitativen Daten und schlagen vor, Änderungen der Reputation in der Zeit zwischen zwei Umfragen jährlich mit Hilfe einer von ihnen entwickelten Gleichung zu prognostizieren.

¹⁴⁾ Der Gini-Koeffizient ist ein Konzentrationsmaß, an dem man ablesen kann, ob absolute Werte eines Indikators auf eine Leistung vieler Mitarbeiter oder nur auf wenige „Stars“ zurückzuführen sind. Bibliometrische Daten werden nur für Natur-, Sozial- und Ingenieurwissenschaften ausgewertet (Quelle: *Institute of Scientific Information*).

¹⁵⁾ Die Antworten korrelieren sehr hoch mit den Antworten auf die Frage nach der wissenschaftlichen Qualität des Personals und basieren nach Auffassung der Kommission nicht auf Kenntnissen der Struktur und Organisation der Promotionsprogramme.

„Good University Guide“ – The Times (UK)

Ähnlich wie in den USA gibt es auch in Großbritannien eine Reihe von Hochschulrankings, die von Publikumszeitschriften angeboten werden und sich an Studienanfänger richten. Unter ihnen ist der jährlich aktualisierte „Times Good University Guide“ am bekanntesten und einflussreichsten. In ihm werden die Studienangebote der britischen Hochschulen in gut 60 Fächern in Ranglisten bewertet. Außerdem wird auch ein globales Ranking der besten 100 Universitäten im Vereinigten Königreich angeboten.

Der *Good University Guide* verwendet für die fachspezifischen Ranglisten nur eine geringe Zahl von Indikatoren:

- Abschneiden im Teaching Quality Assessment (TQA);
- Abschneiden in der Research Assessment Exercise (RAE);
- Durchschnittlicher Schulabschluss der Studienanfänger (jeweils beste drei A-Levels);
- Prozentsatz der Absolventen, die innerhalb von sechs Monaten eine Anstellung finden oder ein weiterführendes Studium aufnehmen.

Die globale Rangliste der 100 besten Universitäten verwendet zusätzliche Indikatoren, darunter das Betreuungsverhältnis, Ausgaben für Bibliotheken und Computer sowie für Sozial- und Freizeitangebote, Anteil sehr guter und guter Abschlüsse sowie die Studienabschlussquoten. Alle Werte werden auf 100 normiert, anschließend die Qualität der Lehre (TQA) mit dem Faktor 2,5, die Qualität der Forschung (RAE) mit dem Faktor 1,5 gewichtet.

Bemerkenswert am *Good University Guide* ist, dass er – wie die Mehrzahl der britischen Konkurrenzprodukte, etwa der *Sunday Times University Guide* oder der *Guardian University Guide* – zu einem entscheidenden Teil auf der Qualitätsbewertung von Forschung und Lehre durch zwei staatliche organisierte Verfahren beruht, durch das *Teaching Quality Assessment* bzw. seine Nachfolgeprozesse und durch die *Research Assessment Exercise*.

„Teaching Quality Assessment“ – UK Quality Assurance Agency

Im Zuge der Reform des Hochschulwesens wurden 1993 für die Finanzierung der Hochschulen in England, Schottland und Wales Hochschulfinanzierungsräte (*Higher*

Education Funding Councils) geschaffen, die zunächst jeweils eigene Verfahren der Qualitätssicherung etablierten. 1997 wurde die *Quality Assurance Agency for Higher Education (QAA)* gegründet, um ein einheitliches System der Qualitätssicherung zu schaffen. Sie wird durch Beiträge der Hochschulen finanziert. Seit dem Jahr 2000 führt die QAA die bis dahin als „Teaching Quality Assessment“ (TQA) bezeichneten Bewertungen der Studiengänge an den Hochschulen als „subject review“ fort, die zusammen mit der „institutional review“ bzw. mit „institutional audits“ die „academic review“ bilden.¹⁶⁾ In der Öffentlichkeit wird nach wie vor der Begriff „Teaching Quality Assessment“ beibehalten, der deshalb auch hier verwendet wird.

Primäres Ziel der TQA ist, die Qualität akademischer Bildungsangebote zu sichern und Anreize zu ihrer Verbesserung zu geben, wobei die Möglichkeit, aufgrund der Ergebnisse öffentliche Mittel umzuverteilen, ebenfalls in Betracht gezogen wird. Ergänzend ist es Absicht der TQA, Informationen über die Qualität tertiärer Bildung öffentlich zugänglich zu machen und damit einer Pflicht zur öffentlichen Rechenschaftslegung nachzukommen. Schließlich ist auch daran gedacht, Studierenden bei der Suche nach einem für sie geeigneten Studienangebot zu helfen.

Im Rahmen der TQA werden individuelle Berichte über die einzelnen Studienangebote der Hochschulen erstellt. Insofern handelt es sich eher um flächendeckende Evaluationen als um Rankings. Die Evaluationen umfassen allerdings eine Bewertung nach einer vorgegebenen Skala im Sinne eines Ratings. Künftig sollen alle Studienangebote in einem sechsjährigen Turnus bewertet werden; die Ergebnisse werden dann in einer Übersicht zusammengefasst, die es erlaubt, die Bewertungen der Angebote aller Hochschulen in dem jeweiligen Fach zu vergleichen.

Grundlage jeder Bewertung ist ein Selbstbewertungsdokument, das die Institution nach einem vorgegebenen Muster anfertigt und das Angaben zur Zielsetzung des betreffenden Studienangebots sowie eine Stellungnahme zur Angemessenheit und Klarheit dieser Zielsetzungen, zur Effektivität und Aktualität der Kurrikula, zur Qualität von Studien- und Lehrbedingungen sowie zur Qualitätssicherung beinhaltet.

¹⁶⁾ Dabei werden nach wie vor in England, Schottland und Wales leicht unterschiedliche Verfahren angewandt. Vgl. *Quality Assurance Agency for Higher Education* (2000).

Auf Basis der Selbstbewertungsdokumente werden die Studienangebote von Fachkomitees bewertet, die größtenteils aus Hochschullehrern bestehen, aber auch Fachleute aus Unternehmen und Verbänden umfassen können. Die Fachkomitees bilden sich ein Urteil darüber, ob die Ziele der Institutionen für den jeweiligen Studiengang relativ zu so genannten „subject benchmarks“ adäquat sind und ob sie ihre selbst gesteckten Ziele erreichen. Sie bewerten das Studienangebot in sechs Dimensionen:

- Aufbau, Inhalt und Organisation des Studiengangs
- Lehre und Studium
- Studienerfolg
- Unterstützung und Betreuung der Studierenden
- Ausstattung für die Lehre
- Qualitätssicherung

In jeder Dimension erfolgt eine Bewertung auf einer vierstufigen Notenskala, die darüber Auskunft gibt, ob die Anstrengungen der Einrichtung in dieser Hinsicht „in vollem Umfang“, „substantiell“, „teilweise“ oder „nicht“ zur Erreichung ihrer Ziele beitragen. Die Mitglieder der Fachkommission sind gehalten, sich jegliche Information zu beschaffen, derer sie zur Substantiierung ihres Urteils bedürfen, wozu sie beispielsweise Ortsbesuche durchführen, an Beiratssitzungen der bewerteten Institution teilnehmen, Studierende befragen, Ergebnisse anderer (interner) Bewertungen heranziehen oder Klausuren, Abschlussarbeiten und Studienmaterialien einsehen können. Der Aufwand soll insgesamt niedrig gehalten werden, wenn die Bewertung eines Studienangebots eindeutig ausfällt, und nur in schwierigen Fällen entsprechend höher sein. In der Regel werden auch quantitative Daten erhoben, ein fester Satz von Indikatoren ist aber nicht vorgesehen. Die Bewertungsberichte münden in einer abschließenden Stellungnahme, in der das Studienangebot als „commendable“, „approved“ oder „failing“ bewertet wird.¹⁷⁾ Ein Vergleich mit den Studienangeboten anderer Hochschulen wird nicht gezogen.

¹⁷⁾ Abweichend in Schottland: „excellent“, „highly satisfactory“, „satisfactory“. Eine eigene Kategorie für ungenügende Angebote ist gemäß der Website des schottischen Finanzierungsrats nicht vorgesehen.

„Research Assessment Exercise“ – UK Funding Bodies

Seit 1986 wird die Forschung an den Hochschulen im Vereinigten Königreich alle fünf Jahre bewertet, um auf dieser Basis über die Verteilung von Grundmitteln für die Forschung zu befinden. Die *Research Assessment Exercise (RAE)* ist damit Teil des so genannten Dual-Support-Systems, in dem Forschungsmittel teils als Grundmittel über die Finanzierungsräte, teils projektgebunden über die Forschungsräte („Research Councils“) vergeben werden. Die Grundmittel für die Lehre werden ebenfalls von den Finanzierungsräten vergeben, jedoch vor allem nach kapazitiven Gesichtspunkten und in geringerem Maße nach Ergebnissen der TQA (s. o.).

Neben der selektiven Vergabe von Forschungsgrundmitteln ist es ein weiteres, nachrangiges Ziel der RAE, verschiedenen Interessenten in den Hochschulen und in der Öffentlichkeit Informationen über die Qualität der Forschung im britischen Hochschulsystem bereitzustellen.

Die Bewertung der Forschung an den Hochschulen erfolgt in 68 disziplinär definierten Bewertungseinheiten („units of assessment“).¹⁸⁾ Es obliegt jeder Hochschule zu entscheiden, in welcher der Bewertungseinheiten sie eine Mitteilung von Forschungsaktivitäten („submission“) macht und welche ihrer Wissenschaftler sie in welcher Bewertungseinheit als forschungsaktiv benennt. Forschungsmittel werden nur für Wissenschaftler vergeben, die in einer Mitteilung erfasst sind, es besteht jedoch keine Verpflichtung, alle Wissenschaftler für die RAE zu melden. Die Bestandteile einer Mitteilung von Forschungsaktivitäten in einer Bewertungseinheit sind in der folgenden Tabelle zusammengefasst:

Kategorie	Inhalt
Personal	<ul style="list-style-type: none">– statistische Informationen zu allen wissenschaftlichen Mitarbeitern– Detailangaben zu forschungsaktiven Mitarbeitern– statistische Informationen zu technischem und Verwaltungspersonal
Forschungsproduktion	<ul style="list-style-type: none">– bis zu vier Produkte der Forschung (Bücher, Aufsätze, CDs, Aufnahmen, Aufführungen) pro forschungsaktivem Mitarbeiter

¹⁸⁾ Alle folgenden Angaben beziehen sich auf die jüngste RAE, deren Ergebnisse im Jahr 2001 veröffentlicht wurden (RAE 2001).

Beschreibung	<ul style="list-style-type: none">- Forschungsumfeld und -organisation- Entwicklungspläne für die Forschung- Qualitative Informationen über die Forschungsleistung und -reputation
Daten	<ul style="list-style-type: none">- Forschungsmittel nach Quellen- Zahl der Forschungsstudenten- Forschungsstipendien nach Quellen- Zahl der Forschungspreise- Indikatoren der Reputation („peer esteem“)

Für jede Bewertungseinheit¹⁹⁾ wird ein Gutachtergremium von neun bis 18 überwiegend akademischen, teilweise aber auch aus der Wirtschaft kommenden Wissenschaftlern zusammengestellt. Um die Bewertung internationaler Exzellenz abzusichern, werden ausländische Experten konsultiert. Die Gutachtergremien bewerten alle Hochschulen, die in ihrer Bewertungseinheit eine Mitteilung gemacht haben, auf Basis dieser Mitteilung, unter besonderer Berücksichtigung der eingereichten Forschungsprodukte. Es finden keine Ortsbesuche statt.

Die Bewertung erfolgt auf einer siebenstufigen Skala von 1 bis 5*, wobei die einzelnen Noten durch den Anteil der Forschungsaktivitäten definiert sind, die nationale respektive internationale Exzellenzstandards erfüllen. Die Gutachtergremien verständigen sich vorab auf eine Gewichtung der verschiedenen Daten und ihre Bewertungsprinzipien. Diese Entscheidung veröffentlichen sie, ggf. gemeinsam mit einer Erklärung über Themen, die in den Mitteilungen der Hochschulen besonders berücksichtigt werden sollten. Auf diese Weise soll unterschiedlichen Bewertungsmaßstäben in den einzelnen Fächern Rechnung getragen werden.

Es wird keine Rangliste gebildet (die RAE ist kein Ranking, sondern ein Rating), anders als bei der TQA werden jedoch alle Hochschulen und alle Fächer gleichzeitig bewertet, so dass die tabellarisch veröffentlichten Noten als Ranggruppen gelesen werden können.

Die in der RAE vergebenen Noten werden in Förderfaktoren übersetzt, die in eine Formel für die Vergabe von Forschungsgrundmitteln durch die Finanzierungsräte

¹⁹⁾ Mit Ausnahme einiger „joint panels“; insgesamt sind 60 Komitees für die 68 Bewertungseinheiten zuständig.

eingehen.²⁰⁾ Basisgröße ist die Zahl der als forschungsaktiv benannten Wissenschaftler, wobei zwischen kostengünstigen und kostenintensiven Bewertungseinheiten mit einem fachspezifischen Faktor zwischen 1 und 1,6 differenziert wird.

Die RAE ist in den Jahren 2002 bis 2004 einer grundlegenden Überprüfung unterzogen worden (HEFCE 2003). Im Ergebnis ist unter anderem ein neues Bewertungssystem beschlossen worden (RAE 2004). Es sieht vor, die Forschungsaktivitäten einer Hochschule in einer Bewertungseinheit nicht mehr pauschal auf einer siebenstufigen Skala zu bewerten, sondern jeweils den prozentualen Anteil von „1 Stern“- bis „4 Sterne“-Forschung in der mitgeteilten Aktivität zu benennen. Es sei nicht Aufgabe des Ratings, dieses so genannte Forschungsprofil auf eine eindimensionale Metrik abzubilden. Vielmehr solle es dem Nutzer überlassen bleiben, ob er eher am Mittelwert der Forschungsqualität, am Anteil der Spitzenforschung oder an der Gesamtkapazität interessiert ist.

Eine weitere empfohlene Veränderung ist, den Bewertungsgruppen eine größere Autonomie in der Definition ihrer Kriterien zu geben. Sie sollen dabei der Entwicklung quantitativer Indikatoren, die auf Basis der üblichen einzureichenden Daten oder auf Basis bereits existierender Datenbanken berechnet werden können, sowie der Entwicklung von Kriterien zur Bewertung anwendungsnahe und praxisorientierter Forschung besondere Aufmerksamkeit widmen. Auch die einzureichenden Materialien und Daten können von jeder Bewertungsgruppe nach Maßgabe ihrer Kriterien modifiziert werden. So soll es künftig möglich sein, dass die Zahl der pro Wissenschaftler einzureichenden Forschungsprodukte geringer angesetzt wird als bisher (zwei oder drei statt vier). Ferner soll es möglich sein, eine Anzahl von Forschungsprodukten als Gemeinschaftswerk einer Gruppe von Forschern einzureichen.

Bei der Überprüfung der RAE stellte sich heraus, dass es schwierig ist, über die verschiedenen Forschungsgebiete hinweg einheitliche Bewertungsstandards aufrecht zu erhalten. Um die Konsistenz der Maßstäbe zu verbessern, soll künftig oberhalb der Ebene der fachbezogenen Gutachtergruppen, von denen es etwa 70 geben soll, eine Ebene von 15–20 „Hauptgutachtergruppen“ geschaffen werden, die sowohl die

²⁰⁾ Für Forschungsaktivitäten mit RAE-Note 3a war der Förderfaktor im Jahr 2003 = 0, für Note 4 = 1, für Note 5 = 2,793, für 5* = 3,357.

Kriterien der eigentlichen Gutachtergruppen kontrollieren als auch die endgültige Entscheidung über die Bewertungen der Forschungsaktivitäten fällen sollen.

Die direkte Koppelung an die Allokation von Forschungsgrundmitteln wurde im Rahmen der Überprüfung als vorrangiges Ziel der RAE bestätigt. Sie solle nach Möglichkeit durch eine Vorabklärung über die Förderfaktoren seitens der Finanzierungsräte noch transparenter gemacht werden.

Forschungsbewertung in den Niederlanden – NWO, VSNU, KNAW

In den Niederlanden werden seit 2003 alle öffentlich geförderten Forschungsaktivitäten nach einer gemeinsamen Initiative der Vereinigung Niederländischer Universitäten (VSNU), der Niederländischen Organisation für Wissenschaftliche Forschung (NWO) und der Königlich Niederländischen Akademie der Wissenschaften (KNAW) im sechsjährigen Turnus bewertet.²¹⁾ Die Initiative hat drei Ziele: die Qualität der Forschung zu verbessern, die Qualität des Forschungsmanagements zu verbessern, und einer Rechenschaftspflicht der Forschungseinrichtungen gegenüber ihren Trägern und Finanziers sowie gegenüber der niederländischen Gesellschaft nachzukommen.²²⁾ Mit Blick auf die erstgenannten Ziele sind die Trägerorganisationen, die Institutsleitungen und verantwortliche Wissenschaftler Adressaten des Prozesses.

Bewertungsgegenstand sind Institute der Universitäten und der außeruniversitären Einrichtungen in der Verantwortung von NWO und KNAW, die wegen der unterschiedlichen Organisationsweisen der Forschung in diesen Sektoren nur sehr allgemein als „Gruppen von Forschern mit einer gemeinsamen Mission, die unter dem gleichen Management operieren“, definiert sind. Die Aktivitäten jedes Instituts werden weiter unterteilt in thematisch kohärente Forschungsprogramme.

Jede Organisation – NWO, KNAW und die einzelnen Universitäten – ist für die Bewertung der in ihrer Zuständigkeit befindlichen Einrichtungen nach den vereinbarten Standards zuständig. Zur Bewertung eines Instituts wird eine Gutachtergruppe eingesetzt, deren Zusammensetzung sich nach der Mission des Instituts richtet.

²¹⁾ Parallel erfolgt an den Universitäten auch eine flächendeckende Qualitätsbewertung der Lehre.

²²⁾ NWO, VSNU & KNAW (2002).

Basis der Bewertung ist ein Selbstevaluationsdokument des jeweiligen Instituts, das zum einen standardisierte Informationen über das ganze Institut und über die einzelnen Forschungsprogramme enthält:

Aggregationsebene	Information
Institut	<ul style="list-style-type: none">- Mission des Instituts- Organisation und Management- Strategie und Taktik- Personal- Ressourcen, Fördermittel, Infrastruktur- Prozesse der Forschung, der internen und der externen Kooperation- Akademische Reputation- Interne Evaluation- Externe Bestätigung der Selbstevaluation
Forschungsprogramm	<ul style="list-style-type: none">- Organisation und Management- Strategie und Taktik- Prozesse der Forschung, der internen und der externen Kooperation- Akademische Reputation- Interne Evaluation- Personal- Ressourcen, Fördermittel, Infrastruktur- drei bis fünf Publikationen oder andere Forschungsprodukte, die die Qualität der Forschung belegen- Auflistung sämtlicher Publikationen- quantitative Übersicht über die Publikationen nach vorgegebenen Kategorien

Ergänzend wird jedes Institut gebeten, eine Selbstbewertung im Sinne einer „SWOT-Analyse“ (Stärken, Schwächen, Chancen und Risiken) vorzulegen und auf dieser Basis seine künftige Strategie zu begründen.

Nachdem sie das Selbstevaluationsdokument erhalten hat, besucht die Gutachtergruppe das Institut und spricht mit dessen Direktor bzw. dem Vorstand, den Leitern der einzelnen Arbeitsbereiche (Forschungsprogramme), dem Beirat sowie weiteren Personen oder Personengruppen, die ein Gespräch mit den Gutachtern wünschen.

Ergebnis der Bewertung ist ein schriftlicher Bewertungsbericht, der einem vorgegebenen Muster entspricht und Bewertungen des ganzen Instituts sowie der einzelnen Forschungsprogramme enthält. Auf der Ebene der Forschungsprogramme wird eine Benotung nach den Kriterien Qualität, Produktivität, Relevanz und Vitalität entlang

einer fünfstufigen Skala vorgenommen. Die Werte der Skala werden verbal definiert, wobei „exzellent“ eine internationale Führungsrolle bedeutet, „sehr gut“ eine international kompetitive, national führende Aktivität. Die einzelnen Kriterien werden in Teilkriterien weiter untergliedert, um eine umfassende Bewertung aller Aspekte der Forschungsaktivität sicherzustellen.

Der Bewertungsbericht wird zusammen mit dem Selbstevaluationsdokument der Leitung der jeweiligen Trägerorganisation (NWO, KNAW) bzw. dem Hochschulrat vorgelegt, die/der daraus Schlussfolgerungen über die Zukunft des Instituts zieht.

Bewertungsbericht, Selbstevaluationsdokument und Stellungnahme der Leitung bilden zusammen das Ergebnis der Forschungsbewertung. Sie sollen möglichst zeitnah veröffentlicht werden.

Durch die Verwendung einer standardisierten Notenskala ist ein Vergleich der verschiedenen Institute und Arbeitsbereiche in einem Fachgebiet möglich. Eine Meta-Evaluation durch eine unabhängige Kommission soll die Einhaltung einheitlicher Standards überprüfen.

A.III. Rankings im deutschen Wissenschaftssystem

Die Geschichte der Rankings im deutschen Wissenschaftssystem reicht bis in die 70er Jahre zurück. Dabei standen zunächst Rankings ganzer Hochschulen im Vordergrund. Eine Vielzahl von Methoden wurde im Laufe der Zeit in einzelnen Fächern oder Fachgebieten erprobt. Seit ungefähr 15 Jahren werden fachlich differenzierte, ein breites Fachspektrum abdeckende Rankings regelmäßig wiederholt. Große öffentliche Aufmerksamkeit zogen vor allem die in Publikumszeitschriften veröffentlichten Rankings auf sich, beginnend mit dem Spiegel-Ranking von 1989.²³⁾ Am bekanntesten ist heute das jährlich in der Zeitschrift Stern als „Studienführer“ erscheinende Hochschulranking des Centrums für Hochschulentwicklung (CHE).

Wie die Mehrzahl der internationalen Beispiele richten sich auch die Ranking-Initiativen deutscher Publikumszeitschriften in erster Linie an Studienanfänger und

²³⁾ Rosigkeit (1997) zitiert als frühestes Beispiel eine Rangliste aus der Zeitschrift *Der Wiener – Zeitschrift für Zeitgeist* von 1987.

ihre Eltern, denen sie bei der Wahl eines Studienortes Entscheidungshilfe geben sollen. Bei bundesweiten Befragungen gab schon vor einigen Jahren gut ein Viertel der Studienanfänger an, dass der – in Rankings dokumentierte – gute Ruf der Hochschule ein „Hochschulwahlmotiv mit großer Bedeutung“ gewesen sei. Eine Auswertung von Daten der Zentralstelle für die Vergabe von Studienplätzen (ZVS) ergab, dass die Zahl der Bewerber an den besser bewerteten Universitäten nach Veröffentlichung von Hochschulranglisten um etwa 20 Prozent steigt.²⁴⁾ Den größten Einfluss auf die Studienortwahl hat freilich zumindest bei Studierenden der Wirtschaftswissenschaften, für die entsprechende Umfragen vorliegen, nach wie vor die Attraktivität des Hochschulstandorts, insbesondere die Nähe zum Heimatort sowie die Lebenshaltungskosten.²⁵⁾

In letzter Zeit sind die Forschungsaktivitäten der Hochschulen vermehrt Gegenstand von Ranking-Initiativen. Nachdem die Deutsche Forschungsgemeinschaft bereits 1997 und 2000 differenzierte Berichte über ihre Bewilligungen abgegeben und dabei einen Zusammenhang zu den Ranglisten der Publikumszeitschriften hergestellt hatte, hat sie diese Berichterstattung weiter fortentwickelt und im Jahr 2003 eine aktuelle Fassung unter Einbeziehung weiterer Daten als „Förder-Ranking“ publiziert, das Basisdaten über die öffentlich geförderte Forschung bereitstellt.²⁶⁾ Im Jahr 2002 hat das CHE auf Basis seines Hochschulrankings ein Forschungsranking veröffentlicht, als dessen primäre Adressaten Wissenschaftler und der wissenschaftliche Nachwuchs genannt werden. In der Fortschreibung des Forschungsrankings (2003) wurde auch der Versuch unternommen, die forschungstärksten Universitäten Deutschlands zu identifizieren.²⁷⁾

Frühe Rankings ganzer Hochschulen

Frühe Beispiele für (Forschungs-)Rankings ganzer Hochschulen in Deutschland sind nach Daniel (1988):

- der Vergleich deutscher Universitäten nach ihren Publikationszahlen und nach einem Zitationsindex, gebildet aus Daten des Science Citation Index, durchgeführt von Spiegel-Rösing (1975) im Auftrag des Bundesministeriums für Bildung und Wissenschaft;

²⁴⁾ Daniel (2001).

²⁵⁾ Büttner et al. (2002), Fabel et al. (2002).

²⁶⁾ Deutsche Forschungsgemeinschaft (2003), S. 5.

²⁷⁾ Centrum für Hochschulentwicklung (2002), Berghoff et al. (2003b).

- die Universitäts-Ranglisten der Enquete „Zur Lage der Forschung an den Universitäten“ nach der – auf fachspezifische Durchschnittswerte normierten – Publikationsproduktivität ihrer Professoren, gebildet auf Basis einer Repräsentativumfrage des Instituts für Demoskopie Allensbach im Wintersemester 1976/77;
- die Beliebtheitsskala deutscher Universitäten, die von der Alexander von Humboldt-Stiftung durch eine Auswertung ihrer Bewilligungsbescheide gebildet und 1981 publiziert wurde.

Alle drei Initiativen eint, dass sie ganze Universitäten vergleichen und dass jeweils nur wenige Indikatoren verwendet werden. Angesichts der geringen Übereinstimmungsvalidität der veröffentlichten Ranglisten (Daniel I. c.), des sporadischen Auftretens der Initiativen sowie der offenen Zielsetzungen kann hier von einer Früh- und Experimentierphase der Rankingmethodik im deutschen Wissenschaftssystem gesprochen werden.

1985 erörterte der Wissenschaftsrat die Einführung von Ranking-Verfahren zur Erhöhung der Transparenz des Wissenschaftssystems als Voraussetzung des Wettbewerbs,²⁸⁾ ohne dass dies zum damaligen Zeitpunkt konkrete Folgen gehabt hätte.

Das Hochschulranking des CHE

Die erstmals 1998 erstellten, seit 1999 als Studienführer der Zeitschrift Stern publizierten Hochschulranglisten des Centrums für Hochschulentwicklung sollen Studienanfänger bei der Studienortwahl unterstützen. Darüber hinaus soll die Angebots- und Leistungstransparenz im Hochschulbereich verbessert werden.²⁹⁾ Das Hochschulranking des CHE legt ein mehrdimensionales Entscheidungsmodell zugrunde und umfasst fachspezifische Daten zur Studiensituation, einige Forschungsindikatoren sowie Urteile von Professoren und Studierenden, ohne sie zu einer Gesamtrangliste zusammenzufassen.³⁰⁾ Ein Gesamtwert für eine Hochschule würde nach Ansicht des CHE Differenzierungen in der Bewertung von Forschung, Lehre, Betreuung, Ausstat-

²⁸⁾ Um die Transparenz des Hochschulsystems zu erhöhen, empfahl der Wissenschaftsrat 1985 zwei Schritte: erstens regelmäßige Selbstdarstellungen der Hochschulen, und zweitens eine vergleichende Bewertung von Leistungen. „Für das Bewertungsverfahren“, erklärte er, „können Methoden des Ranking, die das amerikanische Hochschulsystem entwickelt hat, Hinweise geben“, und präzierte weiter, „dass es sinnvollerweise nur Fächer sein können, die beurteilt werden, nicht ganze Hochschulen. Wichtig ist ferner, dass einzelne Indikatoren nicht überbewertet, Beurteilungen vielmehr auf eine Vielzahl unterschiedlicher Indikatoren gegründet werden. Auch erscheint es nicht erforderlich, auf eine Rangordnung abzielen, in der jeder Fachbereich seinen bestimmten Platz hat: Ranggruppen dürften genügen.“ (Wissenschaftsrat 1985, S. 27).

²⁹⁾ Stern (2003); Berghoff et al. (2003a).

³⁰⁾ Zur wissenschaftlichen Bewertung der Methodik des Hochschulrankings vgl. Hornbostel (2001).

tung etc. verwischen. Auch bei den einzelnen Indikatoren verzichtet das CHE auf Ranglisten im engeren Sinne, sondern stellt die Ergebnisse in Form von Ranggruppen (erstes Quartil, Mittelgruppe, letztes Quartil) dar. Herausgehoben werden allerdings anhand einer Auswahl von Kriterien jeweils eine kleine Zahl von Hochschulen als „Studentipps“ für drei unterschiedliche Typen von Studierenden, die „Zielstrebigen“, „Forscher“ und „Praktiker“.

Einheit der Bewertung sind Studiengänge an den einzelnen Hochschulen, wobei sich die zugrunde gelegten Daten und Umfrageergebnisse teils auf Studienbereiche, teils auf Fachbereiche, teils auf Studiengänge beziehen. Mit der Veröffentlichung des Studienführers 2003 sind 34 Fachgebiete durch CHE-Rankings unterschiedlicher Aktualität abgedeckt.

Die verschiedenen Kriterien des zugrunde gelegten Entscheidungsmodells werden vom CHE in einer Matrix dargestellt:

Studierende	Studienergebnis	Internationale Ausrichtung
Forschung	Studium und Lehre	Ausstattung
Berufsbezug, Arbeitsmarkt	Gesamturteile Studierende, Professorentipp	Studienort und Hochschule

Jeder dieser neun Dimensionen sind mehrere Indikatoren zugeordnet:

1. Dimension „Studierende“: Diese Dimension wird durch Informationen über die Zahl der Studierenden, Studienanfänger und Bewerber sowie durch den Frauenanteil und die Abbrecherquoten charakterisiert.
2. Dimension „Studienergebnis“: In dieser Dimension werden absolute Absolventenzahlen, Notendurchschnitte, die mittlere Fachstudiendauer (Median) und die Verteilung der Absolventen nach Fachsemesterzahl angegeben.
3. Dimension „Internationale Ausrichtung“: Informationen zu dieser Dimension umfassen Angaben zur Möglichkeit von Doppeldiplomen und zur Teilnahme am Eu-

ropean Credit Transfer System, zur Existenz von fachspezifischen Fremdsprachenangeboten, zu obligatorischen Auslandsaufenthalten und zu Austauschprogrammen sowie zum Anteil ausländischer Studierender und Gastprofessoren.

4. Dimension „Forschung“: Umfasst Daten und Urteile. Als Daten werden Drittmittel je Wissenschaftler / je Professor, in einigen Fächern Patente je Professor, Publikationen je Professor sowie Promotionen und zum Teil Habilitationen je Professor ausgewiesen. Zusätzlich werden die Ergebnisse einer Professorenbefragung angegeben, bei der die Professoren des Fachs gebeten werden, (bis zu) drei Hochschulen zu nennen, die ihrer Einschätzung nach in ihrem Fach in der Forschung in Deutschland führend sind.
5. Dimension „Studium und Lehre“: Neben Daten zur Betreuungsrelation (Studierende je Professor), zur Durchführung von Lehrevaluationen sowie zu einer Reihe von studienrelevanten Dienstleistungen beinhalten die Informationen zu dieser Dimension die Ergebnisse von Studierendenbefragungen, die ihrerseits mehrere Aspekte der Studienbedingungen zum Gegenstand haben: die Studienberatung, das Studienangebot, die Betreuung durch Lehrende, die Kommunikation zwischen Studierenden und Lehrenden sowie unter den Studierenden und die Studien- und Prüfungsorganisation.
6. Dimension „Ausstattung“: Auch zu dieser Dimension werden Daten und Studierendurteile angeboten. Die Daten sind fachspezifisch und umfassen beispielsweise Angaben zur Ausstattung mit Computern, zu Laborplätzen, Sachmitteln, Planbetten etc. Die Studierendenbefragungen betreffen die Qualität und Verfügbarkeit von PC-, Arbeits- bzw. Laborplätzen, die Bewertung der Bibliotheken, die Bewertung des Raumangebots und der Ausstattung mit audiovisuellen Medien.
7. Dimension „Arbeitsmarkt- und Berufsbezug“: Hierzu werden Daten über spezifische berufs- und arbeitsmarktbezogene Veranstaltungen (Stundenzahl je 100 Studierende) sowie Urteile von Studierenden über das Angebot an Maßnahmen zur Förderung des Berufs- und Arbeitsmarktbezugs zusammengefasst.
8. Dimension „Gesamturteile“: In diese Dimension gehen Antworten von Studierenden auf eine Frage nach der Studiensituation in ihrem Fach sowie die so ge-

nannten Professorentipps – Professoren werden gebeten, bis zu drei empfehlenswerte Hochschulen in ihrem Fach zu nennen – ein.

9. Dimension „Studienort und Hochschule“: Ergänzend zu den fachspezifischen Informationen stellt das CHE eine Reihe von Informationen über die Bevölkerung, den Studierendenanteil, die Wohnsituation von Studierenden und das Verkehrsangebot sowie über Studierendenzahlen, Semesterbeiträge und allgemeine Dienstleistungen der Hochschule (Studienberatung, Hochschulsport etc.) zusammen.

In der gedruckten Fassung des CHE-Hochschulrankings wird jeweils nur eine kleine, für besonders aussagekräftig gehaltene Auswahl der Indikatoren pro Fach wiedergegeben. Die Nutzer der Internetversion können sich jedoch eine individuelle Rangliste auf Basis von bis zu fünf aus der gesamten Palette der erhobenen Daten ausgewählten Indikatoren erstellen lassen.

Das Forschungsranking des CHE

Seit 2002 erstellt das CHE durch eine Sonderauswertung von Daten, die im Rahmen seiner Hochschulrankings erhoben wurden, auch Forschungsrankings. Ziel ist es nach Angaben des CHE, „die forschungsaktiven Fakultäten in Deutschland zu lokalisieren“.³¹⁾

Das CHE-Forschungsranking von 2003³²⁾ deckt 13 Fächer aus den Natur-, Geistes- und Sozialwissenschaften ab. Rankings für den Maschinenbau und die Elektrotechnik sind zwar erhoben worden, aus unterschiedlichen methodischen Gründen in der aktuellen Publikation aber nicht enthalten.

Das Forschungsranking des CHE umfasst drei Indikatoren für Forschungsaktivität: verausgabte Drittmittel, Publikationen und Promotionen. In vier Fächern werden auch Zitationen erhoben (Biologie, Chemie, Physik, Pharmazie). Bei den Drittmitteln handelt es sich um die in einem Zeitraum von drei Jahren verausgabten Drittmittel einer Fakultät im jeweiligen Fach, die per Umfrage ermittelt wurden; Publikationszahlen wurden durch bibliometrische Analyse geeigneter Datenbanken erhoben, wobei in

³¹⁾ www.dashochschulranking.de/allgemeines_fr.php, 15. 12. 2003.

³²⁾ Berghoff et al. (2003b).

einigen Fächern eine Gewichtung nach Typ und Länge der Publikationen vorgenommen wurde.³³⁾ Promotionszahlen wurden für einen Zeitraum von vier Semestern im Rahmen der Fachbereichsbefragung erhoben. Alle Indikatoren werden im CHE-Forschungsranking jeweils absolut und relativiert auf Wissenschaftler (Drittmittel) respektive auf Professoren (Publikationen, Promotionen) wiedergegeben.

Die Reputation wurde im Rahmen einer Professorenbefragung ermittelt, bei der die Befragten gebeten wurden, drei Universitäten zu nennen, die sie für ein Studium in ihrem Fach empfehlen würden bzw. die nach ihrer Meinung in ihrem Fach in der Forschung führend sind. Im Forschungsranking attestiert das CHE eine hohe Reputation, wenn eine Universität von mehr als 5 % der Befragten genannt wird.

Das CHE definiert forschungsstarke Fakultäten, indem es zunächst für jeden der Indikatoren eine Rangliste aufstellt und anhand dieser Listen Spitzengruppen identifiziert. Bei den absoluten Indikatoren besteht die Spitzengruppe aus denjenigen am höchsten gerankten Universitäten, die zusammen 50 % des jeweiligen Wertes auf sich vereinen; bei den relativen Indikatoren gilt das erste Quartil der Rangliste als Spitzengruppe. Als forschungsstark gelten nun genau die Fakultäten, die bezüglich mindestens der Hälfte der Indikatoren, die für das jeweilige Fach erhoben wurden (absolute und relative, aber ohne Reputation), zur Spitzengruppe gehören.

In einem weiteren Aggregationsschritt definiert das CHE geistes- und naturwissenschaftliche Forschungsuniversitäten als diejenigen Universitäten, deren in den 13 CHE-Rankings erfasste Fakultäten mindestens zur Hälfte als forschungsstark gelten. Nach diesem Kriterium hat das CHE sieben Universitäten in Deutschland als Forschungsuniversitäten identifiziert.

Das Förder-Ranking der DFG

Im Herbst 1996 veröffentlichte die Deutsche Forschungsgemeinschaft auf Anregung einer Gruppe von Rektoren und Präsidenten Daten über die zehn Hochschulen, die in den Jahren 1991–1995 den höchsten Einwerbungserfolg bei der DFG hatten. Angesichts der lebhaften Diskussionen, die diese Veröffentlichung auslöste, entschloss sie sich kurz darauf in Abstimmung mit der Hochschulrektorenkonferenz, in breiterem

³³⁾ Zur Kritik an den Publikationsindikatoren des CHE vgl. Ursprung (2003), dagegen Berghoff & Hornbostel (2003).

Rahmen über ihre Bewilligungen zu berichten und damit einem großen Interesse an Vergleichsdaten Rechnung zu tragen. In der resultierenden Publikation³⁴⁾ wird auf die Eignung von Drittmitteldaten als Indikator für Forschungsaktivität beziehungsweise, sofern es sich wie im Falle der DFG um begutachtete Anträge handelt, auch als Leistungsindikator in qualitativer Hinsicht ausdrücklich hingewiesen. Obwohl die DFG nicht behauptete, ein Forschungsranking der Hochschulen vorzulegen, wurde die Veröffentlichung nicht zuletzt aufgrund der Art der Darstellung – die Hochschulen werden im tabellarischen Anhang nicht etwa alphabetisch, sondern in absteigender Reihenfolge nach dem Einwerbungserfolg, absolut oder je Professur, aufgelistet – in der Öffentlichkeit als ein solches rezipiert.

Nach einem bereits erheblich erweiterten, auch außeruniversitäre Einrichtungen einbeziehenden Nachfolgebericht im Jahr 2000 erschien 2003 eine Publikation, die unter dem Titel „Förder-Ranking“ erstmals nicht nur Daten der DFG selbst und des Statistischen Bundesamtes, sondern auch solche der Alexander von Humboldt-Stiftung und des Deutschen Akademischen Austauschdienstes, der Koordinierungsstelle EG der Wissenschaftsorganisationen sowie bibliometrische Analysen enthält.³⁵⁾

Angesichts der Konjunktur von Ranglisten in Deutschland beabsichtigt die DFG mit ihrer Veröffentlichung, die Diskussion über die Bewertung von Forschung auf eine breitere Grundlage zu stellen und so einen Beitrag zur Definition einer „best practice“ für die Aufstellung von Ranglisten wissenschaftlicher „Exzellenz“ zu leisten. Die Ergebnisse der Studie werden in fünf Rubriken zusammengefasst:

1. Im Mittelpunkt des Förder-Rankings steht eine Analyse der DFG-Bewilligungen an Hochschulen und außeruniversitäre Einrichtungen in den Jahren 1999 bis 2001. Neben den absoluten Bewilligungssummen je Einrichtung, insgesamt sowie differenziert nach Wissenschaftsgebieten und nach Fachgebieten, finden sich für die einzelnen Wissenschaftsgebiete auch Angaben über die Bewilligungen je Hochschule und Programmgruppe³⁶⁾ sowie über Bewilligungen je Hochschule und Professur und je Hochschule und Wissenschaftler. Untersucht wird

³⁴⁾ Deutsche Forschungsgemeinschaft (1997).

³⁵⁾ Deutsche Forschungsgemeinschaft (2003)

³⁶⁾ In den Programmgruppen „Einzelförderung“, „Koordinierte Programme“, „Direkte Nachwuchsförderung“ und „Preise“ fasst die DFG ihre Förderinstrumente unter strukturellen Gesichtspunkten zusammen.

zudem das Verhältnis von DFG-Bewilligungen zu den von den Hochschulen insgesamt eingeworbenen Drittmitteln, wobei sich nach Einrichtungen eine starke Korrelation zeigt, während sich von Fachgebiet zu Fachgebiet deutliche Unterschiede zeigen. Für die außeruniversitären Einrichtungen werden Bewilligungen je Wissenschaftsbereich und Bewilligungen je Programmgruppe angegeben.

2. Neu hinzugetreten ist 2003 eine Analyse der vernetzten Forschung auf Basis der Daten, die über die Förderung in koordinierten Programmen der DFG³⁷⁾ vorliegen. Ausgewertet wird die anhand der institutionellen Adressen der Teilprojektleiter bestimmte gemeinsame Beteiligung von Einrichtungen an koordinierten Programmen. Die Zentralität von Einrichtungen in wissenschaftlichen Netzwerken der einzelnen Wissenschaftsbereiche wird durch die Zahl ihrer Partneereinrichtungen definiert. Indem Verknüpfungen zwischen Einrichtungen dargestellt werden, die an einer (je nach Wissenschaftsgebiet unterschiedlichen) Mindestzahl von koordinierten Programmen gemeinsam beteiligt sind, visualisiert der Bericht Kernstrukturen von Kooperationsnetzwerken in DFG-Projekten. In der Online-Version des Förder-Rankings ist zudem eine dynamisierte Darstellung verfügbar, die es erlaubt, die Kooperationsbeziehungen einzelner Einrichtungen zu verfolgen.
3. Ebenfalls neu ist auch die Auswertung der institutionellen Zugehörigkeit von DFG-Gutachtern, die nach Auffassung der DFG ein guter Indikator für die an einem Ort versammelte Forschungsexpertise ist. Datenbasis stellen die im Erhebungszeitraum von der DFG erbetenen schriftlichen Gutachten dar. Im Förder-Ranking wird die absolute Zahl der Gutachter je Hochschule und Wissenschaftsbereich wie auch je Hochschule und Fachgebiet aufgeführt. Zudem wird die Korrelation von Gutachterzahlen und Bewilligungssummen analysiert.
4. Zur Internationalität von Forschung stellt die DFG eine Reihe von Daten bereit: Die Zahl der von der Alexander von Humboldt-Stiftung geförderten Gastwissenschaftler je Hochschule und Wissenschaftsbereich sowie je Hochschule und Fachgebiet (1997–2001), geförderte Gastwissenschaftler im Verhältnis zu den

³⁷⁾ Es handelt sich um Sonderforschungsbereiche, Schwerpunktprogramme, Forschergruppen, Graduiertenkollegs und Geisteswissenschaftliche Zentren. Forschungszentren waren im Erhebungszeitraum noch nicht eingerichtet.

Professoren oder Wissenschaftlern der jeweiligen Hochschule je Wissenschaftsbereich, analoge Zahlen für den Deutschen Akademischen Austauschdienst (2000–2001) sowie Daten über die Beteiligung der deutschen Hochschulen am 5. Rahmenprogramm der EU (1998–2002). Ergänzend finden sich Analysen zur Herkunft von Gastwissenschaftlern und zur Nationalität von Kooperationspartnern in den EU-Projekten, die jedoch nicht auf einzelne Einrichtungen innerhalb Deutschlands heruntergebrochen werden.

5. Eine fünfte Rubrik im Förder-Ranking machen schließlich die bibliometrischen Befunde aus. Die DFG wertet dazu zwei bereits veröffentlichte Studien³⁸⁾ aus, beschränkt sich aber auf die Frage der Korrelation zwischen DFG-Bewilligungen und Publikationen (Zitationen soweit verfügbar auf der Aggregationsebene Hochschule insgesamt sowie für Hochschulen im Fachgebiet Medizin. Lediglich für DFG-Bewilligungsranggruppen, nicht jedoch für einzelne Hochschulen untersucht sie zudem die Zahl der Publikationen je Professur / je Wissenschaftler und den relativen Zitationsindex im Fachgebiet Medizin.

In der Zusammenfassung stellt die DFG die Ergebnisse der einzelnen Messungen in Form eines Ranggruppenvergleichs der Hochschulen für die wichtigsten erhobenen Kennzahlen (absolut und auf Professuren relativiert) gegenüber und bietet damit ebenfalls ein multidimensionales Ranking an.

Forschungsevaluation an niedersächsischen Hochschulen und außeruniversitären Forschungseinrichtungen

Seit 1999 führt die Wissenschaftliche Kommission Niedersachsen innerhalb des Bundeslandes Niedersachsen ein flächendeckendes Evaluationsverfahren nach einheitlichen Kriterien durch.³⁹⁾ Sie strebt damit an,

- die Hochschulen bei der Entwicklung eines eigenen Forschungsprofils zu unterstützen;
- den Hochschulen Kriterien für die Planung und Durchführung von Maßnahmen zur Qualitätssicherung und -verbesserung an die Hand zu geben;

³⁸⁾ Die DFG zitiert eine Studie des schweizerischen Zentrums für Wissenschafts- und Technologiestudien (CEST) zum Publikationsaufkommen von Hochschulen insgesamt sowie eine Studie des niederländischen Zentrums für Wissenschafts- und Technologiestudien (CWTS) zu Publikationen und Zitationen in der medizinischen Forschung.

³⁹⁾ Wissenschaftliche Kommission Niedersachsen (2002).

- die Profilbildung der Hochschulen von Seiten des Landes gezielter zu gestalten und Leitlinien für die Strukturplanung der Hochschulen zu erarbeiten;
- zur Entwicklung von Kriterien für die qualitätsorientierte Mittelvergabe durch das Land beizutragen;
- die Leistungen der Hochschulen transparenter zu machen.

Die Forschungsevaluationen werden fachbezogen durchgeführt.⁴⁰⁾ Für jedes zu evaluierende Fach wird eine Gutachtergruppe eingerichtet. Die Begutachtungen erfolgen auf Basis von standardisierten Selbstberichten der Hochschulen und Begehungen durch die Gutachtergruppen, in deren Rahmen Gespräche mit der Hochschulleitung, den Fachbereichen oder Fakultäten, Fachvertretern und den einzelnen Forschungseinheiten bzw. Forscherinnen und Forschern erfolgen. In diesem Sinne handelt es sich um ein Verfahren vom Typ „Informed Peer Review“.

Die Gutachter sind gehalten, die Forschungsleistungen der Einrichtungen nach Qualität und Relevanz sowie nach Effektivität und Effizienz zu beurteilen. Hinsichtlich Qualität und Relevanz werden mehrere Unterkriterien angegeben:

- Innovativität der Forschung
- wissenschaftliche Ausstrahlung
- Interdisziplinarität
- Kooperationen mit anderen Forschungseinrichtungen
- internationale Zusammenarbeit
- Nachwuchsförderung
- Kooperationen mit Wirtschaft, Verwaltung, Politik und Kulturinstitutionen

Im Rahmen der Evaluationen wird die Forschungsqualität der einzelnen Forschungseinheiten auf einer Skala von 1 bis 10 bewertet, wobei 10 für beste Forschungsleistungen steht. Auf Gesamtbewertungen der einzelnen Standorte wird verzichtet. Ratinginformationen dienen primär dazu, einheitliche Standards bei der Evaluation der einzelnen Standorte zu gewährleisten.

⁴⁰⁾ Außeruniversitäre Forschungseinrichtungen in Landesfinanzierung werden institutionell evaluiert.

Die Evaluationsergebnisse und Empfehlungen der Gutachter werden in einem Abschlussbericht der Gutachtergruppe zusammengefasst und der Hochschule zur Stellungnahme zugeleitet. Die Lenkungsgruppe des Verfahrens diskutiert die Abschlussberichte und die Stellungnahmen der Hochschulen und leitet sie, ggf. mit ergänzender Stellungnahme, an die Wissenschaftliche Kommission Niedersachsen weiter. Diese berät die Evaluationsergebnisse unter strukturellen Gesichtspunkten und gibt Empfehlungen an das Land und ggf. an einzelne Hochschulen ab.

Hochschulranking nach Gleichstellungsaspekten

Das Kompetenzzentrum Frauen in Wissenschaft und Forschung hat im Jahr 2003 ein Hochschulranking nach Gleichstellungsaspekten veröffentlicht (CEWS 2003). Bezugnehmend auf den Gleichstellungsauftrag, den das Hochschulrahmengesetz den Hochschulen stellt, will es den Hochschulen einen Vergleichsmaßstab liefern, anhand dessen sie ihre Leistungen im Bereich Gleichstellung und Chancengleichheit einordnen können. Es richtet sich in erster Linie an Hochschulleitungen und das Hochschulmanagement, aber auch an Bundes- und Landesministerien, Wissenschaftsorganisationen sowie Politikerinnen und Politiker, die an der Zielerreichung einzelner Hochschulen im Bereich Gleichstellung interessiert sind.

Das Hochschulranking nach Gleichstellungsaspekten bewertet Universitäten, Fachhochschulen und Künstlerische Hochschulen jeweils als ganze Institutionen, aber nach Typen getrennt. Verwendet werden ausschließlich quantitative Indikatoren eines bestimmten Typs, nämlich der Frauenanteil und seine zeitlichen Veränderungen in verschiedenen Statusgruppen. Im Einzelnen werden die Frauenanteile in folgenden Gruppen erfasst:

- Studierende
- Promotionen
- Habilitationen
- hauptberufliches wissenschaftliches und künstlerisches Personal
- Professuren
- Veränderungen beim hauptberuflichen wissenschaftlichen und künstlerischen Personal im Zeitverlauf
- Veränderungen bei den Professuren im Zeitverlauf

Analog zum CHE-Ranking werden für jeden Indikator Ranglisten aufgestellt und dann Spitzengruppe (1. Quartil), Mittelgruppe (2. und 3. Quartil) und Schlussgruppe (4. Quartil) identifiziert. Durch ein einfaches Punktesystem (2 Punkte für jeden Platz in einer Spitzengruppe, 1 Punkt für jeden Platz in einer Mittelgruppe) werden alle sieben Indikatoren zu einer Gesamtbewertung der jeweiligen Hochschule verrechnet. Auf diese Weise werden Universitäten, Fachhochschulen und Künstlerische Hochschulen in Ranggruppen eingeordnet.

A.IV. Vergleich bestehender Verfahren vergleichender Bewertung

Ein tabellarischer Vergleich der lehr- oder forschungsbezogenen Ranking- und Ratingverfahren zeigt, dass dabei die unterschiedlichsten Kombinationen von Zielen und Verfahrensmerkmalen realisiert sind. Das Spektrum reicht von Rankingverfahren, die der eingangs skizzierten Definition formal entsprechen – bspw. das U.S. News Ranking, das allein auf Kennzahlen basiert und in der Bildung von ordinalen Ranglisten mündet – bis zu dem peer-review-basierten niederländischen Verfahren, das zwar flächendeckend vereinheitlicht ist und durch die standardisierte Notenskala Vergleiche ermöglicht, hinsichtlich der Selbstevaluationskomponente, der obligatorischen Ortsbesuche und der Ergebnisdarstellung in Form von Berichten über einzelne Einrichtungen jedoch eher einem Evaluationsverfahren entspricht.

Die Zahl der Kombinationen von Verfahrenseigenschaften würde noch ansteigen, würde man weitere Verfahren einbeziehen. Die in der Tabelle nicht vorkommende Kombination „Forschung – ganze Universitäten – quantitative Indikatoren – Rangliste“ etwa ist mit dem von Wissenschaftlern der Shanghai Jiao Tong University 2003 und erneut 2004 im Internet veröffentlichten Ranking von 500 Universitäten weltweit oder mit der „Champions League der Forschungsinstitutionen“ des schweizerischen Zentrums für Wissenschafts- und Technologiestudien (CEST) mehrfach realisiert. Aus dem internationalen Vergleich allein lassen sich also keine Rückschlüsse darauf ziehen, welche Verfahrenscharakteristika abhängig von der betrachteten Leistungsdimension und den Zielen zwingend wären. Aufschlussreich sind allerdings drei Trends.

Zum einen werden Ranglisten im engeren Sinne immer seltener verwendet. Angesichts der zum Teil unsicheren Datenlage bei einzelnen Indikatoren sowie des Prob-

lems ihrer Gewichtung gilt die Darstellung in Ranglisten als unangemessen, weil der damit implizit erhobene Genauigkeitsanspruch nicht einzulösen ist. Eine Alternative ist, Ranggruppen zu bilden, um eine Überbewertung der ordinalen Abfolge zu verhindern. Das gleiche Ziel kann erreicht werden, indem Vertrauensintervalle für Platzierungen auf einer Rangliste angegeben werden. Derartige Ranggruppen basieren entweder auf Perzentilen, die meist das untere und das obere Viertel der Verteilung abschneiden, oder auf Vertrauensintervallen, die einen auch statistisch signifikanten Unterscheid zwischen Spitzen- und Schlussgruppe sicherstellen. Derartige – auf Konfidenzintervallen beruhende – Ranggruppen wurden in den SPIEGEL- und CHE-Rankings eingesetzt. Im Zuge des Revirements der Methodologie der Bewertung von Promotionsprogrammen durch den National Research Council wurden solche Ranggruppen ebenfalls vorgeschlagen (Ostriker & Kuh (Hrsg.) 2003). Möglich ist schließlich auch ein Rating nach einer vorgegebenen Notenskala, wie es sich insbesondere bei peer-review-basierten Verfahren (s. u.) anbietet.

Einer Informationsverdichtung in Form ordinaler Ranglisten steht zum anderen auch der zweite Trend entgegen, Bewertungen nach verschiedenen Kriterien nicht mehr zu einem einzigen Wert zu verrechnen, sondern für den Nutzer transparent zu machen. Dazu wird häufig eine mehrdimensionale Darstellung gewählt, die es erlaubt, die Position der bewerteten Einheiten hinsichtlich verschiedener Bewertungsdimensionen zu veranschaulichen.

Ein dritter Trend ist die zunehmende Verschränkung von Expertenurteilen und quantitativen Indikatoren. Peer-Review ist international etablierter Standard in der Bewertung wissenschaftlicher Qualität. Allerdings sind der Verlässlichkeit von Peer-Review Grenzen gesetzt.⁴¹⁾ In jedem Fall benötigen Gutachter eine solide Datenbasis, damit die Ergebnisse von Begutachtungen nicht zu reinen Reputationsmaßen mit deren bekannten Nachteilen (Subjektivität, Verzögerung, Halo- und Star-Effekte)⁴²⁾ geraten.

⁴¹⁾ Bornmann & Daniel (2003).

⁴²⁾ Vgl. dazu "Research Doctorate Programs in the U.S.", S. 6 ff.

	U.S. News	Research Doctorate Programs 1993 (U.S. NRC)	Research Assessment Exercise 2001 (UK)	niederländische Forschungsbewertung 2003-2009 (NOW, VSNU, KNAW)	QHE-Hochschulranking	QHE-Forschungsranking	DFG-Förderranking 2003
Leistungsdimension	Lehre	Nachwuchsförderung, Forschung	Forschung	Forschung	Lehre	Forschung	Forschung
Ziele	Studienorientierung	Strategisches Instrument; Basis für Bildungsforschung; Studienorientierung	Effiziente Verteilung von Forschungsmitteln	Strategisches Instrument; Rechenschaftspflicht der Forschungsorganisationen erfüllen	Studienorientierung	Forschungsstarke Fakultäten und Universitäten lokalisieren	Wettbewerb der Forschungseinrichtungen verbessern; Beitrag zur Methodenentwicklung von Forschungsrankings; Rechenschaftspflicht der DFG erfüllen
Objekte	Universitäten Colleges	Promotionsprogramme der Universitäten nach Fächern	Forschungsaktivitäten der Universitäten nach Fächern	Forschungsprogramme der Universitäten und außeruniversitären Forschungseinrichtungen	Fachbereiche, Lehrinhalte, Studiengänge nach Fächern	Fakultäten nach Fächern	Forschungsaktivitäten der Universitäten und außeruniversitären Forschungseinrichtungen nach Fachgebieten
Bewertungsbasis	Quantitative Indikatoren	Reputationsmaß, quantitative Indikatoren	Evidenzbasierte Expertenturteile	Evidenzbasierte Expertenturteile, Begehungen	Quantitative und qualitative Indikatoren, Reputationsmaß	Quantitative Indikatoren, Reputationsmaß	Quantitative Indikatoren
Ergebnisse	Ranglisten (nach Institutionstyp)	Ranglisten (nach Fächern)	Benotung, Ranggruppen (nach Fächern)	Bewertungsberichte, Benotung (nach Leistungskriterien)	Mehrdimensionale Ranglisten (nach Fachgruppen und Indikatoren); Identifikation forschungsstarker Fakultäten und Universitäten	Mehrdimensionale Ranglisten (nach Fachgebieten und Indikatoren), Netzwerkanalysen, Mehrdim. Ranggruppen	Mehrdimensionale Ranglisten (nach Fachgebieten und Indikatoren), Netzwerkanalysen, Mehrdim. Ranggruppen

Gerade bei flächendeckenden, vergleichenden Bewertungen sind quantitative Indikatoren unverzichtbar, um die Beanspruchung der Gutachter und Einrichtungen in Grenzen zu halten und trotz knapper Zeitbudgets zu verlässlichen Resultaten zu kommen. Im Forschungsbereich haben sich neben Drittmittelstatistiken vor allem verschiedene Publikations- und Zitationsindikatoren bewährt.⁴³⁾

Ausschließlich auf Indikatoren basierende Bewertungen werden allerdings immer wieder kritisiert, weil in aller Regel die Indikatoren selbst einer kompetenten Interpretation bedürfen. Trotz ihrer Vorteile (Skalenniveau, geringerer Aufwand, Objektivität) werden quantitative Indikatoren deshalb nur selten als alleinige Basis von Rankings genutzt. Am ehesten geschieht dies dort, wo starke Differenzierung mit geringer Markttransparenz zusammentrifft (U.S. News), wo vorwiegend generelle Informations- und Transparenzziele verfolgt werden (CHE-Forschungsranking) oder wo ein Vergleich auf sehr hoher Aggregationsebene durchgeführt werden soll (Ranking von Wissenschaftlern der Shanghai Jian Tong University 2003). In Verfahren, die explizit intendieren, strategische Prozesse und Allokationsentscheidungen zu stützen (RAE, niederländische Forschungsbewertung), werden quantitative Indikatoren demgegenüber genutzt, um Expertenurteile zu stützen („evidence based peer review“).

⁴³⁾ Van Raan (1996), Hornbostel (1997); als Anwendungsbeispiel Tijssen, van Leeuwen & van Raan (2002).

B. Empfehlungen

B.I. Vorbemerkungen

Die Verfügbarkeit von vergleichenden Qualitätsinformationen über die verschiedenen Leistungen der Hochschulen und außeruniversitären Forschungseinrichtungen trägt wesentlich zu einem effektiven und effizienten Wettbewerb im Wissenschaftsbereich bei. Da Qualitätsbewertungen im Wissenschaftsbereich hoch spezialisierte Kenntnisse voraussetzen, sind systematische Verfahren notwendig, um den Entscheidungsträgern solche Informationen zur Verfügung zu stellen. Vergleichende Leistungsbeurteilungen sind deshalb wesentlicher Bestandteil eines Reformprozesses, der die Autonomie der wissenschaftlichen Einrichtungen stärkt und einen Wandel von der staatlichen Detailsteuerung zu einer Globalsteuerung mit wettbewerblichen Elementen beinhaltet.

Mittelfristig sollten vergleichende Bewertungen alle wichtigen Leistungsdimensionen der Hochschulen und der außeruniversitären Forschungseinrichtungen erfassen, wobei die unterschiedlichen Nutzer der Bewertungsergebnisse zu berücksichtigen sind. Aus methodischen Gründen – unter anderem wegen der unterschiedlichen fachlichen Gliederung von Forschung und Lehre in den Universitäten, die verschiedene Objektdefinitionen erforderlich machen – ist eine gleichzeitige Bewertung aller Leistungen, insbesondere von Forschung und Lehre, in einem einheitlichen Verfahren jedoch nicht zu empfehlen.

Gegenwärtig sind nach Auffassung des Wissenschaftsrates im Bereich der Forschung fundiertere und detailliertere vergleichende Informationen über die Qualität der Leistungen notwendig, damit der Wettbewerb der Universitäten untereinander und mit außeruniversitären Forschungseinrichtungen transparenter wird und zu einer nachhaltigen Leistungssteigerung führen kann. Er empfiehlt deshalb die Etablierung eines Forschungsratings in Deutschland (B.III, S. 42 ff.). Dieses Verfahren sollte aber auch Bewertungen der Förderung des wissenschaftlichen Nachwuchses und des Wissenstransfers umfassen, da diese eng mit der Forschungsleistung zusammenhängen.

Aus Gründen der Qualitätssicherung ist die Existenz alternativer Ranking- bzw. Rating-Initiativen zu begrüßen. Die Akzeptanz bei den Adressaten wie auch bei den bewerteten Einrichtungen hängt entscheidend von der Qualität der Verfahren ab. Durch einen Wettbewerb unter verschiedenen Initiativen kann ein iterativer Lernprozess angestoßen werden. Dabei sollte dafür Sorge getragen werden, die Einrichtungen nicht durch mehrfache, unkoordinierte Datenerhebungen unnötig zu belasten. Deshalb ist es wünschenswert, alle für Zwecke des Leistungsvergleichs erhobenen Daten in geeigneter, ggf. anonymisierter Form für alternative Bewertungen und Re-Analysen zur Verfügung zu stellen. Die Daten sollten auch für wissenschaftliche Untersuchungen zugänglich gemacht werden.

Der mit der Einführung von vergleichenden Bewertungsverfahren verbundene Lernprozess muss nicht bei Null beginnen (vgl. Teil A). Deshalb und auf Grundlage nationaler und internationaler Erfahrungen gibt der Wissenschaftsrat im folgenden Empfehlungen zu vergleichenden Bewertungsverfahren im Wissenschaftssystem ab, die er als Standards guter Rankingpraxis verstanden wissen möchte (B.II). Die verantwortlichen Akteure in Politik und Wissenschaft sollten die Einhaltung dieser Standards aktiv unterstützen und Ranking-Initiativen, die ihnen nicht genügen, kritisieren, um in der Öffentlichkeit ein Bewusstsein für die Qualität von Rankings zu erzeugen.

B.II. Empfehlungen zu vergleichenden Bewertungsverfahren im Wissenschaftssystem

Zielsetzung angeben

Angesichts der Heterogenität möglicher Bewertungsverfahren bedürfen Rankings und verwandte Prozesse⁴⁴⁾ einer Reihe von expliziten Festlegungen, damit die Ergebnisse nicht unsachgemäß interpretiert werden. Zu einer klaren, nutzerfreundlichen Information gehört die Benennung von Zielen und Adressaten. Nutzer, die ein Ranking – in der Regel vor allem die zugrunde liegenden Daten – für andere als die intendierten Zwecke verwenden, sollten auf die damit verbundenen Probleme hingewiesen werden.

Rankings können eine Reihe von Zielen haben:

- Unterstützung der wissenschaftlichen Einrichtungen bei der strategischen Ausrichtung, Profilbildung und Qualitätssicherung;
- Unterstützung der Wissenschaftler bei der Selbsteinschätzung und strategischen Planung;
- Förderung des institutionellen Wettbewerbs;
- Hilfestellung für Studienanfänger, Studierende, Doktoranden, Nachwuchs- oder Gastwissenschaftler oder für Bewerber um eine Professur bei der Wahl einer aufnehmenden Einrichtung für den nächsten Abschnitt ihrer Karriere;
- Verbesserung des Marktes für wissenschaftliche Dienstleistungen und Weiterbildungsangebote.

Je nach Zielsetzung unterscheiden sich auch die Adressaten von Rankings im Wissenschaftssystem:

- Entscheidungsträger in den wissenschaftlichen Einrichtungen;
- Entscheidungsträger in den Ministerien und bei den Fördereinrichtungen, einschließlich der Stiftungen;
- Mitglieder der Universitäten und außeruniversitären Forschungseinrichtungen;
- Studienanfänger, Studierende, Doktoranden, Nachwuchs- oder Gastwissenschaftler sowie Stellenbewerber aus dem In- und Ausland;

⁴⁴⁾ Im Abschnitt B.II. wird für „Rankings und verwandte Prozesse“ oder für „vergleichende Bewertungsverfahren“ aus Gründen der Kürze auch kurz „Rankings“ verwendet. Alle allgemeinen Empfehlungen beziehen sich, soweit nicht explizit von „Rankings im engeren Sinne“ oder aber von „Rankings“ die Rede ist, auf alle derartigen Verfahren.

- Unternehmen und Behörden als Abnehmer von Forschungsleistungen.

Von den Zielen und Adressaten hängen die zu erfassenden Leistungsdimensionen, die Leistungskriterien und die Rankingobjekte ab (s. u.). Auch die Organisation einer Ranking-Initiative sollte sich in vielen Aspekten, etwa hinsichtlich der Veröffentlichungswege und der Wiederholungsintervalle, danach richten, wie die Adressaten optimal unterstützt werden können, ohne die betrachteten Institutionen über Gebühr zu belasten.

Entscheidungsrelevante Objektdefinitionen wählen

Für jedes Ranking muss die Objektmenge durch Benennung des Wissenschaftsbereichs, der betrachteten Institutionen sowie der Aggregationsebene der Bewertung festgelegt werden. Die Auswahl hängt von den Zielen und Adressaten ab.

Hinsichtlich der Institutionentypen ist der Wissenschaftsrat der Auffassung, dass es nicht angemessen wäre, Universitäten, Fachhochschulen und außeruniversitäre Einrichtungen direkt miteinander zu vergleichen, da die Aufgaben und Voraussetzungen dieser verschiedenen Institutionentypen zu unterschiedlich sind. Soweit sie analoge Leistungen erbringen, sollten diese aber nach einheitlichen Kriterien bewertet werden.

Die Wahl der Aggregationsebene ist von den Zielen und Adressaten abhängig und muss außerdem die Verfügbarkeit von Daten respektive den Aufwand ihrer Erhebung berücksichtigen.

Rankings, die Hochschulen zur Gänze und damit fachlich und funktional undifferenziert erfassen, sind angesichts der unterschiedlichen institutionellen Profile problematisch. Ohne einen klaren Handlungsbezug, der wiederum eine benennbare Adressatengruppe voraussetzt, können gerade solche „Hitlisten“ ganzer Universitäten leicht zu einer Fehlsteuerung führen. Fachlich und funktional differenzierte, mehrdimensionale Rankings können demgegenüber den Institutionen helfen, ihr Profil zu schärfen. Voraussetzung ist, dass ihre Profilierungsbemühungen nicht durch eine externe Detailsteuerung konterkariert werden.

Der Wissenschaftsrat rät davon ab, Einzelpersonen öffentlich und institutionenübergreifend zu ranken. Die so erzeugten Ranglisten werden von vielen Nutzern als

vollständige und akkurate Abbildung der wissenschaftlichen Leistungen der Betroffenen gelesen, selbst wenn vor einer solchen Nutzung ausdrücklich gewarnt wird. Dadurch können die Betroffenen, soweit sie nicht zur Spitzengruppe gehören, demotiviert und in ihren Persönlichkeitsrechten verletzt werden. Dieser Schaden wird durch die mögliche Anreizwirkung von öffentlichen, personenbezogenen Rankings nicht aufgewogen. Vergleichende Informationen über die wissenschaftlichen Leistungen von Einzelpersonen sollten deshalb in der Regel nur intern – etwa im Rahmen der leistungsbezogenen Mittelvergabe – verwendet werden. Davon unberührt bleibt selbstverständlich die Möglichkeit, besonders herausragende Leistungen öffentlich zu prämiieren und damit Anreize zu setzen sowie klare Signale beispielsweise für (prospektive) Doktoranden zu geben.

Für fachlich und funktional differenzierte, institutionelle Bewertungen auf mittlerer Aggregationsebene sind hinsichtlich der Objektdefinition zwei Teilfragen zu unterscheiden: Zum einen muss eine Klassifikation wissenschaftlicher Fächer, Fachgebiete oder Disziplinen festgelegt werden. Soweit dies möglich ist, sollte dabei eine der für statistische oder administrative Zwecke etablierten Klassifikationen – die Fachgebiets- und Fächerstruktur der DFG oder die Systematik des Statistischen Bundesamts – zugrunde gelegt werden. Zum anderen ist zu entscheiden, ob die empirisch vorfindliche organisatorische Gliederung der Hochschulen und außeruniversitären Forschungseinrichtungen in Institute, Abteilungen, Fachbereiche, Fakultäten, Studiengänge oder andere Organisationseinheiten abgebildet werden soll oder ob die Wissenschaftler der untersuchten Einrichtungen individuell je einem Fach/Fachgebiet zugeordnet werden sollen.

Die Bewertung empirisch vorfindlicher Organisationseinheiten hat den Vorzug, dass diese Objekte bereits statistisch erfasst und unmittelbar entscheidungsrelevant sind. Ein Nachteil ist jedoch, dass aus historischen Gründen jede Hochschule und Forschungseinrichtung anders gegliedert ist, was die Vergleichbarkeit der so erhobenen Daten beeinträchtigt. Zudem sind die empirischen Organisationseinheiten häufig fachlich heterogen, so dass ein Vergleich nach den Kriterien der an einer Fakultät oder einem Institut dominierenden Disziplin Wissenschaftlern aus anderen Gebieten, die dort ebenfalls institutionell beheimatet sind, nicht gerecht wird. Diese Art von Verzerrung kann durch eine normative Zuordnung von Wissenschaftlern zu Fächern

verhindert werden, allerdings um den Preis, dass die innerhalb der Hochschule bereits vorhandenen Daten, die für ein Ranking als Input-Daten relevant sein könnten, nur eingeschränkt genutzt werden können.

Angesichts dieser Abwägungsfragen kann keine allgemeingültige Empfehlung abgegeben werden. In der Regel sind empirische Organisationseinheiten an Hochschulen eher an der Lehre ausgerichtet und sollten deshalb bei lehrbezogenen Rankings zugrunde gelegt werden, während Forschungskommunikation und Wissenstransfer disziplinären bzw. branchenspezifischen Strukturen folgen und deshalb eher in einer organisationsübergreifend geltenden, normativen Klassifikation adäquat zu erfassen sind (vgl. den Vorschlag für ein Forschungsrating in B.III, S. 42 ff.).

Jede Ranking-Studie sollte ihre Objekte explizit definieren. Im Rahmen einer Studie ist für alle Leistungsdimensionen und -kriterien eine einheitliche Objektdefinition zu verwenden.

Leistungsdimensionen getrennt bewerten

Die Hochschulen erbringen verschiedene wissenschaftliche und wissenschaftsbaasierte Leistungen, unter denen Forschung, Lehre, Förderung des wissenschaftlichen Nachwuchses, Wissenstransfer, Weiterbildung und – in den Universitätsklinika – Krankenversorgung nur die wichtigsten sind. Auch die außeruniversitären Forschungseinrichtungen sind nicht auf eine einzige, klar umrissene wissenschaftliche Funktion beschränkt. Hinzu kommen weitere, nicht wissenschaftsspezifische Leistungen wie die Förderung der Chancengleichheit von Frauen und Männern oder die Integration von Menschen mit Migrationshintergrund.

Die verschiedenen Leistungen, die teils aneinander gekoppelt sind, teils im Zuge der Differenzierung des Wissenschaftssystems an einzelnen Institutionen mehr oder weniger stark ausdifferenziert werden, sollten unabhängig voneinander bewertet werden. Es ist nicht sinnvoll, die Institutionen einfach in eine Rangfolge von „besser“ und „schlechter“ bringen zu wollen. Wo ein Ranking mehrere Leistungen umfasst, sollten diese als unterschiedliche Leistungsdimensionen klar getrennt und in separaten Ranglisten bewertet werden. Zusammenhänge zwischen verschiedenen Leistungsdimensionen können gegebenenfalls durch eine mehrdimensionale Darstellung anschaulich gemacht werden.

Leistungskriterien offen legen

Innerhalb einer Leistungsdimension kann die Bewertung nach unterschiedlichen Kriterien erfolgen. So kann beispielsweise die Forschungsleistung hinsichtlich der in der Spitze erzielten Qualität, hinsichtlich der (qualitätsgewichteten) Quantität oder hinsichtlich der Effizienz der Leistungserbringung (Quantität pro Input) beurteilt werden.

Welche Kriterien verwendet werden, hängt von der Zielsetzung und den Interessen des jeweiligen Adressaten ab. Während es für einen Nachwuchswissenschaftler wichtig sein mag, unter guten Bedingungen mit einem oder wenigen besonders hoch reputierten Kollegen zusammenarbeiten zu können, ist es für einen jüngeren Studierenden entscheidend, einen Lehrkörper mit durchgängig gutem Niveau vorzufinden; ein Wissenschaftspolitiker mag vor allem an der nationalen und internationalen Reputation eines Standorts interessiert sein, während unter fiskalischen Gesichtspunkten größere Bedeutung haben kann, wo Leistungen besonders effizient erbracht werden. Unterschiedliche (implizite) Kriterien dürften beispielsweise für die typischerweise divergierenden Bewertungen von Hochschulen durch Professoren und Studierende mit verantwortlich sein; auch Qualität, Effektivität und Effizienz koinzidieren nicht zwangsläufig. Bewertungen nach widersprüchlichen Kriterien zu addieren oder zu mitteln, ergäbe sinnlose Resultate. In dem Maße, in dem eine mehrdimensionale Darstellung es erlaubt, auch unterschiedliche Kriterien unabhängig voneinander darzustellen, kann ein größerer Adressatenkreis bedient werden.

Gelegentlich werden auch indirekte Kriterien, z. B. zur Ausstattung, herangezogen. Derartige Informationen über infrastrukturelle Voraussetzungen begründen die Erwartung, dass eine gute Leistung erbracht werden könnte, ohne selbst Leistungen zu belegen. Dennoch sind solche Informationen für manche Nutzer – etwa für Nachwuchswissenschaftler oder Bewerber um eine Professur – wertvoll.

Die Kriterien für die Bewertung unterscheiden sich ferner auch zwischen den wissenschaftlichen Fachgebieten, weshalb ein Vergleich über verschiedene Fachgebiete hinweg oder von fachlich heterogenen Einheiten zu Scheinergebnissen führt. Dies gilt nicht nur in der Forschung, sondern auch in der Lehre, wo die Erwartungen an Absolventen in unterschiedlichen Fachgebieten ebenfalls stark voneinander abweichen. Der Wissenschaftsrat hält deshalb fachlich differenzierte Rankings für geboten.

Eine Bewertung der Qualität akademischer Leistungen erfordert die Einbeziehung von Fachkollegen, die die allgemeinen Leistungskriterien eines Rankings fachspezifisch konkretisieren und operationalisieren, da Qualität und Leistung von Fach zu Fach unterschiedlich definiert sind.

Indikatoren nach Leistungskriterien auswählen

Jede Bewertung wissenschaftlicher Leistungen muss auf geeigneten quantitativen und qualitativen Indikatoren fußen. Welche Indikatoren verwendet werden, hängt dabei von der Definition der Leistungskriterien ab und ist damit zum Teil nur fachspezifisch zu klären. Es entspricht im Verständnis des Wissenschaftsrates nicht guter Rankingpraxis, lediglich diejenigen Indikatoren zusammenzustellen, die auf Basis leicht verfügbarer Daten berechnet werden können. Auf Indikatoren, die nicht eindeutig einem Leistungskriterium zugeordnet werden können, sollte verzichtet werden.

Soweit möglich, sollten Indikatoren gewählt werden, die anreizkompatibel und nicht manipulierbar sind. Von öffentlichen Bewertungen gehen immer Verhaltenswirkungen aus, selbst wenn sie nicht direkt an Ressourcenverteilungen gekoppelt sind. Dies sollte schon bei der Definition von Leistungskriterien und noch mehr bei der Konstruktion von Indikatorensystemen berücksichtigt werden. Wo bekannt ist, dass bestimmte Indikatoren ein einseitiges Optimierungsverhalten nahe legen, sollten die Indikatoren mit entsprechender Vorsicht genutzt oder so modifiziert bzw. kombiniert werden, dass dysfunktionale Anreizeffekte minimiert werden.

Grundsätzlich sollte jede Ranking-Initiative einen begleitenden Prozess vorsehen, in dem Indikatoren im Lichte der Erfahrungen, die mit ihnen gemacht wurden, überprüft und gegebenenfalls vor einer erneuten Anwendung revidiert werden.

Hohe Datenqualität und optimale Auswertbarkeit anstreben

An die Datenqualität von Rankings sind insbesondere auch wegen der erwarteten Steuerungseffekte höchste Ansprüche zu stellen. Grundsätzlich sind die Erhebungsmethoden und Daten in einer wissenschaftlichen Standards genügenden Weise zu veröffentlichen. Um Kontrolle, Nachvollzug und Zweitverwertungen zu ermöglichen, sollten auch die Mikrodaten unter Einhaltung datenschutzrechtlicher Standards für Re-Analysen verfügbar gemacht werden. So kann auch die Gesamtbelastung der

Hochschulen und außeruniversitären Forschungseinrichtungen durch Datenerhebungen in Grenzen gehalten werden.

Gleichzeitig sollte berücksichtigt werden, dass Rankings nicht nur von Experten rezipiert werden. Es gehört deshalb zur guten Rankingpraxis, die Erhebungsmethoden in einer auch für Laien verständlichen Weise darzustellen und dabei auf Faktoren hinzuweisen, die für ein Urteil über die Verlässlichkeit der Daten und die Bewertung der Indikatoren entscheidend sind.

Adäquate Ergebnisdarstellung wählen

Eine Rangliste erzeugt beim Leser den Eindruck, dass zwischen den einzelnen Rangplätzen – nicht nur an der Spitze, sondern auch im mittleren und unteren Bereich – signifikante Unterschiede bestehen. Um den Adressaten die gewünschte Entscheidungsunterstützung zu geben, ist eine solche Differenzierung jedoch nur selten erforderlich. Deswegen ist die Bildung von Ranggruppen, die in sich alphabetisch sortiert sind, in der Regel vorzuziehen. Ist die Bewertung der Qualität eines akademischen Outputs gefordert und wird dazu ein Peer-Review-Verfahren gewählt, ist ein Rating mit prädefinierten Qualitätsstufen in der Regel sinnvoller als ein eigentliches Ranking. Falls ein Ranking im Sinne einer ordinalen Rangliste für bestimmte Ziele für notwendig erachtet wird, sollte in jedem Fall geprüft werden, ob Unterschiede in einzelnen Indikatoren, die für die Platzierung in der Rangliste entscheidend sind, überhaupt signifikant sind.

Unabhängige Trägerschaft und Bewertung sicherstellen

Der Träger eines Rankings sollte in erster Linie den Adressaten verantwortlich, von den bewerteten Einrichtungen hingegen unabhängig sein. Zugleich muss er das Vertrauen beider Seiten genießen.

Für flächendeckende Rankings in Deutschland empfiehlt der Wissenschaftsrat eine gewichtige Beteiligung ausländischer Wissenschaftler, um adäquate, international gültige Bewertungsmaßstäbe zu sichern.

Wettbewerb und Autonomie wahren

Rankings sind geeignet, Leistungsdifferenzierung im Wissenschaftssystem zu dokumentieren und dadurch Wettbewerb und Profilbildung zu fördern. Von den Hochschu-

len und außeruniversitären Forschungseinrichtungen können sie als Informationsgrundlage für Entwicklungsplanungen verwendet und damit als Instrument der Selbststeuerung genutzt werden.

Eine Verteilung staatlicher Ressourcen nach einer Formel, die wesentlich auf Rankingergebnissen beruht (vgl. dazu die RAE im Vereinigten Königreich, S. 12 ff.), wäre zwar transparent, würde jedoch, da Rankings nur bereits erbrachte Leistungen bewerten, zu einer rückwärtsgewandten anstelle einer zukunftsorientierten Steuerung führen. Rankings sind eine wichtige Informationsgrundlage für eine leistungsorientierte Steuerung, sollten aber so verwendet werden, dass die Autonomie der wissenschaftlichen Einrichtungen gestärkt wird und Spielräume für die Ausgestaltung individueller Profile bleiben. Der Wissenschaftsrat spricht sich deswegen gegen eine vollständig oder überwiegend auf Rankingergebnissen basierende Steuerung des Wissenschafts- oder des Hochschulsystems und für ihre Nutzung durch die Entscheidungsträger in Verbindung mit anderen Verfahren der Qualitätssicherung und strategischen Planung aus.

B.III. Empfehlungen zu einem Forschungsrating

Die Profilierung der Universitäten und ihre Leistungsdifferenzierung, die zu einem verstärkten Wettbewerb und einer Orientierung an internationalen Maßstäben führen soll, erfolgt heute vor allem in der Forschung. Die Forschungslandschaft in Deutschland ist wesentlich auch von den außeruniversitären Forschungseinrichtungen geprägt. Universitäten und außeruniversitäre Forschungseinrichtungen ergänzen sich in ihren Aufgaben und sollten dies bei ihren strategischen Entscheidungen im Auge behalten. Um sie darin zu unterstützen und den Wettbewerb in der Forschung effektiver und effizienter zu machen, bedarf es zuverlässiger vergleichender Qualitätsinformationen. Deshalb empfiehlt der Wissenschaftsrat eine deutschlandweite, vergleichende Bewertung von Forschungsleistungen, die eine Standortbestimmung nach internationalen Maßstäben leistet. Damit die Universitäten bei der Profilierung unterstützt werden, ist eine fachgebietsspezifische und mehrdimensionale Bewertung nach verschiedenen Kriterien angezeigt.

Angesichts der internationalen Erfahrungen (vgl. A.II und das Resümee in A.IV, S. 29ff.) schließt der Wissenschaftsrat eine ausschließlich auf quantitativen Indikato-

ren basierende Forschungsbewertung ebenso aus wie eine reine Reputationsmessung. Für einen Vergleich der Qualität von Forschungsleistungen ist eine fachgebietsspezifische Bewertung durch Fachwissenschaftler („Peer-Review“) auf Basis einheitlicher Daten und quantitativer Indikatoren („Informed Peer Review“) nach einer vorgegebenen Bewertungsskala angezeigt. In diesem Sinne handelt es sich bei dem vom Wissenschaftsrat empfohlenen Verfahren um ein Forschungsrating.

Das vorliegende Gesamtkonzept eines Forschungsratings hat folgende Komponenten: eine Steuerungsgruppe (vgl. III.7, S. 52 ff.) setzt Bewertungsgruppen (III.5, 50 ff.) ein, die die Leistungen der bewerteten Einrichtungen in ihrem jeweiligen Fachgebiet in den Dimensionen Forschung, wissenschaftlicher Nachwuchs und Wissenstransfer nach allgemein vorgegebenen, von jeder Gruppe fachgebietsspezifisch zu konkretisierenden Kriterien (III.2, S. 45 ff.) bewerten. Zu diesem Zweck werden ihnen fachgebietsspezifische Forschungsprofile der Einrichtungen (III.3, S. 48 ff.) und quantitative Indikatoren (III.4, S. 49 f.) vorgelegt. Die Ergebnisse werden als Ranggruppen von Universitäten und außeruniversitären Forschungseinrichtungen in den einzelnen Fachgebieten dargestellt (III.6, S. 52 f.). Daneben sollen auch Qualitätsprofile der einzelnen Einrichtungen gebildet werden. Auf mittlere Sicht soll in Kooperation mit anderen Ländern, in denen vergleichbare Verfahren etabliert sind, die Möglichkeit eines internationalen Benchmarkings erkundet werden (III.9, S. 56 f.).

III.1. Ziele, Adressaten, Objekte

Das vom Wissenschaftsrat empfohlene Forschungsrating hat zum Ziel, Universitäten und außeruniversitäre Forschungseinrichtungen im Rahmen ihrer jeweiligen Mission und in Verbindung mit anderen Verfahren der Qualitätssicherung und strategischen Planung in ihren strategischen Entscheidungen und bei der Qualitätssicherung in der Forschung zu unterstützen und den Qualitätswettbewerb zu fördern. Dazu sollen vergleichende Informationen über ihre Forschungsleistungen bereitgestellt und nach internationalen Maßstäben bewertet werden. Wichtige, mit der Forschungsleistung eng verbundene Aspekte sind die Förderung des wissenschaftlichen Nachwuchses und der Wissenstransfer. Sie sind für den Beitrag der jeweiligen Einrichtung zum Gedeihen der Wissenschaft in Deutschland entscheidend und sollen deshalb ebenfalls bewertet werden, um den Wettbewerb auch in diesen Dimensionen anzuregen.

Adressaten des Forschungsratings sind Entscheidungsträger in den Universitäten – sowohl in den Universitätsleitungen als auch auf Fachbereichs- und Institutsebene – und in den außeruniversitären Forschungseinrichtungen sowie in den Ministerien. Durch das Zusammenwirken dieser Ebenen soll der Wettbewerb in der Forschung zu einer dauerhaften Steigerung des Qualitätsniveaus und zur Herausbildung von Spitzenleistungen führen. Die durch das Forschungsrating erzielte Transparenz ist dafür eine wichtige Rahmenbedingung, indem sie hilft, lokale Maßstäbe zu vermeiden.

Neben den Entscheidungsträgern in den Organisationen wird das Forschungsrating auch für Studierende, Doktoranden und Nachwuchswissenschaftler von Interesse sein. Ähnliches gilt für Gastwissenschaftler und Bewerber um eine Professur, für die allerdings andere, individuellere Wege der Informationsbeschaffung größeres Gewicht haben dürften.

Gegenstand des Forschungsratings sind Forschungsaktivitäten der Universitäten sowie der außeruniversitären Forschungseinrichtungen, die als eigene Kategorie getrennt aufgeführt werden sollen. Serviceeinrichtungen der Forschung werden im Rahmen des Forschungsratings nicht bewertet.

Ein Forschungsrating muss fachlich differenziert sein, da in den einzelnen Fächern und Fachgebieten unterschiedliche Qualitätskriterien gelten. Der Wissenschaftsrat empfiehlt der Steuerungsgruppe (vgl. III.7, S. 52 f.), die die endgültige Taxonomie der Fachgebiete in Konsultation mit Fachvertretern festlegen sollte, dabei von der aktuellen Taxonomie der Fachgebiete, Fachkollegien und Fächer der DFG auszugehen. Dadurch wird auch eine Mehrfachverwertung der Daten möglich. Im Einzelfall kann es sinnvoll sein, für Zwecke des Ratings Fächergruppen zusammenzufassen oder entlang der Grenzen der Fächer im Sinne der DFG-Klassifikation weiter zu unterteilen. Die endgültige Taxonomie sollte nicht mehr als 50 Fachgebiete umfassen.

In ihrer Gesamtheit ergeben die Ratings für eine Einrichtung in den verschiedenen Fachgebieten ein Qualitätsprofil ihrer Forschungsleistungen. Der Wissenschaftsrat sieht die Möglichkeit, auf Basis solcher Qualitätsprofile zu einer Rangliste von Universitäten und außeruniversitären Forschungseinrichtungen in Deutschland zu kommen. Dies setzt allerdings eine Gewichtung von Fachgebieten und Leistungsdimensionen voraus, die von den Interessen des jeweiligen Nutzers abhängt, weshalb der

Wissenschaftsrat keine allgemeine Formel für eine solche Gewichtung empfiehlt. Wenn auf der Grundlage des Forschungsratings eine institutionenbezogene Rangliste in dem jeweiligen Fach gebildet werden soll, sollten für die Interpretation der Rangliste Informationen über die Mission, Forschungsstrategie und Organisationsstruktur der ganzen Einrichtung sowie höher aggregierte Daten, bspw. über den institutionellen Haushalt, hinzugezogen werden.

III.2. Allgemeine Leistungsdimensionen und Rahmenkriterien

Das angestrebte Forschungsrating soll Bewertungen der Universitäten und außeruniversitären Forschungseinrichtungen in den drei Leistungsdimensionen Forschung, Nachwuchsförderung und Wissenstransfer nach einer Reihe von Kriterien umfassen, die im Folgenden allgemein beschrieben werden. Die Konkretisierung und Operationalisierung dieser Rahmenkriterien und die Auswahl von Kriterien ist fachspezifisch und von der Verfügbarkeit zuverlässiger Daten abhängig. Sie erfolgt durch die für die einzelnen Fachgebiete einzusetzenden Bewertungsgruppen (III.5). Grundlage sind Forschungsprofile, die von den Organisationen eingereicht werden (III.3), und quantitative Indikatoren (III.4).

Leistungsdimension Forschung

Zentrales Kriterium in der Dimension Forschung ist die Qualität. Die weiteren Kriterien Effektivität und Effizienz hängen mit der Qualität zusammen; ihre getrennte Bewertung ermöglicht es jedoch den Bewertungsgruppen, ein dem Einzelfall angemesseneres, differenziertes Urteil abzugeben.

1. Kriterium: Qualität – Dieses Kriterium umfasst die Aktualität und Relevanz der Fragestellungen für das Forschungsgebiet, Neuheit und Originalität der Forschungsergebnisse sowie die Eignung und Verlässlichkeit der Methoden. Daten über die fachliche Resonanz der Forschungsergebnisse (z. B. normierte relative Zitationsindikatoren) können als Indikatoren genutzt werden.

2. Kriterium: Effektivität – Hier sollte bewertet werden, welchen Beitrag die Einrichtung zur Entwicklung des Fachgebiets gemessen an internationalen Standards leistet. Geeignete Indikatoren sollten ein Maß für die qualitätsgewichtete Menge der Forschungsprodukte darstellen.

3. Kriterium: Effizienz – Die qualitätsgewichtete Menge der Forschungsprodukte im Sinne der Effektivität ist hier in Relation zum Aufwand zu bewerten. Für die Erfassung des Aufwands können z. B. der Personaleinsatz oder die eingesetzten Mittel berücksichtigt werden.

Leistungsdimension Nachwuchsförderung

4. Kriterium: Prozesse der Nachwuchsförderung – Hier sollten die Maßnahmen der Einrichtung zur Nachwuchsförderung bewertet werden. Indikatoren sind u. a. die Einrichtung strukturierter Promotionsprogramme, die durchschnittlichen Promotionszeiten (Median) und Nachwuchswissenschaftler in wissenschaftlich selbständigen Positionen.

5. Kriterium: Erfolg der Nachwuchsförderung – Bei diesem Kriterium geht es um eine qualitätsgewichtete Bewertung des Beitrags, den die Einrichtung im Bewertungszeitraum zur Förderung des wissenschaftlichen Nachwuchses im Fachgebiet geleistet hat. Der wissenschaftliche Erfolg ihrer Absolventen kann etwa an ihren Publikationen (bzw. deren fachlicher Resonanz) gemessen werden. Ein breiterer Indikator für den Ausbildungserfolg ist, wie vielen Postgraduierten es gelingt, eine adäquate berufliche Stellung zu erreichen. Welches Spektrum an beruflichen Möglichkeiten als adäquat angesehen wird, ist von Fachgebiet zu Fachgebiet verschieden. Neben akademischen Karrieren sind entsprechend auch Fach- und Führungspositionen in der Wirtschaft, im Kultur- und Bildungsbereich sowie in Politik und Verwaltung positiv zu werten. Eine Bewertung kann allerdings erst mit großer zeitlicher Verzögerung erfolgen.

Derzeit können nur wenige Universitäten in Deutschland über den Erfolg ihrer Absolventen fundierte Angaben machen, so dass Befragungsergebnisse im ersten Durchgang nur mit Vorbehalt bewertet werden sollten. Der Wissenschaftsrat hält es jedoch für dringend erforderlich, den Informationsstand über den Verbleib von Nachwuchswissenschaftlern erheblich zu verbessern, sowohl um die Qualität der Nachwuchsförderung der einzelnen Einrichtungen zu bewerten, als auch um den Ausbildungs- und Arbeitsmarkt für Nachwuchswissenschaftler transparenter zu machen. Er spricht

sich deshalb dafür aus, Daten hierzu zu erheben.⁴⁵⁾ Er erhofft sich davon einen starken Anreiz für die Universitäten, durch eine verbesserte Alumnibetreuung und durch Absolventenstudien den Erfolg ihrer Nachwuchsförderung zu belegen.

Analog sollten auch die außeruniversitären Forschungseinrichtungen den Karriereverlauf der dort betreuten Doktoranden und von Nachwuchswissenschaftlern dokumentieren.

*Leistungsdimension Wissenstransfer*⁴⁶⁾

6. Kriterium: Relevanz – Hier ist zu bewerten, ob die Forschungsergebnisse über das engere Fachgebiet hinaus für den wissenschaftlichen Fortschritt auch in anderen Disziplinen relevant sind.

7. Kriterium: Wirtschaftliche Umsetzung – Die Umsetzung von relevanten Forschungsergebnissen in neue Produkte oder Dienstleistungen ist ein wichtiges Erfolgskriterium vor allem für praxisnahe Gebiete.

8. Kriterium: Fort- und Weiterbildung – Angebote der Fort- und Weiterbildung können ein wichtiges Mittel für die Universitäten sein, forschungsnahes Wissen in der Gesellschaft zur Geltung zu bringen. Neben der Qualität und Quantität der Angebote sollte auch ihr Modellcharakter beurteilt werden.

9. Kriterium: Forschungsbasierte Beratung, Wissenschaftskommunikation – In vielen Fachgebieten sind forschungsbasierte Beratungsdienstleistungen für Unternehmen und Behörden zentrales Mittel des Wissenstransfers. Auch die Vermittlung wissenschaftlicher Methoden und Erkenntnisse an die Öffentlichkeit („Wissenschaft im Dialog“, Ausstellungen etc.) ist eine wichtige Transferleistung von Universitäten und außeruniversitären Forschungseinrichtungen.

⁴⁵⁾ Die Bewertungsgruppen sollten auch vorliegende Absolventenstudien nutzen, die nicht von den Universitäten selbst, sondern von Hochschulforschern erstellt worden sind. Ein Vergleich verschiedener Standorte ist allerdings auf Basis dieser Studien in der Regel nicht möglich. Auch in der Europäischen Hochschulabsolventenstudie CHEERS mit 40.000 befragten Absolventen ist die Hochschule, an der der Abschluss erworben wurde, keine untersuchte Variable (vgl. http://www.uni-kassel.de/wz1/TSEREGS/metho_e.htm und die dort verzeichnete Literatur).

⁴⁶⁾ Die Interpretation der Rahmenkriterien und die Verfügbarkeit von Daten, die der Bewertung zugrunde gelegt werden können, sind in dieser Leistungsdimension von Fach zu Fach unterschiedlich. Es sollte den Bewertungsgruppen freigestellt werden, unter den Rahmenkriterien 6.-9. eine fachlich adäquate Auswahl zu treffen bzw. sie entsprechend zu ergänzen.

III.3. Forschungsprofile der Einrichtungen

Grundlage der Bewertung sind Forschungsprofile der Einrichtungen mit qualitativen und quantitativen Daten sowie weitere, aus bereits vorliegenden Daten gebildete quantitative Indikatoren (s. B.III.4, S. 49). Jede Einrichtung reicht für jedes Fachgebiet, das an ihr vertreten ist, ein Forschungsprofil ein; in Abhängigkeit von ihrem Fächerspektrum kann also jede Einrichtung bis zu ca. 50 Forschungsprofile einreichen.⁴⁷⁾

Jedem Forschungsprofil soll eine kurze Darstellung zur Rolle des Fachgebiets in der Gesamtstrategie der Einrichtung („Mission“) und zu ihrer für dieses Fachgebiet spezifischen Strategie vorangestellt werden. Daran schließt sich eine standardisierte Dokumentation von Forschungsaktivitäten in dem betreffenden Gebiet an (vgl. Anhang, S. 62 f.), die eine Bewertungsgrundlage für jedes Rahmenkriterium enthält. Diese Bewertungsgrundlage kann durch separat erhobene, quantitative Indikatoren (kursiv, vgl. III.4, S. 49) ergänzt werden.

Auf Wunsch der Bewertungsgruppen können über das standardisierte Forschungsprofil hinaus weitere Informationen erhoben werden, soweit dies mit vertretbarem Aufwand seitens der bewerteten Einrichtungen möglich ist. Außerdem sollte den Einrichtungen die Möglichkeit gegeben werden, zusätzliche Informationen zu übermitteln, die aus ihrer Sicht für die Bewertung unverzichtbar sind (Nennung besonderer, forschungsbasierter Aktivitäten, die im Standardforschungsprofil nicht erfasst sind).

Rahmenkriterium (III.2)	Bewertungsgrundlage im Forschungsprofil (Anhang)
<i>Dimension Forschung</i>	
1. Qualität	Forschungsprodukte (2.) Begutachtete Drittmittel (4.a) Wissenschaftliche Kooperationen (8.) <i>Ggf. relative Zitationsindikatoren, Anteil an viel zitierten Arbeiten</i>
2. Effektivität	Qualitätsgewichtete Publikationszahlen (3.a/b), <i>ggf. Vorträge auf international bedeutenden Tagungen</i>

⁴⁷⁾ In besonders begründeten Fällen kann eine Einrichtung auch mehr als ein Forschungsprofil in einem bestimmten Fachgebiet einreichen. Dies muss vorab mit der Bewertungsgruppe vereinbart werden.

	<i>Ggf. absolute Zitationsindikatoren, impactgewichtete Publikationszahlen</i>
3. Effizienz	Zähler: Qualitätsgewichtete Publikationszahlen (3.a/b) Nenner: Zahl der Wissenschaftler (1.b), Eingesetzte Mittel inkl. Drittmittel (4.)
<i>Dimension Nachwuchsförderung</i>	
4. Prozesse der Nachwuchsförderung	Strukturierte Promotionsprogramme, Median der Promotionsdauer (5.) extern finanzierte Stipendien (4.) Zahl der selbständigen Nachwuchsgruppen (6.)
5. Erfolg der Nachwuchsförderung	Verbleib von Promovierten, Postdoktoranden (5., 6.) Publikationen von Nachwuchswissenschaftlern (3.a)
<i>Dimension Wissenstransfer</i>	
6. Relevanz	Forschungsprodukte (2.) Kooperationen (8.)
7. Wirtschaftliche Umsetzung	Industriemittel (3.) Schutzrechte, Lizenzen (7.) Firmengründungen (7.) Kooperationen (8.)
8. Fort- und Weiterbildung	Beschreibung von Fort- und Weiterbildungsangeboten (9.)
9. Forschungsbasierte Beratung, Wissenschaftskommunikation	Beschreibung von forschungsbasierten Beratungsleistungen und Aktivitäten der Wissenschaftskommunikation (10.)

Tabelle: Zuordnung von Komponenten der Forschungsprofile zu Rahmenkriterien

III.4. Quantitative Indikatoren

In vielen Fachgebieten sind heute quantitative, vor allem bibliometrische Indikatoren zur Messung des Volumens und der Qualität von Forschungsleistungen anerkannt. Die Bewertungsgruppen sollten solche Indikatoren verwenden, sofern in ihrem Fachgebiet die Voraussetzungen dafür erfüllt sind. Dabei ist in der Leistungsdimension Forschung vor allem an folgende Indikatoren zu denken:

- Absolute Publikations- und Zitationszahlen als Maß für die Effektivität einer Einrichtung im betreffenden Forschungsgebiet;⁴⁸⁾
- mit Zitationen (oder alternativ/ergänzend einem Qualitätsfaktor für die jeweilige Zeitschrift) gewichtete Publikationszahlen als Indikator für die Effektivität;⁴⁹⁾
- ein fachspezifisch normiertes, relatives Zitationsmaß⁵⁰⁾ als Qualitätsindikator;
- ein dem Anteil an besonders viel zitierten Papieren (oberstes Perzentil) im Fachgebiet entsprechender Quotient als Qualitätsindikator.⁵¹⁾

Da es in dem Forschungsrating darum geht, das aktuelle Forschungspotential der Einrichtungen möglichst zeitnah zu vergleichen, sollten bibliometrische Daten nach der „Current potential“-Methode (auf Basis der im Forschungsprofil aufgeführten Namen der zum Stichtag beschäftigten Wissenschaftler) erhoben werden.

Die Durchführung von bibliometrischen Analysen in den Fachgebieten, für die geeignete Datenbanken vorliegen und ein Zitationsfenster von fünf Jahren ausreichend ist, ist im Ablauf- und Zeitplan und der Finanzierung des Forschungsratings angemessen zu berücksichtigen. Die bewerteten Einrichtungen sollten mit dem durchführenden Institut bei der Bereinigung der Autorenadressen in den verwendeten Datenbanken kooperieren.

III.5. Fachgebietsspezifische Bewertung durch Bewertungsgruppen

Die Bewertung erfolgt durch fachgebietsspezifische Bewertungsgruppen, denen bis zu sechs Fachwissenschaftler, davon in der Regel zwei aus dem Ausland, sowie zwei außerhalb des öffentlich geförderten Wissenschaftssektors (abhängig vom Fachgebiet: Wirtschaft, Kultur- und Bildungsbereich, Politik, Verwaltung) tätige Ex-

⁴⁸⁾ Empirische Zitationszahlen sollten für das Rating nur in Fachgebieten genutzt werden, in denen das Zeitfenster von maximal fünf Jahren (für die am längsten zurückliegenden Publikationen, entsprechend weniger für die jüngeren Publikationen) ausreicht, um die Wirkung von Veröffentlichungen verlässlich zu beurteilen.

⁴⁹⁾ Aufgrund der stark asymmetrischen Verteilung von Zitationen sagt der häufig zur Gewichtung verwendete Impactfaktor einer Zeitschrift wenig über die Qualität der einzelnen in ihr erscheinenden Publikationen aus. Auf der institutionellen Bewertungsebene des Forschungsratings mit der entsprechenden Stichprobengröße kann die Verwendung eines Qualitätsfaktors, der an den Impactfaktor angelehnt ist, jedoch sinnvoll sein.

⁵⁰⁾ Zitationen pro Publikation (=„relativer Zitationsindikator“) geteilt durch (=„normiert auf“) die *durchschnittliche* Anzahl von Zitationen pro Publikation *in dem Fachgebiet*. Ein Wert von 1 zeigt bei einem normierten Maß eine durchschnittliche, ein Wert über 1 eine überdurchschnittliche Rezeption der Publikationen der bewerteten Einrichtung an.

⁵¹⁾ Tijssen, Visser & van Leeuwen (2002); für einen Vergleich mit herkömmlichen bibliometrischen Indikatoren vgl. van Leeuwen et al. (2003). Bei dieser Art von Indikator sind reine Methodenpapiere niedriger zu gewichten.

perten und ein fachferner Beisitzer angehören. Größe und Zusammensetzung der einzelnen Gruppen hängen von der Breite des Fachgebiets ab und werden von der Steuerungsgruppe bestimmt.

Jeder fachferne Beisitzer nimmt an den Beratungen von vier bis fünf Bewertungsgruppen teil. Seine Aufgabe ist es, für einheitliche Bewertungsmaßstäbe zu sorgen.

Jede Bewertungsgruppe verständigt sich auf ihrer konstituierenden Sitzung, wie sie die Bewertungskriterien interpretieren, anhand welcher Indikatoren sie sie beurteilen und wie sie die Indikatoren gewichten will. Eine Erklärung darüber wird veröffentlicht und den zu bewertenden Einrichtungen zugesandt.

Die Bewertungen erfolgen auf Basis der von den Universitäten und außeruniversitären Forschungseinrichtungen einzureichenden Forschungsprofile für jedes Fachgebiet (III.3 und Anhang). Darüber hinaus sollten die Bewertungsgruppen bereits vorliegende oder mit vertretbarem Aufwand zu erhebende quantitative Indikatoren zur Stützung und Kontrolle ihres Urteils hinzuziehen, soweit dies den Gepflogenheiten in ihrem Fach entspricht (III.4). Die Bewertungsgruppen sollten die Möglichkeit haben, in schwierigen Fällen Gespräche mit Vertretern einzelner bewerteter Einrichtungen zu führen oder Begehungen vorzunehmen.

Die Bewertungsgruppen werden gebeten, die Forschungsprofile nach jedem der oben aufgeführten Kriterien auf einer siebenstufigen Skala zu bewerten. Für das Hauptkriterium „Forschungsqualität“ werden Eckwerte der Notenskala verbal definiert:

7	Mehr als die Hälfte der Forschungsaktivitäten gehören zur internationalen Spitze, alle weiteren Aktivitäten zur nationalen Spitze und sind international konkurrenzfähig.
5	Mehr als die Hälfte der Forschungsaktivitäten gehören zur nationalen Spitze und sind international konkurrenzfähig. Einzelne Aktivitäten können zur internationalen Spitze gehören.
3	Die Forschungsaktivitäten entsprechen zum überwiegenden Teil nationalen Qualitätsstandards. Einzelne Aktivitäten gehören zur nationalen Spitze und sind international konkurrenzfähig.
1	Die Forschungsaktivitäten entsprechen nicht oder nur zu einem geringen Anteil nationalen Qualitätsstandards.

Die zwischen den definierten Werten liegenden Ziffern dienen der feineren Abstimmung der Bewertung. Die Bewertungsgruppen werden aufgefordert, das Notenspektrum auszuschöpfen. Die Bestnote wird bei Einhaltung internationaler Exzellenzstandards nicht in jedem Fachgebiet vergeben werden.

III.6. Ergebnisdarstellung

Das Rating resultiert in multidimensionalen Bewertungen der Forschungsaktivitäten von Universitäten respektive der außeruniversitären Forschungseinrichtungen für jedes der betrachteten Fachgebiete. Die adäquate Veröffentlichungsform ist eine webbasierte, dynamische Veröffentlichung mit zwei verschiedenen Abfragemöglichkeiten.

Zum einen sollte es dem Nutzer möglich sein, Aktivitäten der Universitäten in einem Forschungsgebiet nach einem der neun vorgegebenen Bewertungskriterien zu sortieren. Die Ausgabe nach einem Bewertungskriterium sollte dabei in sieben Ranggruppen (entsprechend der Bewertungsskala) erfolgen, innerhalb derer die Universitäten alphabetisch sortiert sind. Analog können auch für die außeruniversitären Forschungseinrichtungen in einem Forschungsgebiet Ranggruppen nach geeigneten, ihren jeweiligen Aufgaben angemessenen Kriterien gebildet werden.

Zum anderen sollte es möglich sein, eine bestimmte Einrichtung (universitär oder außeruniversitär) auszuwählen und alle Bewertungen anzeigen zu lassen, die sich auf ihre gesamten Forschungsaktivitäten beziehen. Auf diese Weise entsteht eine strategisch nutzbare Darstellung des Forschungsportfolios der betreffenden Einrichtung.

Die von den Einrichtungen abgefragten Grunddaten zur Personalstruktur, zu Drittmitteln und zu Promotionen sollten ebenfalls publiziert werden.

III.7. Trägerschaft, Organisation und Finanzierung

Träger des Forschungsratings sollte eine wissenschaftsnahe Organisation sein, die das Vertrauen von Hochschulen und außeruniversitären Forschungseinrichtungen sowie von Bund und Ländern genießt, über organisatorische und methodische Kom-

petenz in der Forschungsbewertung verfügt und von den bewerteten Einrichtungen unabhängig ist.

Die Verfahrenshoheit über das Forschungsrating muss eine Steuerungsgruppe haben, die im Kern aus anerkannten Wissenschaftlern aus allen Wissenschaftsgebieten besteht. Ferner sollten die großen Wissenschaftsorganisationen in geeigneter Weise vertreten sein. Vertreter von Bund und Ländern können der Steuerungsgruppe als Gäste angehören.

Zentrale Aufgaben der Steuerungsgruppe sind:

- Festlegung der Taxonomie der Fachgebiete in Konsultation mit den Fachgesellschaften,
- Ernennung der Vorsitzenden der Bewertungsgruppen,
- Ernennung der Mitglieder der Bewertungsgruppen,
- Verabschiedung von Richtlinien für die Erstellung von Forschungsprofilen,
- Verabschiedung von Richtlinien für die Arbeit der Bewertungsgruppen,
- Sicherung einheitlicher, konsistenter Bewertungsmaßstäbe,
- Gesamtverantwortung für das Forschungsrating.

Die Steuerungsgruppe und die Bewertungsgruppen müssen von einer Geschäftsstelleneinheit unterstützt werden. Diese Einheit sollte administrativ beim Träger angesiedelt, inhaltlich jedoch allein der Steuerungsgruppe verantwortlich sein.

Die Erhebung der Forschungsprofile bei den Universitäten und außeruniversitären Forschungseinrichtungen sollte auf elektronischem Wege erfolgen. Die Hochschulleitungen sollten die für das Rating aufbereiteten Forschungsprofile noch einmal überprüfen können, bevor sie den Bewertungsgruppen vorgelegt werden. Soweit möglich und mit angemessenem Aufwand zu erreichen sollten quantitative Angaben anhand externer Quellen validiert werden. Die Ergebnisse der Erhebung sollten für Zweitauswertungen verfügbar gemacht und nutzerfreundlich aufbereitet werden. Sobald die erste Aufbauphase des geplanten Instituts für Forschungsinformation und Qualitätssicherung (IFQ)⁵² abgeschlossen ist, sollte geprüft werden, inwieweit das Institut

⁵² Die Deutsche Forschungsgemeinschaft hat dieses Institut als sogenannte Hilfseinrichtung der Forschung geschaffen, um Informationen über ihr Förderhandeln und weitere Daten über das Wissenschaftssystem so aufzubereiten, dass in einer ersten Phase ihr Förderhandeln verbessert, in einer zweiten Phase auch weitere Akteure im Wissenschaftssystem unterstützt werden können.

mit der Erhebung von Daten und der Auswertung bereits existierender Datenquellen (Drittmittelstatistiken der Geber, Bibliometrie) für das Rating beauftragt werden und an der Entwicklung quantitativer Indikatoren mitwirken kann.

Die Bewertung in einem Fachgebiet wird von der Einberufung der Bewertungsgruppe über die Operationalisierung der Kriterien und die Erhebung der Forschungsprofile bis zur Veröffentlichung der Bewertungen etwa 15 bis 18 Monate in Anspruch nehmen. Aus organisatorischen Gründen und wegen der Belastung der Universitäten und außeruniversitären Forschungseinrichtungen ist es kaum praktikabel, alle (ca. 50) Fachgebiete parallel zu bewerten. Da das Rating, anders als etwa die RAE im Vereinigten Königreich, nicht dazu dienen soll, die Grundfinanzierung der Universitäten zu steuern, ist eine gleichzeitige Bewertung aller Fachgebiete auch nicht zwingend erforderlich. Stattdessen sollte ein rollierendes System gewählt werden, bei dem jährlich eine bestimmte Anzahl von Forschungsgebieten bewertet wird.

Um die mit dem Forschungsrating beabsichtigten Lerneffekte zu erzielen, ist eine turnusmäßige Wiederholung des Prozesses anzustreben. Nach internationalen Erfahrungen ist dafür ein fünf- bis sechsjähriges Intervall sinnvoll. Wenn zwölf bis 15 Fachgebiete pro Jahr bewertet werden, wird die Gesamtheit der Wissenschaftsgebiete innerhalb von vier Jahren abgedeckt, so dass etwa ein Jahr Zeit bleibt, um die Ergebnisse zu resümieren, den Prozess zu evaluieren und gegebenenfalls notwendige Änderungen zu beschließen.

Auf Seiten der Universitäten und außeruniversitären Forschungseinrichtungen ist mit großem Klärungsbedarf sowohl hinsichtlich der Zielsetzung als auch der Umsetzung des Forschungsratings zu rechnen. Es bedarf deshalb einer umfassenden Kommunikationsstrategie, die Informationsveranstaltungen, eine Hilfefunktion auf der Website des Ratings sowie eine telefonische Beratung einschließen sollte. Zusätzlich empfiehlt der Wissenschaftsrat, im Zuge der Umsetzung des Forschungsratings Regionalkonferenzen abzuhalten, um die Universitäten und außeruniversitären Forschungseinrichtungen mit den Zielen und der Methodik vertraut zu machen. Dies ist dringend notwendig, um die Qualität der erhobenen Daten und die Akzeptanz des Ratings sicherzustellen.

Die direkten Kosten für die Durchführung des Forschungsratings setzen sich zusammen aus den Aufwandsentschädigungen und Reisekosten für die Gutachter, den Kosten für bibliometrische und andere Datenbankanalysen, den Kosten für die Erhebung der Forschungsprofile und den Personalkosten für die Projektleitung, für die Betreuung der Steuerungsgruppe und der Bewertungsgruppen sowie für die Öffentlichkeitsarbeit. Sie werden sich nach Schätzungen des Wissenschaftsrates auf mindestens 2,6 Mio. Euro pro Jahr belaufen.

Darüber hinaus ist mit erheblichen indirekten Kosten zu rechnen, die den bewerteten Einrichtungen durch die Erstellung der standardisierten Forschungsprofile und die dafür notwendigen Datenerhebungen entstehen. Sie können die direkten Kosten um ein Mehrfaches übersteigen.

Angesichts der großen Bedeutung, die ein verlässliches, methodisch anspruchsvolles Forschungsrating für den wissenschaftlichen Wettbewerb in Deutschland hat, ist der Aufwand nach Ansicht des Wissenschaftsrates gerechtfertigt.

III.8. Pilotstudie

Der Wissenschaftsrat empfiehlt, die Eignung der hier vorgeschlagenen Methodik in einer Pilotstudie zu überprüfen. Neben zahlreichen operativen Aspekten ist auch eine Reihe von grundsätzlicheren Fragen erst in diesem Rahmen zu klären:

- Es ist zu prüfen, wie detailliert die Taxonomie und die Kriterienliste unbedingt sein müssen, um die Ziele des Forschungsratings zu erreichen. Ziel sollte ein möglichst einfaches Verfahren sein.
- Die Kriterien müssen fachspezifisch konkretisiert und operationalisiert werden, ohne indes die Einheitlichkeit des Verfahrens aufzugeben.
- Die Möglichkeit einer weiteren Informationsverdichtung durch eine vorgegebene Gewichtung der Kriterien sollte geprüft werden.
- Es sollte überprüft werden, ob ein Teil der Daten in kürzeren Zeitabständen aktualisiert werden kann, um den Adressaten zeitnahe Informationen bereitzustellen.

- Für die Bewertung interdisziplinärer Arbeitseinheiten und Institute müssen geeignete Regeln etabliert werden, um trotz der primär fachbezogenen Struktur des Ratings adäquate Ergebnisse zu erhalten.
- Soweit Universitäten und außeruniversitäre Forschungseinrichtungen über Evaluationsverfahren verfügen, ist zu prüfen, inwieweit dafür bereits erhobene, aktuelle Daten für das Rating verwendet werden können, um die Belastung durch Erhebungen möglichst gering zu halten.
- Die Kosten-/Nutzen-Relation des Verfahrens ist zu prüfen.

Für die Pilotstudie sollten zwei Fachgebiete ausgewählt werden, die sich in methodischer Hinsicht deutlich unterscheiden und die es erlauben, möglichst viele der absehbaren Probleme zu untersuchen. Zugleich sollte es sich um Fachgebiete handeln, deren Abgrenzung möglichst unkontrovers ist, damit die Ergebnisse in die für das eigentliche Forschungsrating gewählte Taxonomie eingeordnet werden können. Geeignet wären die Fachgebiete Informatik und Soziologie.

Um die Pilotstudie im Anschluss an die Verabschiedung der vorliegenden Empfehlungen möglichst rasch beginnen zu können, sollte eine Arbeitsgruppe des Wissenschaftsrates mit der Operationalisierung des Verfahrens und der Einberufung der Bewertungsgruppen beauftragt werden. Diese könnte zugleich als Kern der künftigen Steuerungsgruppe dienen, um die Kontinuität des Verfahrens zu gewährleisten. Die großen Wissenschaftsorganisationen sind bereits in der Pilotphase zu beteiligen.

Schon im Rahmen der Pilotstudie ist es wichtig, eine aktive Informations- und Kommunikationspolitik zu betreiben, um das Verständnis für das Verfahren bei den Universitäten und außeruniversitären Forschungseinrichtungen zu verbessern und eine optimale Kooperation bei der Datenerhebung sicherzustellen.

Die Steuerungsgruppe sollte die im Rahmen der Pilotstudie gemachten Erfahrungen auswerten und dem Wissenschaftsrat darüber berichten, der seinerseits die Wissenschaftsorganisationen konsultieren wird. Der Wissenschaftsrat behält sich vor, im Anschluss an die Pilotstudie zur Umsetzung eines Forschungsratings Stellung zu nehmen.

III.9. Internationales Benchmarking

Das vorgeschlagene Forschungsrating ermöglicht eine vergleichende Bewertung der Forschungsleistungen von Universitäten und von außeruniversitären Forschungseinrichtungen in Deutschland. Die Bewertungsgruppen sind gehalten, dabei internationale Maßstäbe anzulegen. Aspekte des Verfahrens, die dies gewährleisten sollen, sind die internationale Besetzung der Bewertungsgruppen, Definition der Bewertungsskala und entsprechende Instruktion der Gutachter sowie die Verwendung international normierter Indikatoren.

Ein internationales Benchmarking von Universitäten und außeruniversitären Forschungseinrichtungen lässt sich nur in Kooperation mit anderen Ländern durchführen, da die erforderlichen Daten nicht frei zugänglich sind. Neben einer wechselseitigen Kontrolle der Bewertungsmaßstäbe kann eine derartige Kooperation dazu dienen, in methodischer Hinsicht voneinander zu lernen.

Das vom Wissenschaftsrat vorgeschlagene Verfahren weist Ähnlichkeiten zu den Bewertungsverfahren in den Niederlanden und dem Vereinigten Königreich von Großbritannien und Nordirland auf. Aufgrund dessen empfiehlt der Wissenschaftsrat, gemeinsam mit Vertretern der Niederlande und des Vereinigten Königreichs zu erkunden, wie nach Abschluss der Pilotstudie des Forschungsratings ein Benchmarking der Bewertungsverfahren und eine wechselseitige Kontrolle der Bewertungsmaßstäbe in diesen drei Ländern realisiert werden kann. Dies würde es den beteiligten Ländern ermöglichen, den Rang ihrer Forschungseinrichtungen noch überzeugender zu bestimmen.

Anhang

Literaturverzeichnis

Bayer, Ch. R. (1999): „Hochschul-Ranking. Übersicht und Methodenkritik“. *Beiträge zur Hochschulforschung*, Sonderheft.

Berghoff, S. et al. (2003a): *Das Hochschulranking. Vorgehensweise und Indikatoren*. Centrum für Hochschulentwicklung Arbeitspapier 46, Gütersloh.

Berghoff, S. et al. (2003b): *Das CHE-Forschungsranking deutscher Universitäten 2003*. Centrum für Hochschulentwicklung, Gütersloh.

Berghoff, S. & S. Hornbostel (2003): „Das CHE Hinter Den Sieben Bergen“ *Perspektiven der Wirtschaftspolitik* 4, 191–195.

Bornmann, L. & H.-D. Daniel (2003) „Begutachtung durch Fachkollegen in der Wissenschaft – Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens.“ In S. Schwarz & U. Teichler (Hrsg.), *Universität auf dem Prüfstand - Konzepte und Befunde der Hochschulforschung*. Campus Verlag, Frankfurt/New York, S. 207–225

Büttner, Th., M. Kraus & J. Rincke (2002): *Hochschulranglisten als Qualitätsindikatoren im Wettbewerb der Hochschulen*. ZEW Discussion Paper 02-78, Mannheim.

Carnegie Foundation for the Advancement of Teaching (2001): *The Carnegie classification of institutions of higher education. 2000 edition*. Menlo Park/Ca.

Centrum für Hochschulentwicklung (2002): „CHE-Forschungsranking. Forschungsstarke Fakultäten an deutschen Universitäten.“ *DUZ – das unabhängige Hochschulmagazin*, Beilage vom 8. 11. 2002

CEWS – Kompetenzzentrum Frauen in Wissenschaft und Forschung (2003): *Hochschulranking nach Gleichstellungsaspekten*. cews.publik.no.5, Bonn.

Daniel, H.-D. (1988): „Forschungsleistungen wissenschaftlicher Hochschulen im Vergleich“, in Daniel & Fisch (Hrsg.), S. 93–104.

Daniel, H.-D. (2001): „Was bewirken Hochschul-Rankings? Wer orientiert sich an ihnen?“ in D. Müller-Böling et al. (Hrsg. 2001), S. 121–124.

Daniel, H. D. & R. Fisch (Hrsg. 1988): *Evaluation von Forschung*. Universitätsverlag Konstanz, Konstanz.

Deutsche Forschungsgemeinschaft (1997): *Bewilligungen nach Hochschulen. Bewilligungsvolumen 1991 bis 1995. Anzahl kooperativer Projekte im Jahr 1996*. Bonn.

Deutsche Forschungsgemeinschaft (2003): *Förder-Ranking 2003. Institutionen - Regionen – Netzwerke*. Bonn.

Engel, U. (Hrsg. 2001): *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. Campus Verlag, Frankfurt/M.

Fabel, O., E. Lehmann & S. Warning (2002): „Der relative Vorteil deutscher wirtschaftswissenschaftlicher Fachbereiche im Wettbewerb um studentischen Zuspruch: Qualität des Studiengangs oder des Studienstandortes?“ *Zeitschrift für betriebswirtschaftliche Forschung* 54, S. 509–526.

Goldberger, M. L., B. A. Maher & P. E. Flattau (Hrsg. 1995): *Research-Doctorate Programs in the United States. Continuity and Change*. National Academy Press, Washington, D.C.

HEFCE – Higher Education Funding Council for England (2003): *Joint consultation on the review of research assessment*. Bristol.

Hornbostel, S. (1997): *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Westdeutscher Verlag, Opladen.

Hornbostel, S. (2001): „Der Studienführer des CHE – ein multidimensionales Ranking“, in: Engel (Hrsg.), S. 83-120.

Jones, L. V., G. Lindzey & P. E. Coggeshall (Hrsg. 1982): *An Assessment of Research-Doctorate Programs in the United States*. National Academy Press, Washington, D.C.

Leeuwen, Th. N. van, M. S. Visser, H. F. Moed, T. J. Nederhof, A. F. J. van Raan (2003): „The Holy Grail of Science Policy: Exploring and combining bibliometric tools in search of scientific excellence“. *Scientometrics* 57, S. 257–280.

Leszczensky, M. & F. Dölle (2003): *Ausstattungs-, Kosten- und Leistungsvergleich von Universitäten. Ergebnisse für Norddeutschland und Berlin für das Jahr 2000*. HIS Hochschul-Informations-GmbH, Hannover.

Max-Planck-Gesellschaft (2002): *Evaluation. Die Verfahren der Max-Planck-Gesellschaft*. MPG, München.

Müller-Böling, D., et al. (Hrsg. 2001): *Hochschulranking. Aussagefähigkeit, Methoden, Probleme*. Verlag Bertelsmann Stiftung, Gütersloh.

NWO, VSNU & KNAQ (2002): *Standard Evaluation Protocol 2003 – 2009. For public research organisations*. <http://www.vsnul.nl/show?id=42615&langid=246>

Ostriker, J. P. & Ch. Kuh (Hrsg. 2003): *Assessing Research-Doctorate Programs. A Methodology Study*. The National Academies Press, Washington, D.C.

Quality Assurance Agency for Higher Education (2000): *Handbook for academic review*. QAA, Gloucester.

Raan, A. F. J. van (1996): „Advanced Bibliometric Methods as Quantitative Core of Peer Review Based Evaluation and Foresight Exercises“, *Scientometrics* 36, S. 397–420.

Research Assessment Exercise (2001): *A Guide to the 2001 Research Assessment Exercise*. www.hero.ac.uk/rae/pubs

Research Assessment Exercise (2004): *RAE 2008. Initial decisions by the UK funding bodies*. Bristol, Februar 2004.

Rosigkeit, A. (1997): „Hochschul-Ranking. Hintergründe und kritische Anmerkungen zu einem modernen Bewertungsverfahren“. *Beiträge zur Hochschulforschung* 1, S. 23–49.

Stern (2003): *Der Studienführer 2003. Stern Spezial Campus & Karriere* 1/2003.

Tijssen, R. J. W., Th. N. van Leeuwen & A. F. J. van Raan (2002): *Mapping the Scientific Performance of German Medical Research*. Schattauer Verlag, Stuttgart.

Tijssen, R. J. W., S. Visser, Th. N. van Leeuwen (2002): „Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference?“ *Scientometrics* 54, S. 381–397.

Ursprung, H. W. (2003): „Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators“. *Perspektiven der Wirtschaftspolitik* 4, S. 177–189.

Wissenschaftliche Kommission Niedersachsen (2002): *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Grundzüge des Verfahrens*. Fassung vom 20. Dezember 2002.

Wissenschaftsrat (1985): *Empfehlungen zum Wettbewerb im deutschen Hochschulsystem*. Köln.

Schema der Forschungsprofile

Form und Inhalt der Forschungsprofile sollten im Zuge der Pilotstudie konkretisiert und im Zuge der Umsetzung des Forschungsratings von der Steuerungsgruppe festgelegt werden, wobei die Möglichkeit fachspezifischer Abweichungen auf Empfehlung der Bewertungsgruppen besteht. Das folgende Schema dient der Veranschaulichung der Empfehlungen des Wissenschaftsrates.

Forschungsprofil der Einrichtung ... im Fachgebiet ...

Erhebungszeitraum: ... bis ... (letzte fünf Jahre)

Stichtag: ...

Kurze Angaben zur Mission der Einrichtung, zur Rolle des Fachgebiets im Rahmen ihrer Mission und zu ihrer Strategie für das Fachgebiet

1. Wissenschaftliches Personal

a) Namensliste der wissenschaftlichen Mitarbeiter aus dem Erhebungszeitraum mit Kennzeichnung der am Stichtag aktuell beschäftigten Mitarbeiter, getrennt nach Personalkategorien und Finanzierung (Grundmittel/Drittmittel), ggf. mit Angabe von Teilzeitbeschäftigungen

b) Personalzahlen nach Kategorien und Finanzierung tabellarisch (in Vollzeitäquivalenten)

c) Arbeitseinheiten (Lehrstühle / Institute / Abteilungen) mit zugehörigen Mitarbeitern (fakultativ)

Erläuterung: Die Bewertung erfolgt unabhängig von dieser Zuordnung für die gesamten Forschungsaktivitäten in dem Fachgebiet. Es ist jedoch möglich, einzelne Arbeitseinheiten prozentual mehreren Fachgebieten zuzuordnen. Dies hat Auswirkungen auf die Bewertung der Effizienz.

2. Ausgewählte Forschungsprodukte: Liste von ausgewählten Forschungsprodukten aus dem Erhebungszeitraum. Dabei kann es sich um Bücher, Buchbeiträge, Artikel in

Fachzeitschriften, Tagungsbeiträge, Ausstellungen bzw. Ausstellungskataloge, multimediale Produktionen oder andere Produkte handeln, vorausgesetzt, sie gelten im jeweiligen Fachgebiet als Forschungsprodukt und sind vor dem Stichtag publiziert.

Für jedes Fachgebiet sind mindestens drei Forschungsprodukte zu benennen. Sind mehr als drei Wissenschaftler in dem Fachgebiet tätig, ist für je angefangene drei weitere Wissenschaftler ein weiteres Forschungsprodukt zu benennen.

Die Institutionen verpflichten sich, der Bewertungsgruppe auf Anforderung ein Exemplar jedes aufgelisteten Forschungsprodukts zur Verfügung zu stellen.

3. *Publikationen:*

a) Vollständige Publikationsliste aus dem Erhebungszeitraum, getrennt nach Medientypen (Monographien, Beiträge zu Sammelwerken, Aufsätze in referierten und nicht referierten Zeitschriften, Publikationen im Eigenverlag, elektronische Publikationen, Vorträge auf international bedeutenden Tagungen); Kennzeichnung von Publikationen, die aus Dissertationen hervorgegangen sind

b) Tabellarische Aufstellung von Publikationszahlen aus dem Erhebungszeitraum, getrennt nach Medientypen.

4. *Eingesetzte Mittel*

a) Verausgabte Drittmittel: Tabellarische Aufstellung der im Erhebungszeitraum jährlich verausgabten Drittmittel nach Drittmittelgebern (DFG, Bund, Länder, EU, Stiftungen, Wirtschaft, Sonstige); Aufstellung von extern finanzierten Stipendien

b) Eingesetzte Grundmittel (soweit fachspezifisch verfügbar)

5. *Betreute Promotionen:* Anzahl der gemeldeten Doktoranden zum Stichtag; Anzahl von Promotionen im Erhebungszeitraum; Median der Promotionsdauer; Verbleib der Promovierten ein Jahr nach dem Abschluss (akademisch/Wirtschaft/Medien/Verwaltung/arbeitslos/unbekannt, jeweils Inland/Ausland); Umfang, Finanzierung und Profil strukturierter Promotionsprogramme.

6. *Promovierter Nachwuchs:* Anzahl nach Kategorien (wissenschaftliche Mitarbeiter, selbständige Nachwuchsgruppenleiter, Emmy-Noether-Stipendiaten – ergibt sich aus

der Auflistung unter 1.); ausgeschieden im Erhebungszeitraum; Verbleib ein Jahr nach dem Ausscheiden (Professur/anderweitig akademisch/Wirtschaft/Medien/Verwaltung/arbeitslos/unbekannt, jeweils Inland/Ausland)

7. *Transfer in die Wirtschaft*: Angaben zu Industriemitteln (siehe 3.), Schutzrechten und Lizenzen, die unmittelbar aus Forschungsarbeiten hervorgegangen sind, Firmengründungen von Angehörigen der Einrichtung oder durch die Einrichtung selbst sowie Unternehmensbeteiligungen

8. *Kooperationen*: Nennung der wichtigsten Kooperationspartner aus Wissenschaft und Wirtschaft mit Angaben zur vertraglichen und finanziellen Basis der Kooperationen; Beteiligung an Projekten der Verbundförderung

9. *Fort- und Weiterbildung*: Kurze Beschreibung von Fort- und Weiterbildungsangeboten im Fachgebiet mit Angaben zum Kurrikulum, Art und Zahl der Teilnehmer, Gebühren, Ressourcen, Beurteilung durch Angebotsnutzer

10. *Forschungsbasierte Beratung und Wissenschaftskommunikation*: Angaben zu Beratungsleistungen für Wirtschaft, Politik und Verwaltung (Zahl der Verträge, Einnahmen); Angaben zu Aktivitäten der Wissenschaftskommunikation (Ausstellungen, Tag der offenen Tür, Schüler-Universität, Wissenschaft im Dialog etc.)