

# Kriterien zur Messung der Forschungsleistung an der Heinrich-Heine-Universität Düsseldorf

Abschlussbericht der AG Leistungsparameter

*Torsten Grapatn • Johannes Muck • Pascal Siegers • Jost Sieweke*



## ZUSAMMENFASSUNG

Die *Arbeitsgruppe Leistungsparameter* wurde vom Rektorat der Heinrich-Heine-Universität Düsseldorf beauftragt, Indikatoren für die Bewertung der Forschungsleistung an der Juristischen, Philosophischen und Wirtschaftswissenschaftlichen Fakultät zu entwickeln.

Basierend auf einer Umfrage unter 87 Universitäten sowie 129 philosophischen, 87 wirtschaftswissenschaftlichen und 30 juristischen Fakultäten in Deutschland und nach intensivem Studium der einschlägigen Literatur zum Thema Forschungsevaluation gelangt die AG Leistungsparameter zu folgenden Vorschlägen:

- 1) Die Messung der Forschungsleistung sollte grundsätzlich anhand der drei Indikatoren Nachwuchsförderung, eingeworbene Drittmittel und angefertigte Publikationen erfolgen.
- 2) Für jeden dieser drei Indikatoren sind verschiedene Verfahrensvorschläge zur konkreten Umsetzung denkbar. Die Auswahl des aus Sicht der Arbeitsgruppe am besten geeigneten Verfahrensvorschlags für jeden der drei Indikatoren erfolgte auf Grundlage vorab definierter Kriterien. So sollen geeignete Indikatoren zur Messung der Forschungsleistung erstens die spezifischen Wissenschaftskulturen in den einzelnen Fächern in angemessener Weise berücksichtigen, zweitens sollen sie dennoch über die Fächer hinweg konzeptionell vergleichbar aufgebaut sein und drittens sollen die Indikatoren effizient, d.h. mit einem angemessenen Arbeits- und Zeitaufwand, zu erheben sein (S. 52).
- 3) Für die Messung der Forschungsleistung im Bereich der Nachwuchsförderung empfiehlt die Arbeitsgruppe, die Zahl der Promotionen bzw. Habilitationen pro Kopf dividiert durch den fachspezifischen Bundesdurchschnitt (Promotionsquote, Habilitationsquote) zu verwenden. Dabei sollten bei den Promotionen sowohl interne als auch externe und bei den Habilitationen auch erfolgreiche Evaluationen von Juniorprofessoren/innen sowie Rufe auf W2- bzw. W3-Stellen berücksichtigt werden (S. 52).
- 4) Für die Messung der Forschungsleistung im Bereich eingeworbener Drittmittel lautet die Empfehlung der Arbeitsgruppe, an allen drei Fakultäten die in einem Jahr eingeworbenen Drittmittel pro Kopf durch die Höhe der deutschlandweit in diesem Fachgebiet durchschnittlich eingeworbenen Drittmittel zu teilen. Berücksichtigt werden sollten dabei alle Drittmittelarten (z. B. von DFG, Stiftungen, Bundes- und Landesministerien, Unternehmen usw.) (S. 53).
- 5) Hinsichtlich der Verwendung von Publikationen zur Erfassung der Forschungsleistung spricht sich die Arbeitsgruppe für ein fakultätsspezifisch differenziertes Vorgehen aus (S. 53f.). Sowohl an der Philosophischen als auch an der Wirtschaftswissenschaftlichen Fakultät sollte eine Qualitätsgewichtung der Publikationen anhand von Ranglisten erfolgen. Da für philosophische Fakultäten solche Ranglisten bislang nicht existieren, schlägt die Arbeitsgruppe die Ausarbeitung spezifischer Ranglisten für die einzelnen Fächer der Philosophischen Fakultät vor (S. 57f.). Diese sollen sowohl Monografien als auch Zeitschriftenaufsätze berücksichtigen. Für die Wirtschaftswissenschaftliche Fakultät wird empfohlen, für die Betriebswirtschaftslehre das JOURQUAL-Ranking und für die Volkswirtschaftslehre das Handelsblatt-Ranking zur Bewertung von Zeitschriftenaufsätzen zu verwenden (S. 59f.). Zur Berücksichtigung von Monografien wird empfohlen, eine Qualitätsgewichtung basierend auf dem Verlag, in welchem die Monografie veröffentlicht wurde, vorzunehmen. Aufgrund der differenzierten Publikationskultur und spezifischen Publikationsarten in den Rechtswissenschaften ist eine Evaluation der Forschungsleistung anhand von Publikationen in der Juristischen Fakultät derzeit nicht möglich (S. 61f.).

- 6) Vor der Einführung eines Systems zur Forschungsevaluation sind aus Sicht der Arbeitsgruppe noch vier zentrale Voraussetzungen zu erfüllen. So muss erstens entschieden werden, für welche Zwecke das System zur Forschungsevaluation eingesetzt werden soll, da für bestimmte Zwecke (z.B. leistungsorientierte Mittelvergabe zwischen den Fakultäten) die verwendeten Kriterien bestimmten Ansprüchen (z.B. fakultätsübergreifende Vergleichbarkeit) genügen müssen (S. 64f.). Zweitens muss festgelegt werden, ob die verwendeten Kriterien zu einem Forschungsprofil zusammengefasst oder zu einem Forschungs-Score verdichtet werden, wobei jedes Verfahren mit spezifischen Vor- und Nachteilen verbunden ist (S. 65f.). Drittens sollte das bestehende Forschungsinformationssystem ausgebaut und universitätsweit implementiert werden (S. 66f.). Idealerweise sollte dieses Forschungsinformationssystem neben den an der Heinrich-Heine-Universität verfassten Publikationen auch die Höhe und die Laufzeit der eingeworbenen Drittmittel sowie die Zahl der erfolgreichen Promotionen und Habilitationen je Professor/in beinhalten. So wäre gewährleistet, dass die für die Forschungsevaluation benötigten Daten zentral an einem Ort vorliegen. Viertens sollte das System zur Forschungsevaluation selbst in regelmäßigen Abständen im Rahmen eines Audits hinsichtlich seiner Effektivität und Effizienz evaluiert werden (S. 67f.).



---

## VORWORT

Mit diesem Gutachten legt die *Arbeitsgruppe Leistungsparameter* ihren Abschlussbericht vor. Er soll der weiteren Abstimmung zwischen Rektorat und Fakultäten als Grundlage dienen und zur Umsetzung eines ausgeglichenen und wissenschaftsadäquaten Systems der Forschungsevaluation an der Heinrich-Heine-Universität Düsseldorf beitragen.

Die Mitglieder der Arbeitsgruppe bedanken sich beim Rektorat, namentlich beim Prorektor für Strukturentwicklung, *Prof. Dr. Klaus Pfeffer*, für die Initiierung, Leitung und fortwährende Unterstützung des Projekts sowie bei *PD Dr. Thomas Bruhn* für die Projektkoordination. Ebenso gilt unser Dank den Dekanen der beteiligten Fakultäten, *Prof. Dr. Jan Busche* und *Prof. Dr. Ralph Alexander Lorz* für die Juristische Fakultät, *Prof. Dr. Bruno Bleckmann* für die Philosophische Fakultät sowie *Prof. Dr. Bernd Günter* für die Wirtschaftswissenschaftliche Fakultät. Sie haben die Arbeit der Gruppe während des gesamten Projektzeitraums intensiv begleitet und inhaltlich bereichert.

Darüber hinaus gilt unser Dank dem Vorsitzenden des Senats der Heinrich-Heine-Universität Düsseldorf, *Prof. Dr. Michael Baurmann*, der es der Arbeitsgruppe ermöglicht hat, einen Expertenworkshop auszurichten, welcher eine Basis für den wissenschaftlichen Austausch bereitgestellt und zugleich durch seine Diskussionsbeiträge und -ergebnisse zur Vervollständigung des Abschlussberichtes beigetragen hat. Bezugnehmend auf diesen Workshop bedanken wir uns bei den Referenten, *LMR Helmut Fangmann* (MIWF NRW), *Dr. Jürgen Güdler* (DFG), *Dr. Elke Lütkemeier* (Wissenschaftsrat), *Prof. Dr. Torben Schubert* (Fraunhofer ISI), *Dr. Simon Sieweke* (Finanzbehörde Hamburg) und *Dr. Matthias Winterhager* (IWT – Universität Bielefeld).

Düsseldorf, Juni 2012

*Torsten Grapatin, Johannes Muck, Pascal Siegers und Jost Sieweke*

## INHALT

Zusammenfassung

Vorwort

Inhalt

|  |    |
|--|----|
| 1. Einleitung.....   | 7  |
| 2. Grundlagen der Forschungsevaluation in Deutschland.....   | 9  |
| 3. Rechtliche Rahmenbedingungen der Forschungsevaluation in Deutschland.....   | 10 |
| 3.1. Verfassungsrechtliche Grundlagen .....  | 10 |
| 3.2. Bundesrechtliche Grundlagen .....   | 11 |
| 3.3. Landesrechtliche Grundlagen .....   | 12 |
| 3.4. Verfassungsrechtliche Zulässigkeit der Forschungsevaluation.....  | 15 |
| 3.5. Zusammenfassung.....  | 19 |
| 4. Ergebnisse einer Befragung zur Verbreitung und Ausgestaltung der Forschungsevaluation an deutschen Universitäten und Fakultäten ..... | 22 |
| 4.1. Methodik.....   | 22 |
| 4.2. Ergebnisse der Umfrage .....  | 23 |
| 4.2.1. Beschreibung der Daten.....   | 23 |
| 4.2.2. Beschreibung der eingesetzten Kriterien.....  | 25 |
| 4.2.3. Verwendung und Auswirkungen der forschungsbezogenen Leistungsevaluation .....   | 29 |
| 4.2.4. Hinderungsgründe forschungsbezogener Leistungsmessung .....   | 32 |
| 4.3. Zusammenfassung der Ergebnisse und Schlussfolgerungen.....  | 34 |
| 5. Indikatoren zur Messung der Forschungsleistung .....  | 36 |
| 5.1. Forschungsevaluation anhand von Dissertationen und Habilitationen.....  | 36 |
| 5.1.1. Anzahl der Promotionen und Habilitationen pro Kopf .....  | 37 |
| 5.1.2. Anzahl der Promotionen und Habilitationen pro Kopf relativ zum Bundesdurchschnitt .....   | 37 |
| 5.2. Forschungsevaluation anhand von Drittmittelinwerbungen .....  | 38 |
| 5.2.1. Drittmittel pro Kopf .....  | 39 |
| 5.2.2. Drittmittelfinanzierte Mitarbeiter/innenstellen .....   | 40 |
| 5.2.3. Drittmittel pro Kopf relativ zum fachspezifischen Bundesdurchschnitt.....   | 40 |
| 5.3. Forschungsevaluation anhand von Publikationen.....  | 42 |
| 5.3.1. Anzahl der Publikationen pro Kopf als nicht wissenschaftsadäquater Indikator zur Messung der Forschungsleistung.....              | 43 |
| 5.3.2. Qualitätsgewichtung anhand einfacher Ranglisten.....  | 44 |
| 5.3.3. Qualitätsgewichtung anhand von Journal-Impact-Faktoren .....  | 45 |
| 5.3.4. Zitationsanalysen .....   | 47 |
| 5.3.5. Behandlung von Buchpublikationen im Rahmen der Forschungsevaluation.....  | 48 |
| 5.4. Forschungsevaluation mittels Informed-Peer-Review .....   | 50 |



---

|  |    |
|--|----|
| 6. Empfehlungen bezüglich geeigneter Kriterien zur Messung der Forschungsleistung in der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät..... | 52 |
| 6.1 Anforderungen an geeignete Leistungsparameter .....  | 52 |
| 6.2 Bewertung der Leistungsindikatoren.....  | 52 |
| 6.3. Fakultätsspezifische Ausgestaltung der Forschungsevaluation .....   | 54 |
| 6.3.1 Philosophische Fakultät.....   | 55 |
| 6.3.2 Wirtschaftswissenschaftliche Fakultät .....  | 59 |
| 6.3.3 Juristische Fakultät.....  | 60 |
| 7. Voraussetzungen zur Einführung einer Forschungsevaluation an der HHU .....  | 64 |
| 7.1. Definition der Anwendungsgebiete der Forschungsevaluation .....   | 64 |
| 7.2. Zusammenfassung der Indikatoren zur Messung der Forschungsleistung: Forschungsprofil oder Forschungsscore.....  | 65 |
| 7.3. Erweiterung der bestehenden Infrastruktur .....   | 66 |
| 7.4. Einführung eines regelmäßigen Audits des Systems zur Forschungsevaluation .....   | 67 |
| Literatur.....   | 69 |
| Anhang 1: Drittmittel je Professor/in nach Fächergruppen und Fachbereichen.....  | 75 |
| Anhang 2: Erläuterung des Source Normalized Impact per Paper (SNIP).....   | 77 |
| Anhang 3: Befragung der Professoren/innen der Philosophischen Fakultät .....   | 79 |
| Anhang 4: Zur Befragung der Rektorate und Fakultäten verwendeter Fragebogen.....   | 89 |

---





## 1. EINLEITUNG

In den vergangenen ca. 15 Jahren haben in der deutschen Hochschullandschaft tiefgreifende Veränderungen stattgefunden. Die Veränderungen zielen in erster Linie darauf, „weniger Staat“ und „mehr Wettbewerb“ sowie einen stärkeren Marktbezug zu ermöglichen (vgl. Ridder 2006, S. 101); sie sind eng verknüpft mit Reformen, die seit längerer Zeit im gesamten öffentlichen Sektor unter Begriffen wie „New Public Management“ oder „Neue Steuerung“ diskutiert und umgesetzt werden (vgl. Sieweke 2010, S. 24-34). Auf dieser Grundlage haben Schlagworte wie Hochschulautonomie, Globalbudgets, Zielvereinbarungen oder leistungsorientierte Mittelvergabe Einzug in das Management deutscher Hochschulen gehalten (vgl. Bogumil/Heinze 2009). Innerhalb der Hochschulen werden infolgedessen Globalbudgets zunehmend leistungsorientiert verteilt (vgl. Jaeger 2009), wodurch eine stärkere Orientierung der Mittelvergabe an erwarteten, vereinbarten oder tatsächlich geleisteten Outputs ermöglicht werden soll.

Insgesamt sollen diese Veränderungen dazu beitragen, die Leistungserstellung von Hochschulen, Fakultäten und einzelnen Wissenschaftlern/innen stärker als in der Vergangenheit an Wirtschaftlichkeitskriterien auszurichten. Dies wird mit der Notwendigkeit der Profilbildung in Forschung und Lehre sowie dem zunehmenden Wettbewerb um Studierende, Wissenschaftler/innen und Drittmittel begründet (vgl. Meier 2009, S. 132). Dass damit die Einführung verschiedener Elemente einer leistungsorientierten Steuerung verbunden ist, ist unumstritten. So sind leistungsorientierte Mittelvergaben und Leistungsevaluationen an nahezu allen Hochschulen verpflichtend eingeführt worden (vgl. Jaeger et al. 2005, S. 12).

Auch an der Heinrich-Heine-Universität Düsseldorf wurde eine leistungsorientierte Mittelvergabe auf Ebene der Fakultäten eingeführt. Diese hat sich bislang an dem Kriterium der eingeworbenen kompetitiven Drittmittel orientiert. Insbesondere von der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät wurde jedoch kritisiert, dass dieses Kriterium allein kaum geeignet ist, um Forschungsleistung in diesen Fachbereichen umfassend abzubilden. Aus diesem Grund setzte das Rektorat gemeinsam mit den Dekanen die *Arbeitsgruppe Leistungsparameter* ein. Diese hatte das Ziel, geeignete Kriterien zur Messung der Forschungsleistung an der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät zu entwickeln.

Die Entwicklung des Konzepts zur forschungsbezogenen Leistungsmessung an der HHU erfolgt in vier Schritten

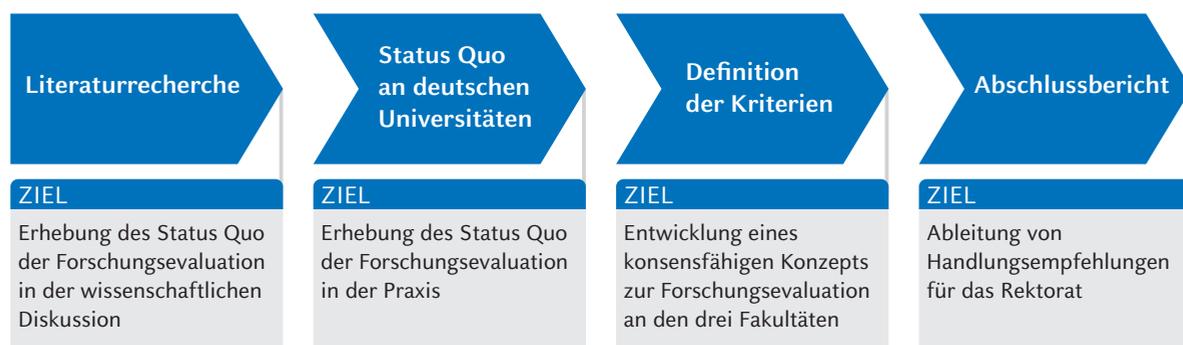


Abb. 1: Arbeitsschritte und Zwischenziele der Arbeitsgruppe Leistungsparameter bei der Entwicklung von Kriterien zur Bewertung der Forschungsleistung an der Juristischen, Philosophischen und Wirtschaftswissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf (von April 2011 bis Mai 2012)

Der vorliegende Abschlussbericht ist das Ergebnis dieser Arbeit. Er ist wie folgt aufgebaut: Nach der Einleitung werden in Kapitel 2 die Grundlagen und in Kapitel 3 die rechtlichen Rahmenbedingungen der Forschungsevaluation in Deutschland erörtert. Das 4. Kapitel stellt die Ergebnisse einer Umfrage zur Verbreitung und Ausgestaltung der Forschungsevaluation an deutschen Universitäten und Fakultäten dar. Basierend auf diesen Ergebnissen gibt Kapitel 5 einen Überblick über die Indikatoren der Forschungsleistung und zeigt verschiedene Möglichkeiten ihrer Anwendung auf. In Kapitel 6 wird erläutert, welche dieser Verfahrensvorschläge sich aus Sicht der Arbeitsgruppe jeweils am besten zur Messung der Forschungsleistung an der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät eignen. Abschließend gibt Kapitel 7 einen Überblick über die Voraussetzungen, die noch zu erfüllen sind, bevor ein System zur Messung der Forschungsleistung an der HHU eingeführt werden sollte.



## 2. GRUNDLAGEN DER FORSCHUNGSEVALUATION IN DEUTSCHLAND

Während in den USA, Australien, Großbritannien und den Niederlanden die Evaluation der Forschungsleistung von Universitäten, Fakultäten und Professoren/innen bereits seit mehreren Jahrzehnten etabliert ist, hat die Debatte in Deutschland erst in den vergangenen ca. 15 Jahren eingesetzt. Zwar lassen sich bereits in den 80er und den frühen 90er Jahren vereinzelt Arbeiten finden, die sich mit Forschungsevaluation auseinandersetzen (vgl. Heiber 1986; Pommerehne/Renggli 1986; Daniel/Fisch 1990), jedoch hatte insbesondere die Orientierung der Hochschulpolitik am sogenannten „New Public Management“ eine Steigerung der Relevanz dieses Themas zur Folge (vgl. Bogumil/Heinze 2009, S. 7f.; Meier/Schimank 2009, S. 43).

Forschungsevaluation umfasst die nachträgliche Einschätzung der Leistung einer Universität, einer Fakultät, eines Instituts oder eines Forschers/einer Forscherin hinsichtlich der Entwicklung neuer wissenschaftlicher Erkenntnisse (vgl. Frey 2007, S. 207). Sie ist jedoch kein Selbstzweck, sondern dient bestimmten Zielen. Der Wissenschaftsrat gibt beispielsweise als Ziel seiner Forschungsevaluation der Fächer Soziologie und Chemie an, dass die beurteilten Forschungseinrichtungen durch die Ergebnisse der Evaluation bei strategischen Entscheidungen unterstützt werden sollen. Darüber hinaus soll die Forschungsevaluation zu mehr Transparenz in der deutschen Forschungslandschaft führen, wodurch der Wettbewerb zwischen den evaluierten Einrichtungen gestärkt (vgl. Wissenschaftsrat 2007, S. 9) und langfristig die Qualität verbessert werden soll. Daneben wird auch die leistungsorientierte Mittelvergabe als Ziel der Forschungsevaluation genannt (vgl. Kieser 1998; Jansen et al. 2007); Ressourcen sollen an forschungsstärkere Einrichtungen verteilt werden, da bei diesen von einer effizienteren Nutzung ausgegangen wird (vgl. Kieser 1998; Jansen et al. 2007; Kreysing 2008).

Obwohl sich die Forschungsevaluation in Deutschland etabliert hat, wird sie in der wissenschaftlichen Diskussion nicht unkritisch gesehen (vgl. u.a. Kieser 1998; Süß 2006; Frey 2007). Ein grundsätzlicher Kritikpunkt ist, dass Evaluationen Anreize verändern und sich die Evaluierten zukünftig stärker an den Evaluationskriterien ausrichten, da primär das als Leistung zählt, was auch in der Evaluation gemessen wird (vgl. Frey 2007, S. 209).

Darüber hinaus wird befürchtet, dass intrinsische Motivation verdrängt werden könnte, was sich negativ auf die Forschungsleistung auswirken würde. So argumentiert Frey, dass durch eine regelmäßige Evaluation von Forschungsleistungen die Qualität und Originalität der Forschung sinkt, da innovative wissenschaftliche Arbeit intrinsischer Motivation bedarf, welche durch die Evaluation verdrängt werde (vgl. Frey 2007, S. 211). Auch eine Anreizverschiebung wird kritisiert, wenn aufgrund von Evaluationen Leistungen in der Forschung als wichtiger erachtet werden als in der Lehre und Selbstverwaltung (vgl. Dilger 2001, S. 138; Pull 2009, S. 311).

Neben diesen kritischen Stimmen gibt es aber auch Personen innerhalb und außerhalb des Wissenschaftsbetriebs, die eine systematische Evaluation der Forschungsleistungen anhand transparenter Kriterien begrüßen, zum einen, um dem Problem relativ wirkungsloser Anreize zu begegnen, zum anderen, um Bewertungsverfahren und Entscheidungen über die Allokation von Ressourcen transparenter zu gestalten (vgl. z. B. Bommer/Ursprung 1998; Ursprung 2003).

Zusammenfassend kann festgehalten werden, dass sich die Forschungsevaluation zunehmend in der deutschen Hochschullandschaft etabliert hat. Mittlerweile werden Universitäten (vgl. Kieser 1998; Valadkhani/Worthington 2006), Forschungsinstitute (vgl. Oppenländer 2002), Fachbereiche (vgl. Litzenberger/Sternberg 2005; Dyckhoff et al. 2009) und einzelne Forscher/innen (vgl. Dyckhoff et al. 2005; Rost/Frey 2011) hinsichtlich ihrer Forschungsleistung evaluiert.

### 3. RECHTLICHE RAHMENBEDINGUNGEN DER FORSCHUNGSEVALUATION IN DEUTSCHLAND

#### 3.1. Verfassungsrechtliche Grundlagen

Aus juristischer Perspektive betrachtet, betrifft die Evaluation von Forschungsleistungen im hier gegebenen Kontext vor allem das Grundrecht der Wissenschaftsfreiheit aus Art. 5 Abs. 3 Satz 1 GG. Die Wissenschaftsfreiheit hat ihren Ursprung in der Aufklärung und wurde – nach den Rückschlägen durch die Karlsbader Beschlüsse 1819 – erstmals in Art. 17 der sogenannten „oktroierten“ Verfassung Preußens vom 5. Dezember 1848, der von der Frankfurter Paulskirchenversammlung übernommen wurde, verbürgt; daher lautete § 152 der Reichsverfassung vom 28. März 1849: „Die Wissenschaft und ihre Lehre ist frei.“ Dieser Wortlaut ging auch in Art. 20 der Preußischen Verfassung von 1850 und später – nahezu unverändert – in Art. 142 der Weimarer Reichsverfassung ein. Freilich blieben diese Garantien in der Praxis oft eingeschränkt und gewannen erst in den 1920er Jahren im Zusammenhang mit der institutionellen Relevanz der Wissenschaftsfreiheit zunehmend an Bedeutung, bevor es in der Zeit des Dritten Reichs erneut zu tiefgreifenden Einschnitten kam (vgl. Pernice 2004, Rn. 1 f. m. w. N.). In Art. 5 Abs. 3 gewährleistet das Grundgesetz: „Kunst und Wissenschaft, Forschung und Lehre sind frei. Die Freiheit der Lehre entbindet nicht von der Treue zur Verfassung.“ Der Inhalt der Wissenschaftsfreiheit ist seit 1949 in Abhängigkeit zu den gesellschaftlichen Prozessen einerseits und dem wissenschaftlichen Fortschritt andererseits einem steten Wandel unterworfen. Bestimmten anfänglich noch die abwehrrechtliche und die institutionelle Bedeutung den Charakter des Grundrechts, ist die Wissenschaft heute ein „offener sozialer Prozeß vielfältiger Kommunikationen“ (Schulze-Fielitz 1995, Rn. 17).

Grundrechtsdogmatisch gewährleistet Art. 5 Abs. 3 Satz 1 GG neben der klassischen Funktion eines subjektiven Abwehrrechts auch objektiv-rechtliche Gehalte. Er stellt „nach Wortlaut und Sinngehalt eine objektive, das Verhältnis von Wissenschaft, Forschung und Lehre zum Staat regelnde wertentscheidende Grundsatznorm auf“ (BVerfGE 35, 79, 112; 111, 333, 353; 127, 87, 114; vgl. auch BVerfGE -2 BvL 4/10- Ur. v. 14.02.2012, abrufbar über Iuris, dort Rn. 159). Diese „Wertentscheidung“ schließt „das Entstehen des Staates, der sich als Kulturstaat versteht, für die Idee einer freien Wissenschaft und seine Mitwirkung an ihrer Verwirklichung ein und verpflichtet ihn, sein Handeln positiv danach auszurichten, d.h. schützend und fördernd einer Aushöhlung dieser Freiheitsgarantie vorzubeugen“ (BVerfGE 35, 79, 114; 111, 333, 353; 127, 87, 114).

Dazu gehört auch, „zur Pflege der freien Wissenschaft und ihrer Vermittlung an die nachfolgenden Generationen personelle, finanzielle und organisatorische Mittel“ bereitzustellen (BVerfGE 88, 129, 136 f., im Anschluss an BVerfGE 35, 79, 114 f.; s. auch BVerfGE 93, 85, 95; 94, 268, 285 f.). Insbesondere „im Bereich des mit öffentlichen Mitteln eingerichteten und unterhaltenen Wissenschaftsbetriebs“ hat er „durch geeignete organisatorische Maßnahmen dafür zu sorgen, dass das Grundrecht der freien wissenschaftlichen Betätigung so weit unangetastet bleibt, wie das unter Berücksichtigung der anderen legitimen Aufgaben der Wissenschaftseinrichtungen und der Grundrechte der verschiedenen Beteiligten möglich ist“ (BVerfGE 35, 79, 114 f.; 85, 360, 384; 88, 129, 136 f.; 111, 333, 353; 127, 87, 114).

Dem/der einzelnen Grundrechtsträger/in „erwächst aus dieser Wertentscheidung ein Recht auf solche staatlichen Maßnahmen auch organisatorischer Art, die zum Schutz seines grundrechtlich gesicherten Freiheitsraums unerlässlich sind, weil sie ihm freie wissenschaftliche Betätigung überhaupt erst ermöglichen“ (vgl. BVerfGE 35, 79, 116; 111, 333, 354; 127, 87, 115). Zwar ist ein Recht des Hochschullehrers/der Hochschullehrerin auf eine „angemessene Minimalausstattung“ (Denninger 2001, Art. 5 Abs. 3 I, Rn. 50) bisher nur innerhalb des Vergaberahmens anerkannt (BVerfGE 43, 242, 285), doch „grundrechtlich verbürgt ist damit eine personelle und sachliche Grundausrüstung“ (BVerfG [1. Kammer des Ersten Senats], NVwZ-RR 1998, 175



[Leitsatz 1 b]; BVerfGE 54, 363, 390; 111, 333, 362). Dies gilt „insbesondere für die nicht ‚marktgängige‘ und kurzfristig kaum angemessen zu bewertende Grundlagenforschung“. „Bei der Verteilung der verfügbaren Mittel müssen jedenfalls die Personal- und Sachmittel zugewiesen werden, die es überhaupt erst ermöglichen, wissenschaftliche Forschung und Lehre zu betreiben“ (BVerfGE 43, 242, 285; 54, 363, 390; 111, 333, 362). In jedem Falle muss die Möglichkeit zur wissenschaftlichen Arbeit gewahrt bleiben (BVerfGE 43, 242, 285 f.; 47, 327, 416 f.).

Die Schwierigkeit besteht mithin in der Notwendigkeit, den positiven Schutz der Wissenschaft vor staatlicher Bevormundung mit ihrer zugleich gegebenen Abhängigkeit von staatlicher Förderung zu einem Ausgleich zu bringen, der einerseits den Freiheitsanspruch der Hochschulangehörigen, andererseits aber auch die organisatorische Funktionsfähigkeit der Hochschule angemessen berücksichtigt (vgl. Erichsen/Scherzberg 1990, S. 8 ff.; vgl. ferner Groß 2002, S. 313 f.).

Dieses Verhältnis wird nach Ansicht des Bundesverfassungsgerichts maßgeblich über Strukturbedingungen gewährleistet (BVerfGE 35, 79, 121; 111, 333, 354 ff.), wobei Art. 5 Abs. 3 Satz 1 GG keine bestimmte Rechts- und Organisationsform des Wissenschaftsbetriebes vorsieht (BVerfGE 35, 79, 116, 120; 111, 333, 355). Da dem Gesetzgeber in diesem Bereich ein weiter Einschätzungsspielraum zusteht (BVerfGE 111, 333, 356), ergeben sich aus Art. 5 Abs. 3 Satz 1 GG auch keine konkreten Förderungspflichten. Vielmehr muss die staatliche Förderung grundsätzlich global erfolgen. Insofern wird in der Vereinbarung von Globalhaushalten eine Stärkung der von Art. 5 Abs. 3 Satz 1 GG beabsichtigten Autonomie und Planungssicherheit, aber auch der Wettbewerbsfähigkeit der Hochschulen gesehen. Problematisch wird es indes dann, wenn die notwendige Grundausstattung durch themenspezifische Projektförderungen ersetzt werden soll (vgl. auch BVerfGE 111, 333, 359). Hier droht der Wissenschaftsfreiheit ein Verlust ihrer Basis insbesondere dann, wenn die öffentliche Hochschulfinanzierung zunehmend durch den Rückgriff auf eine private Drittmittelförderung abgelöst wird. Die so auf private Förderer transferierte Steuerung des Wissenschaftsprozesses steht im Widerspruch zu der durch Art. 5 Abs. 3 Satz 1 GG vorgegebenen staatlichen Verantwortung für die Förderung und Organisation freier Wissenschaft (vgl. Pernice 2004, Rn. 60; BVerfGE 111, 333, 359).

Die hier skizzierte Grundrechtsdogmatik der Wissenschaftsfreiheit entwickelt sich analog zu den politischen Reformprozessen organisatorischer und finanzieller Hochschulstrukturen (vgl. zur Kontextabhängigkeit objektiv-rechtlicher Dimensionen und den Folgen Groß 2002, S. 313, 328 f.). Aktuell verändert sich auch das verfassungsrechtliche Verständnis der Wissenschaftsfreiheit in ihrer objektiv-rechtlichen Dimension zunehmend entlang des Leitbildes der Managementuniversität (vgl. zur Prozeduralisierung des Grundrechtsschutzes Ladeur 2005, S. 753, 761 ff.). Charakterisiert wird diese Organisationsform der Universität vor allem durch ihre Ausrichtung an Kriterien der Wirtschaftlichkeit und an der Stärkung zentraler Leitungsinstanzen. Die Managementuniversität ist die Weiterentwicklung und Neustrukturierung der Gruppenuniversität, die ihrerseits in den 60er und 70er Jahren die Ordinarienuniversität ablöste und durch die Einführung demokratischer Entscheidungsstrukturen geprägt war (vgl. zur Entwicklung der Hochschulmodelle auch Groß 2002, S. 313, 316 ff.).

### 3.2. Bundesrechtliche Grundlagen

Der Bundesgesetzgeber hat mit der 4. Novelle des Hochschulrahmengesetzes (HRG) vom 20.08.1998 (BGBl. I S. 2190) die maßgeblichen Weichen für die hochschulpolitischen Entwicklungen der vergangenen 15 Jahre gestellt. Unter anderem fielen im Zuge dieser Reform die hochschulorganisationsrechtlichen Vorschriften der §§ 60 ff. HRG weg, sodass den Ländern hier zugleich neue Spielräume eröffnet wurden. Durch die am 01.09.2006 in Kraft getretene Föderalismusreform wurde schließlich die gesamte Rahmengesetzgebungs-

kompetenz des Bundes, die sich auch auf die „allgemeinen Grundsätze des Hochschulwesens“ (Art. 75 Abs. 1 Nr. 1a GG a.F.) erstreckte, abgeschafft. Lediglich für „die Hochschulzulassung und die Hochschulabschlüsse“ (Art. 74 Abs. 1 Nr. 33 GG) wurden neue konkurrierende Gesetzgebungszuständigkeiten begründet, hinsichtlich derer sich allerdings aus Art. 72 Abs. 3 S. 1 Nr. 6 i.V.m. Art. 125b Abs. 1 GG zugleich auch eine Abweichungskompetenz der Länder ergibt. Alle anderen Bereiche des HRG gelten gemäß Art. 125a Abs. 1 S. 1 GG fort, können aber gemäß Art. 125a Abs. 1 S. 2 GG durch Landesrecht ersetzt werden.

Folgerichtig hat die Bundesregierung am 09.05.2007 den Entwurf eines *Gesetzes zur Aufhebung des Hochschulrahmengesetzes* beschlossen. Dieser Entwurf, der dem Bundestag (BT-Drs. 16/6122) und Bundesrat (BR-Drs. 352/07) im Sommer 2007 zugeleitet wurde, sah ein Außerkrafttreten des HRG zum 01.10.2008 vor. Dieser Termin konnte jedoch nicht eingehalten werden (vgl. BT-Drs. 16/11355 und BT-Drs. 16/11550). Über einen neuen Termin wurde anschließend nicht mehr entschieden. Auch die aus der Bundestagswahl vom 27.09.2009 hervorgegangene Regierung aus CDU, CSU und FDP beabsichtigt laut Koalitionsvertrag die Abschaffung des Hochschulrahmengesetzes (Koalitionsvertrag für die 17. Legislaturperiode vom 26.10.2009, S. 61 Abschn. 1.5 a.E.).

Das Hochschulrahmengesetz in der Fassung der Bekanntmachung vom 19.01.1999 (BGBl. I S. 18), zuletzt geändert durch Art. 2 des Gesetzes vom 02.04.2007 (BGBl. I S. 506), sieht die leistungsorientierte Mittelvergabe und die Evaluation der Forschungsleistung in §§ 5 und 6 vor:

#### **§ 5 Staatliche Finanzierung**

*Die staatliche Finanzierung der Hochschulen orientiert sich an den in Forschung und Lehre sowie bei der Förderung des wissenschaftlichen Nachwuchses erbrachten Leistungen. Dabei sind auch Fortschritte bei der Erfüllung des Gleichstellungsauftrags zu berücksichtigen.*

#### **§ 6 Bewertung der Forschung, Lehre, Förderung des wissenschaftlichen Nachwuchses und der Gleichstellung der Geschlechter**

*Die Arbeit der Hochschulen in Forschung und Lehre, bei der Förderung des wissenschaftlichen Nachwuchses sowie der Erfüllung des Gleichstellungsauftrags soll regelmäßig bewertet werden. Die Studierenden sind bei der Bewertung der Qualität der Lehre zu beteiligen. Die Ergebnisse der Bewertungen sollen veröffentlicht werden.*

Im Einklang mit diesen hochschulpolitischen und -rechtlichen Entwicklungen haben die Länder ihre Hochschulgesetzgebung reformiert (vgl. Sandberger (2002), S. 125, 131 ff.; ferner Groß (2002), S. 313, 325 f.; Schenke (2005), S. 1000).

### **3.3. Landesrechtliche Grundlagen**

Der nordrhein-westfälische Landesgesetzgeber hat das Landeshochschulrecht im Hochschulfreiheitsgesetz (HFG NRW) vom 31.10.2006 (GV. NRW. S. 473), das am 01.01.2007 in Kraft getreten ist, grundlegend neu geregelt. Art. 1 HFG NRW enthält das Hochschulgesetz Nordrhein-Westfalen (HG NRW). Im Allgemeinen Teil der Gesetzesbegründung zum HG NRW wird unter anderem ausgeführt:

*„Wissenschaft und Forschung schreiten rasant voran und stehen in einem immer stärkeren Wettbewerb. Die Entwicklungszyklen in der Forschung werden immer kürzer. Zudem erfordert exzellente Forschung einen*



steigenden Mitteleinsatz. Damit sich die nordrhein-westfälischen Hochschulen in diesem Umfeld erfolgreich behaupten und ihre Wettbewerbsfähigkeit steigern können, müssen daher die Rahmenbedingungen deutlich verbessert werden.

*Damit die nordrhein-westfälische Hochschul- und Forschungslandschaft noch leistungsfähiger und national wie international noch wettbewerbsfähiger wird, bedarf es gut ausgestatteter Hochschulen mit Profil, Exzellenz und Internationalität sowie einer gezielten und wettbewerbsorientierten Forschungs- und Technologieförderung. Wissenschaftlich exzellente und qualitativ hochwertig ausbildende Hochschulen sind eine unabdingbare Voraussetzung für mehr Innovation, für zukunftssichere Arbeitsplätze und für die kulturelle und ökonomische Wertschätzung Nordrhein-Westfalens.<sup>1</sup>*

Weiter heißt es dort zur Gewährleistung der Wissenschaftsfreiheit:

*„Angesichts der einfachgesetzlich in § 4 HG niedergelegten Wissenschafts-, Lehr- und Forschungsfreiheit sind Eingriffe in Wissenschaft und Forschung als einen grundsätzlich von Fremdbestimmung freien Bereich autonomer Verantwortung und damit strukturelle Gefährdungen von Wissenschaft, Lehre und Forschung nicht zu befürchten.“<sup>2</sup>*

Zur Durchführung von Evaluationen und zur Wettbewerbsfähigkeit der Universitäten wird weiter ausgeführt:

*„Mit dem Hochschulfreiheitsgesetz wird zudem das Verhältnis zwischen Staat und Hochschulen auf eine neue Grundlage gestellt, die zu deutlichen Autonomiegewinnen für die Hochschulen führen und damit ihre Forschungsstärke und Innovationskraft in der selbstverwalteten Hochschule erhöhen wird. Planerische Einzelentscheidungen des Staates wird es künftig nicht mehr geben. Vielmehr werden die Hochschulen künftig über Ziel- und Leistungsvereinbarungen sowie Methoden des Leistungscontrollings und Qualitätssicherung, wie Akkreditierung und Evaluation, ergebnisorientiert gesteuert. Hinzu kommt die aufgaben- und leistungsbezogene Finanzierung, die sich nur noch innerhalb der Vorschriften des Hochschulgesetzes bewegen wird. In diesem Rahmen müssen sich die Hochschulen zukünftig im Wettbewerb behaupten.“<sup>3</sup>*

Die Durchführung von Evaluationen hat der Landesgesetzgeber in § 7 Abs. 2 und 3 HG NRW geregelt. Dort heißt es:

### **§ 7 Qualitätssicherung durch Akkreditierung und Evaluation**

(1) <sup>1</sup>Die Studiengänge sind nach den geltenden Regelungen zu akkreditieren und zu reakkreditieren. <sup>2</sup>Die Aufnahme des Studienbetriebs setzt den erfolgreichen Abschluss der Akkreditierung voraus; die aus dem Akkreditierungsverfahren resultierenden Auflagen sind umzusetzen. <sup>3</sup>Die Akkreditierung erfolgt durch Agenturen, die ihrerseits akkreditiert worden sind. <sup>4</sup>Ausnahmen von den Sätzen 1 bis 3 bedürfen der Genehmigung durch das Ministerium.

(2) <sup>1</sup>Zur Qualitätsentwicklung und -sicherung überprüfen und bewerten die Hochschulen regelmäßig die Erfüllung ihrer Aufgaben, insbesondere im Bereich der Lehre. <sup>2</sup>Die Evaluationsverfahren regeln die Hochschulen

<sup>1</sup> Hochschulrecht Nordrhein-Westfalen, Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, 10/2007, S. 220

<sup>2</sup> Hochschulrecht Nordrhein-Westfalen, Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, 10/2007, S. 223

<sup>3</sup> Hochschulrecht Nordrhein-Westfalen, Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, 10/2007, S. 223

in Ordnungen, die auch Bestimmungen über Art, Umfang und Behandlung der zu erhebenden, zu verarbeitenden und zu veröffentlichenden personenbezogenen Daten der Mitglieder und Angehörigen enthalten, die zur Bewertung notwendig sind.<sup>3</sup>Die Evaluation soll auf der Basis geschlechtsdifferenzierter Daten erfolgen.<sup>4</sup>Die Ergebnisse der Evaluation sind zu veröffentlichen.

(3) <sup>1</sup>Das Ministerium kann hochschulübergreifende, vergleichende Begutachtungen der Qualitätssicherungssysteme der Hochschulen (Informed-Peer-Review) sowie Struktur und Forschungsevaluationen veranlassen.<sup>2</sup>Die Evaluationsberichte werden veröffentlicht.

(4) Alle Mitglieder und Angehörigen der Hochschule haben die Pflicht, an Akkreditierung und Evaluation mitzuwirken.

Für die Durchführung der Evaluation sind das Präsidium (§ 16 Abs. 1 S. 6 HG NRW) und die Dekane (§ 27 Abs. 1 S. 2 HG NRW) verantwortlich. Gemäß § 21 Abs. 1 S. 2 Nr. 4 HG NRW nimmt der Hochschulrat zu den Evaluationsberichten Stellung. Empfehlungen und Stellungnahmen des Senats zu den Evaluationsberichten ergehen nach § 22 Abs. 1 Nr. 4 HG NRW. Die Mitwirkung der Gleichstellungsbeauftragten im Rahmen der leistungsorientierten Mittelvergabe ist in § 24 Abs. 1 S. 2 HG NRW geregelt. Gemäß § 5 Abs. 1 S. 1 HG NRW orientiert sich die staatliche Finanzierung der Hochschulen an ihren Aufgaben, den vereinbarten Zielen und den erbrachten Leistungen:

#### § 5 Finanzierung und Wirtschaftsführung

(1) Die staatliche Finanzierung der Hochschulen orientiert sich an ihren Aufgaben, den vereinbarten Zielen und den erbrachten Leistungen.

In der Gesetzesbegründung zu § 5 Abs. 1 HG NRW heißt es:

*„In der Vorschrift werden die grundlegenden Aspekte und Verfahren genannt, nach denen sich künftig die Finanzierung der Hochschulen richten soll. Damit wird die Abkehr von der bisherigen Orientierung am prognostizierten Mittelverbrauch hin zu einer ergebnisorientierten Mittelzuweisung verdeutlicht. Das Finanzierungsmodell setzt durch Instrumente wie Zielvereinbarungen und leistungsbezogene Mittelvergabe auf Steuerung über Anreize und stärkt damit die Autonomie und Selbstverantwortung der Hochschule. Dabei wird auch ihrer besonderen Kostenstruktur Rechnung getragen. Insbesondere die große Bedeutung der Personalkosten sowohl für die Hochschulen wie das Land erfordert im Hinblick auf die künftige Entwicklung eine gewisse Planungssicherheit. In der Gesamtfinanzierung sollen deshalb besoldungs- und tarifrechtliche Veränderungen entsprechend berücksichtigt werden. Bei der Mittelvergabe werden nach § 5 Satz 1 Landesgleichstellungsgesetz auch Fortschritte bei der Erfüllung des Gleichstellungsauftrages aus Artikel 3 Abs. 2 Grundgesetz berücksichtigt.“<sup>4</sup>*

Zur Gewährleistung der Wissenschaftsfreiheit und insbesondere zur Absicherung der verfassungsrechtlich erforderlichen Grundausrüstung hat der Landesgesetzgeber in dem bereits angesprochenen § 4 HG NRW folgende Regelung getroffen:

#### § 4 Freiheit in Wissenschaft, Forschung, Lehre und Studium

(1) <sup>1</sup>Das Land und die Hochschulen stellen sicher, dass die Mitglieder der Hochschule bei der Erfüllung ihrer Aufgaben die durch Artikel 5 Abs. 3 Satz 1 des Grundgesetzes und durch dieses Gesetz verbürgten Rechte in

<sup>4</sup> Hochschulrecht Nordrhein-Westfalen, Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen, 10/2007, S. 232



*Lehre und Forschung wahrnehmen können. <sup>2</sup>Die Hochschulen gewährleisten insbesondere die Freiheit, wissenschaftliche Meinungen zu verbreiten und auszutauschen.*

(2) [...]

(3) <sup>1</sup>Die Freiheit der Forschung, der Lehre, der Kunstausübung und des Studiums entbindet nicht von der Treue zur Verfassung. <sup>2</sup>Entscheidungen der zuständigen Hochschulorgane sind zulässig, soweit sie sich auf die Organisation des Forschungsbetriebs sowie des Lehr- und Studienbetriebs sowie dessen ordnungsgemäße Durchführung beziehen. <sup>3</sup>Darüber hinaus sind sie zulässig, soweit sie sich auf die Förderung und Abstimmung von Forschungsvorhaben, die Bildung von Forschungsschwerpunkten und auf die Bewertung der Forschung gemäß § 7 Abs. 2, auf die Aufstellung und Einhaltung von Studien- und Prüfungsordnungen, die Erfüllung des Weiterbildungsauftrages und auf die Bewertung der Lehre gemäß § 7 Abs. 2 sowie auf die Gewährleistung eines ordnungsgemäßen Studiums beziehen. <sup>4</sup>Entscheidungen nach den Sätzen 2 und 3 dürfen die Freiheit der Forschung und der Lehre nicht beeinträchtigen. [...]

Nach den Landtagswahlen am 9. Mai 2010 hatte die Landesregierung Absichten zur Reform des bestehenden Hochschulrechts erkennen lassen. Die neue Ministerin für Innovation, Wissenschaft und Forschung, Svenja Schulze, stand verschiedenen Regelungen, die unter ihrem Vorgänger Prof. Dr. Andreas Pinkwart getroffen worden waren, kritisch gegenüber. Es ist davon auszugehen, dass mit der neu angestrebten Hochschulreform der rot-grünen Landesregierung einzelne Gestaltungsspielräume von der Hochschule wieder in den Bereich der Politik zurückverlagert werden sollten. Schulze sprach unter anderem von „konsequenterer Steuerung“ und der Setzung von „Leitplanken“<sup>5</sup>. In diesen Zusammenhang dürfte auch bereits der in § 10 der Ziel- und Leistungsvereinbarungen IV bezogene Rahmenkodex „Gute Arbeit an Hochschulen“ einzuordnen sein<sup>6</sup>. Gleiches gilt für den Austausch des Parameters *Promotionen* gegen den Parameter *Gleichstellung* als Grundlage der leistungsorientierten Mittelvergabe<sup>7</sup>. Am 14.03.2012 hat sich der nordrhein-westfälische Landtag aufgelöst. Die Ergebnisse der Landtagswahlen vom 13.05.2012 deuten auf eine Fortführung der rot-grünen Regierungskoalition und der dargestellten hochschulpolitischen Entwicklungen hin.

### 3.4. Verfassungsrechtliche Zulässigkeit der Forschungsevaluation

Das BVerfG hat in seinem Beschluss vom 26.10.2004 zur Verfassungsmäßigkeit verschiedener Vorschriften des Brandenburgischen Hochschulgesetzes (BbgHG) vom 20.05.1999 (GVBl. I S. 130) bzw. zu der durch das Erste Gesetz zur Änderung des Brandenburgischen Hochschulgesetzes vom 22.03.2004 geänderten Fassung (GVBl. I S. 394) Stellung genommen. Die zur Prüfung gestellten Normen betreffen unter anderem auch die Durchführung von Evaluationen und deren Berücksichtigung bei der leistungsorientierten Mittelzuweisung (insb. §§ 7 Abs. 1, 65 Abs. 1 S. 4 Nr. 4 und 5, 73 Abs. 3 S. 1 BbgHG) aus hochschulorganisatorischer Perspektive. Zwar handelt es sich bei dem überprüften Gesetz um brandenburgisches Landesrecht, doch lassen die Ausführungen des Bundesverfassungsgerichtes allgemeine – und in der Umsetzung auch für die nordrhein-westfälischen Hochschulen relevante – Rückschlüsse für die Durchführung von Evaluationen und die unter Beachtung der Wissenschaftsfreiheit hieran zu stellenden Anforderungen zu.

<sup>5</sup> <http://www.rp-online.de/politik/nrw/nrw-unis-fuerchten-kurze-leine-1.2637926>

<sup>6</sup> Die Ziel- und Leistungsvereinbarungen IV (2012-2013) zwischen der HHU und dem MIWF NRW wurden am 18.01.2012 unterzeichnet und sind unter folgendem Link abrufbar:

[http://www.wissenschaft.nrw.de/objekt-pool/download\\_dateien/hochschulen\\_und\\_forschung/ZLV\\_2012/Uni\\_Duesseldorf.pdf](http://www.wissenschaft.nrw.de/objekt-pool/download_dateien/hochschulen_und_forschung/ZLV_2012/Uni_Duesseldorf.pdf)

<sup>7</sup> Für den Parameter *Gleichstellung* wurde ein neuer eigenständiger Verteilungstopf mit einem Volumen von 10 Prozent der Verteilungsmasse (rd. 52 Mio. € in 2012) eingeführt. Der zuvor stattdessen angewandte Parameter *Promotion* war ebenfalls mit 10 Prozent gewichtet worden. Die „Hinweise und Erläuterungen zur leistungsorientierten Mittelverteilung an den Hochschulen in Trägerschaft des Landes NRW“ 2012-2013 sind unter folgendem Link abrufbar: [http://www.wissenschaft.nrw.de/objekt-pool/download\\_dateien/hochschulen\\_und\\_forschung/Hinweise\\_zur\\_LOM\\_Stand\\_Dez\\_2011.pdf](http://www.wissenschaft.nrw.de/objekt-pool/download_dateien/hochschulen_und_forschung/Hinweise_zur_LOM_Stand_Dez_2011.pdf)

In einem Leitsatz des angesprochenen Beschlusses stellt das Bundesverfassungsgericht klar, dass die leistungsorientierte Mittelvergabe – sofern eine wissenschaftsadäquate Leistungsbewertung gewährleistet ist – mit Art. 5 Abs. 3 Satz 1 GG vereinbar ist:

*„2. Art. 5 Abs. 3 Satz 1 GG enthält kein Verbot, an die Bewertung wissenschaftlicher Qualität Folgen bei der Mittelvergabe anzuknüpfen. Die Entscheidung des Gesetzgebers, die Verteilung von Mitteln im Hochschulbereich auch leistungsorientiert vorzunehmen, ist verfassungsrechtlich nicht zu beanstanden, wenn eine wissenschaftsadäquate Bewertung der Leistung hinreichend gewährleistet ist.“ (BVerfGE 111, 333)*

In den Entscheidungsgründen nimmt das Bundesverfassungsgericht zur verfassungsrechtlichen Zulässigkeit hochschulorganisatorischer Regelungen Stellung und führt in diesem Zusammenhang aus, welche Anforderungen an die als Grundlage leistungsorientierter Mittelvergabe verwendeten Evaluationskriterien zu stellen sind.

Dazu heißt es in BVerfGE 111, 333, 358 ff.:

*„2. Auch die Kompetenz der Leitungsorgane zur Evaluation von Lehre und Forschung sowie zur Berücksichtigung der Ergebnisse bei der Ressourcenverteilung nach § 65 Abs. 1 Satz 4 Nr. 4 und 5, sowie § 73 Abs. 3 Satz 1 BbgHG ist bei verfassungsgemäßer Auslegung mit Art. 5 Abs. 3 Satz 1 GG vereinbar.*

*a) Für die Wissenschaft als Bereich autonomer Verantwortung, der nicht durch bloße gesellschaftliche Nützlichkeits- und politische Zweckmäßigkeitvorstellungen geprägt sein darf (vgl. BVerfGE 47, 327 <370>), birgt diese Kompetenz allerdings nicht nur unerhebliche Gefahren. Der von ihrer Wahrnehmung möglicherweise ausgelöste Druck zur Orientierung an extern gesetzten Bewertungskriterien kann zu Fehlentwicklungen führen. Die für die Evaluation benutzten Bewertungskriterien müssen hinreichenden Raum für wissenschaftseigene Orientierungen belassen. Dieses Erfordernis gilt unabhängig davon, ob solche Kriterien hochschulextern oder -intern festgesetzt werden; bei einer externen Festsetzung besteht aber ein erhöhtes Risiko der Vernachlässigung wissenschaftsadäquater Belange, etwa durch Nutzung der Evaluation zur Erreichung wissenschaftsfremder Zwecke. Evaluationskriterien haben eine gesteigerte Bedeutung, wenn die Verteilung öffentlicher Mittel an die Evaluationsergebnisse geknüpft wird, weil die Hochschulangehörigen auf den öffentlichen Wissenschaftsbetrieb und dessen Ressourcen angewiesen sind.*

*Ein Verbot der Bewertung wissenschaftlicher Qualität oder ein Verbot, an die Bewertung Folgen bei der Ressourcenverteilung zu knüpfen, lässt sich Art. 5 Abs. 3 Satz 1 GG allerdings nicht entnehmen. Forschungsleistungen und Forschungsvorhaben werden seit jeher nicht nur in Prüfungen und Qualifikationsverfahren, sondern auch in Berufungsverfahren und bei der Vergabe von Drittmitteln bewertet. Ebenso zulässig ist die Bewertung im Rahmen hochschulinterner Ressourcenverteilung. Jede Ressourcenverteilung, auch die nicht an eine Evaluation anknüpfende Verteilung durch Organe der Selbstverwaltung, kann wissenschaftsfremden Einflüssen ausgesetzt sein. Die Absicht des Gesetzgebers, Allokationsentscheidungen möglichst rational und im Interesse der Effektivierung der Ressourcenverwendung auch leistungsorientiert zu steuern, ist bei wissenschaftsadäquater Bewertung der in der Forschung erbrachten und zu erwartenden Leistungen verfassungsrechtlich nicht zu beanstanden.*

*b) Für die Sicherung wissenschaftsadäquater Evaluationskriterien ist zur Vermeidung wissenschaftsinadäquater Steuerungspotentiale eine angemessene Beteiligung der Vertreter der Wissenschaft im Verfahren der Festlegung der Kriterien unabdingbar. Dabei ist auch darauf Rücksicht zu nehmen, dass diese Kriterien in den verschiedenen Disziplinen unterschiedlich sein können und gegebenenfalls auch sein müssen. Außerdem sind disziplinübergreifende Unterschiede in Rechnung zu stellen, etwa hinsichtlich einer abstrakt-theoretischen*



*Grundlagenforschung mit (ungewissem) langfristigem Ertrag gegenüber einer kurzfristig ausgerichteten anwendungs- und nachfrageorientierten wissenschaftlichen Tätigkeit. Eine Evaluation allein oder ganz wesentlich anhand eines einzigen Kriteriums, etwa eingeworbener Drittmittel, würde dem nicht gerecht. Soweit die Einwerbung von Drittmitteln als Bewertungskriterium dient, darf es sich nicht um Drittmittel handeln, deren Entgegennahme Anreize für eine auftrags- und ergebnisorientierte Forschung setzt.*

[...]

*Jedenfalls im gegenwärtigen Stadium der Diskussion, Erprobung und erst allmählichen Herausbildung bewährter Praktiken der Wissenschaftsevaluation auf lokaler, nationaler und internationaler Ebene ist der Gesetzgeber verfassungsrechtlich noch nicht gehalten, solche Kriterien festzuschreiben. Vielmehr kann er im Rahmen seines Einschätzungs- und Prognosespielraums ein Modell etablieren, in dem die Herausarbeitung von solchen Kriterien einem inneruniversitären Prozess überlassen bleibt. Dieser Prozess ist durch Regelungen des Brandenburgischen Hochschulgesetzes gestaltet, und im Ergebnis besteht keine strukturelle Gefahr für die Möglichkeit, Wissenschaft frei auszuüben. Den Gesetzgeber trifft aber insoweit eine Beobachtungs- und gegebenenfalls eine Nachbesserungspflicht, sobald Gefahren für die Wissenschaftsfreiheit, zum Beispiel durch die Verwendung wissenschaftsinadäquater Kriterien, auftreten (vgl. BVerfGE 95, 267 <314> m. w. N.).“*

Unter Bezugnahme auf die in § 4 Abs. 2 BbgHG gewährleistete Garantie der Freiheit der Forschung (vgl. § 4 HG NRW) hat das Bundesverfassungsgericht weiter ausgeführt:

*„bb) Eine weitere Begrenzung ergibt sich aus der Garantie der Freiheit der Forschung in § 4 Abs. 2 BbgHG. Insbesondere für die nicht unmittelbar ‚marktgängige‘ und kurzfristig kaum angemessen zu bewertende Grundlagenforschung müssen durch eine Grundausstattung gesicherte Bereiche bleiben. Bei der Verteilung der verfügbaren Mittel müssen jedenfalls die Personal- und Sachmittel zugewiesen werden, die es überhaupt erst ermöglichen, wissenschaftliche Forschung und Lehre zu betreiben [...].“*

*Das Brandenburgische Hochschulgesetz enthält [...] keine ausdrückliche Regelung über die Gewährleistung einer angemessenen Beteiligung an den zur Verfügung stehenden Sach- und Personalmitteln. Art. 5 Abs. 3 Satz 1 GG und auch die einfachrechtliche Garantie der Freiheit der Forschung in § 4 Abs. 2 BbgHG fordern aber, dass die Möglichkeit zu wissenschaftlichem Arbeiten für jeden Grundrechtsträger auch bei einer Ressourcenverteilung aufgrund der Evaluationsergebnisse bestehen bleibt.“ (BVerfGE 111, 333, 362)*

Der Hochschulreformprozess der vergangenen 15 Jahre ist auch von der rechtswissenschaftlichen Literatur intensiv und kritisch begleitet worden. Dabei wird durchaus anerkannt, dass der Gesetzgeber zunächst mit der Hochschul-, später auch mit der Föderalismusreform der landesrechtlichen Gestaltung des Hochschulwesens neue Perspektiven eröffnet hat.

Im Fokus der Kritik steht hingegen, dass sich mit der Implementierung dieser Perspektiven eine Tendenz hin zum dienstleistungsorientierten Hochschulmanagement auch in den wissenschaftlichen Tätigkeitsbereichen abzeichnet (Gärditz 2005, S. 407; Schenke 2005, S. 1000). Indiziell werden in diesem Zusammenhang die Einsetzung externer Organe und – insbesondere in Verbindung mit Konzepten der leistungsorientierten Mittelvergabe – die Bewertungen von Forschungs- und Lehrleistungen genannt (Gärditz 2005, S. 407). Zwar werden die Motive des Gesetzgebers zur Schaffung effizienter Entscheidungs- und Mittelvergabestrukturen vor dem Hintergrund seiner Budgetverantwortung als berechtigt angesehen (Gärditz 2009, S. 353, 366 f.; Groß 2002, S. 313, 324; Mehde 2011, S. 179, 191), in Frage gestellt wird aber, ob die hiermit verbundene organisatorische Neuordnung und die Anleihe ökonomischer Konzepte nicht im Widerspruch zu den Anforderungen der grundrechtlich geschützten Wissenschaftsfreiheit stehen (Bull/Mehde 2000, S. 650, 655; Gärditz 2005,

S. 407; Geis 2004, S. 2, 8; Görisch 2003, S. 583, 584 ff.; Hoffacker 2003, S. 92, 97 ff.; Hufeld 2002, S. 309, 310 ff.; Müller-Terpitz 2011, S. 236 ff.; Sandberger 2011, S. 118 ff.; Schenke 2005, S. 1000, 1002; Schlüter 2011, S. 264 ff.; Seidler 1999, S. 261 ff.; ferner Groß 2002, S. 313, 324).

In Bezug auf den oben vorgestellten Beschluss des Bundesverfassungsgerichts wird kritisiert, dass das durch den brandenburgischen Landesgesetzgeber eingeführte Organisationsmodell unter Verweis auf die verfassungs- und rechtskonforme Gesetzesanwendung durch die Leitungsorgane sowie den bestehenden Rechtsschutz nur beiläufig besprochen wird (BVerfGE 111, 333, 355 ff.; dazu Gärditz 2005, S. 407). Darüber hinaus bliebe das zentrale Verträglichkeitskriterium der Wissenschaftsadäquanz zu unkonkret und sei mithin zu einer Abgrenzung von verfassungsrechtlich unzulässigen Organisationsformen und -verfahren weitestgehend untauglich (vgl. Gärditz 2005, S. 407; Ladeur 2005, S. 753, 761; vgl. zum Maßstab der Wissenschaftsadäquanz als Verweis auf die Eigengesetzlichkeit der Wissenschaft bereits Groß 2002, S. 313, 320 ff.; zur Überprüfbarkeit Mehde 2011, S. 179, 206).

Bedenken werden auch im Zusammenhang mit der leistungsorientierten Mittelvergabe erhoben, da diese eine Evaluation von Forschung und Lehre voraussetzt und auf Grund der finanziellen Implikationen zugleich hohe Anforderungen an die verwendeten Kriterien stellt (Gärditz 2005, S. 407, 409; Schenke 2005, S. 1000, 1002 f.). Denn die Rationalität der neuen Verteilungsschlüssel hängt maßgeblich von der Entwicklung angemessener Evaluationsverfahren ab (Gärditz 2009, S. 353, 358; Groß 2002, S. 313, 324). Angesichts der dazu in dem Beschluss angesprochenen, „nicht nur unerheblichen Gefahren“ (BVerfGE 111, 333, 358) wird kritisiert, dass das Bundesverfassungsgericht eine gesetzliche Festschreibung von Evaluationskriterien bzw. -verfahren für entbehrlich hält und deren Entwicklung dem inneruniversitären Prozess überantwortet (BVerfGE 111, 333, 360; dazu Gärditz 2005 S. 407, 409; vgl. auch Mehde 2011, S. 179, 193; zu den Gefahren Ladeur 2005, S. 753, 760).

Nach Ansicht der Literatur lässt die Verlagerung der Kriterienkompetenz von der legislativen Ebene auf die Ebene der universitären Selbstverwaltung die in Anbetracht der verfassungsrechtlichen Implikationen durch den Gesetzgeber zu gewährleisten Rechtssicherheit vermissen (vgl. Gärditz 2009, S. 353, 388; vgl. zu den mit der reformierten Hochschulorganisation verbundenen Legitimationsproblemen Bull/Mehde 2000, S. 650, 656 f.; Hoffacker, 2003, S. 92, 102; Schenke 2005, S. 1000, 1004 ff.). Der Gesetzgeber steht zudem in der Kritik, Evaluationspflichten zu normieren und daran Rechtsfolgen bei der Mittelvergabe zu knüpfen, ohne dass die hierzu erforderlichen wissenschaftsadäquaten Verfahren hinreichend entwickelt seien (vgl. Groß 2002, S. 313, 326 f.).

In der Folge kommt den durch das zuständige Ministerium festgelegten Vergabekriterien auch auf universitärer Ebene eine Orientierungswirkung zu (vgl. zum Parameterwechsel in NRW Fn. 2; vgl. ferner Gläser et al. 2009, S. 329, 350 f.). Zugleich führt die leistungsorientierte Mittelvergabe ohne die Vorgabe bestimmter Kriterien zu Interessenkonflikten zwischen den Gremien und Organisationseinheiten auf universitärer Ebene sowie zur Förderung von Partikularinteressen einzelner Wissenschaftler/innen und wissenschaftlicher Einheiten (vgl. ferner Groß 2002, S. 313, 324). Diese Defizite wirken sich beispielsweise auf die interfakultäre Vergleichbarkeit und Mittelverteilung aus (vgl. ferner Gläser et al. 2009, S. 329, 336 ff.; dazu Gärditz 2009, S. 353, 355; ferner Mehde 2011, S. 179, 193) und können zu einer an politischen, finanziellen Vorgaben orientierten und im Ergebnis wissenschaftsinadäquaten Vergabesystemik führen.

Verstärkt werden diese Problemstellungen durch die Komplexität der Kriterienfindung selbst (vgl. Gärditz 2009, S. 353, 376; Schenke 2005, S. 1000, 1002 f.). Verfassungsrechtlich obliegt die Bewertung von Forschungsleistungen dem wissenschaftlichen Diskurs (BVerfG (2. Kammer des Ersten Senats), NJW 2000, S. 3635; ferner Groß 2002, S. 313, 325 f.).



Dieser lässt sich jedoch nicht messen und kann seinerseits Veränderungen unterworfen sein (vgl. Gärditz 2009, S. 353, 377; vgl. grundlegend zu den Optionen und Problemen der Kriterienfindung Kuhlmann/Heinze 2004, S. 53, 56 ff, 67; darüber hinaus zur Datenverfügbarkeit und institutionellen Perspektiven dies. 2004, S. 125, 129 ff., 142 ff.; ferner Löwer 2000, S. 302, 325 ff.).

Der mithin notwendige Rückgriff auf inhaltsneutrale Kriterien vermag indes keine qualitativen Aussagen zu treffen. Er muss sich daher quantitativ messbarer Ersatzindikatoren als Surrogat für qualitative Bewertungen bedienen (Bull/Mehde 2000, S. 650, 651; Gärditz 2009, S. 353, 366, 381 ff.; Hoffacker 2003, S. 92, 99 ff.; Seidler 1999, S. 261, 264 f.). Die Verlagerung hin zu einer quantitativen Erhebung der Forschungsleistung fördert indes die Entwicklung einer outputorientierten Forschung, die qualitative Ansprüche zwangsläufig zurückstellt und zugleich Gefahr läuft, andere Bereiche des Wissenschaftsbetriebes, wie etwa die Lehre oder das Engagement in der akademischen Selbstverwaltung, zu vernachlässigen (Gärditz 2005, S. 407, 409; ferner Groß 2002, S. 313, 326; Seidler 1999, S. 261, 265; vgl. zu einschlägigen Erfahrungen mit dem australischen Evaluationssystem Gläser et al. 2009, S. 329, 347; dazu Gärditz 2009, S. 353, 355, zur demotivierenden Wirkung von Evaluationen S. 377 f.).

Hinsichtlich des Evaluationskriteriums der Nachwuchsförderung – insbesondere der Erhebung von abgeschlossenen Promotionen und Habilitationen – wird hier bereits kritisch vom Prinzip „bounty hunter“ gesprochen (Gärditz 2005, S. 407, 409; 2009, S. 353, 382; Hoffacker, 2003, S. 92, 99 ff.; ferner Bull/Mehde 2000, S. 650, 652). Darüber hinaus ergeben sich auch aus der heterogenen Struktur wissenschaftlicher Forschung Schwierigkeiten dahingehend, dass sich etwa die Grundlagenforschung mit rein quantitativen Kriterien kaum adäquat messen lässt (vgl. auch BVerfGE 111, 333, 362; zu den Konsequenzen Gärditz 2009, S. 353, 379). Soweit das Bundesverfassungsgericht in diesem Kontext zwar auf die nach wie vor notwendige „Grundausstattung“ des/der einzelnen Forschers/Forscherin verweist, lässt es jedoch offen, welche Ressourcen letztlich als unabdingbar betrachtet werden dürfen (vgl. Gärditz 2005, S. 407, 409 m.w.N.; 2009, S. 353, 370, zur staatlichen Sicherstellung einer abstrakten Drittmittelfähigkeit S. 371 f., zum fächerspezifischen Mittelbedarf S. 373, zur Grundfinanzierung S. 384 f.; dazu auch Seidler 1999, S. 261, 266). Die vollständige Überantwortung der Festlegung von Evaluationskriterien und der daran orientierten Messung der Forschungsleistung auf die Wissenschaftsgemeinde „wird der staatlichen Verantwortung für eine wissenschaftsadäquate Hochschulorganisation nicht gerecht“ (Gärditz 2005, S. 407, 409). Der objektiv-rechtliche Gehalt der Wissenschaftsfreiheit wird in Ermangelung eines wirksamen Grundrechtsschutzes in der verfahrenstechnischen und organisatorischen Ausgestaltung zunehmend auf die individuell-abwehrrechtliche Dimension verlagert (Gärditz 2005, S. 407, 409). Auch in diesem Bereich setzen sich aber die oben angesprochenen Problemstellungen – unter Hinzukommen der Problematik der Grundgehaltssätze in der W-Besoldung (vgl. BVerfG -2 BvL 4/10- Urt. v. 14.02.2012) – fort. Es besteht mithin die Gefahr, dass die Universitäten „zunächst einmal Experimentierfeld für die verschiedensten Instrumente ökonomischer Effizienzsteigerung bleiben“ (Gärditz 2005, S. 407, 410; 2009, S. 353, 354).

### 3.5. Zusammenfassung

Die angesprochenen Kritikpunkte werfen wichtige Fragen auf, die konstruktiv in den Implementierungsprozess der Forschungsevaluation einfließen sollten.

Rechtlich relevant sind hingegen die gegebenen gesetzlichen Grundlagen sowie die verfassungsgerichtlichen Vorgaben. In seinem Beschluss vom 26.10.2004 zum Brandenburgischen Hochschulgesetz (BVerfGE 111, 333 ff.) hat das Bundesverfassungsgericht keinen Verstoß gegen die durch Art. 5 Abs. 3 Satz 1 GG gewährleistete Wissenschaftsfreiheit festgestellt. Insbesondere „ein Verbot der Bewertung wissenschaftlicher Qualität oder

ein Verbot, an die Bewertung Folgen bei der Ressourcenverteilung zu knüpfen, lässt sich Art. 5 Abs. 3 Satz 1 GG [...] nicht entnehmen“ (BVerfGE 111, 333, 359). „Die Entscheidung des Gesetzgebers, die Verteilung von Mitteln im Hochschulbereich auch leistungsorientiert vorzunehmen, ist verfassungsrechtlich nicht zu beanstanden, wenn eine wissenschaftsadäquate Bewertung der Leistung hinreichend gewährleistet ist“ (BVerfGE 111, 333).

#### Anforderung an Kriterien zur Forschungsevaluation



Abb. 2: Übersicht über die Anforderungen an die Ausgestaltung von Kriterien der Leistungsbewertung aus juristischer Perspektive

Unter diesem Gesichtspunkt hat das Bundesverfassungsgericht folgende Anforderungen aufgestellt, die an die Evaluationskriterien zu stellen sind<sup>8</sup>:

- „Die für die Evaluation benutzten Bewertungskriterien müssen *hinreichenden Raum für wissenschaftseigene Orientierungen* belassen.“ (BVerfGE 111, 333, 358)
- „Für die Sicherung wissenschaftsadäquater Evaluationskriterien ist zur Vermeidung wissenschaftsinadäquater Steuerungspotentiale eine *angemessene Beteiligung der Vertreter der Wissenschaft im Verfahren der Festlegung der Kriterien* unabdingbar.“ (BVerfGE 111, 333, 359)
- „Dabei ist auch darauf Rücksicht zu nehmen, dass diese Kriterien *in den verschiedenen Disziplinen unterschiedlich sein können und gegebenenfalls auch sein müssen*. Außerdem sind *disziplinübergreifende Unterschiede in Rechnung zu stellen*, etwa hinsichtlich einer abstrakt-theoretischen Grundlagenforschung mit (ungewissem) langfristigem Ertrag gegenüber einer kurzfristig ausgerichteten anwendungs- und nachfrageorientierten wissenschaftlichen Tätigkeit.“ (BVerfGE 111, 333, 359)
- „Eine *Evaluation allein oder ganz wesentlich anhand eines einzigen Kriteriums*, etwa eingeworbener Drittmittel, *würde dem nicht gerecht*.“ (BVerfGE 111, 333, 359)
- „Soweit die Einwerbung von Drittmitteln als Bewertungskriterium dient, *darf es sich nicht um Drittmittel handeln, deren Entgegennahme Anreize für eine auftrags- und ergebnisorientierte Forschung setzt*.“ (BVerfGE 111, 333, 359)

<sup>8</sup> Hervorhebungen in den Zitaten durch die Autoren



- „Insbesondere für die nicht unmittelbar ‚marktgängige‘ und kurzfristig kaum angemessen zu bewertende Grundlagenforschung müssen durch eine Grundausrüstung gesicherte Bereiche bleiben. Bei der Verteilung der verfügbaren Mittel müssen jedenfalls die Personal- und Sachmittel zugewiesen werden, die es überhaupt erst ermöglichen, wissenschaftliche Forschung und Lehre zu betreiben.“ (BVerfGE 111, 333, 362)
- „Art. 5 Abs. 3 Satz 1 GG und auch die einfachrechtliche Garantie der Freiheit der Forschung in § 4 Abs. 2 BbgHG fordern [...], dass die Möglichkeit zu wissenschaftlichem Arbeiten für jeden Grundrechtsträger auch bei einer Ressourcenverteilung aufgrund der Evaluationsergebnisse bestehen bleibt.“ (BVerfGE 111, 333, 362)

## 4. ERGEBNISSE EINER BEFRAGUNG ZUR VERBREITUNG UND AUSGESTALTUNG DER FORSCHUNGSEVALUATION AN DEUTSCHEN UNIVERSITÄTEN UND FAKULTÄTEN

Um bei der Entwicklung eines Konzepts zur Messung der Forschungsleistung an der HHU die Erfahrungen anderer Hochschulen im Umgang mit diesem Thema zu berücksichtigen, wurde eine Umfrage durchgeführt, die sich sowohl an die Rektorate aller deutschen Universitäten als auch die Dekanate aller wirtschaftswissenschaftlichen und juristischen Fakultäten sowie an die Dekanate ausgewählter philosophischer Fakultäten richtete. Die Ergebnisse dieser Umfrage werden im Folgenden dargestellt. In Abschnitt 4.1. wird zunächst das methodische Vorgehen erläutert, bevor in Abschnitt 4.2 die Ergebnisse der statistischen Auswertungen präsentiert werden. Anschließend werden in Abschnitt 4.3 die zentralen Schlussfolgerungen hinsichtlich geeigneter Indikatoren zur Messung der Forschungsleistung an der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät gezogen.

### 4.1 Methodik

Die Durchführung der Umfrage verlief in drei Schritten:

Im ersten Schritt wurden alle staatlichen und privaten deutschen Universitäten sowie alle Fakultäten mit philosophischen, wirtschaftswissenschaftlichen und juristischen Fächern identifiziert. Die Identifikation der Universitäten erfolgte anhand des Hochschulkompasses der Hochschulrektorenkonferenz. Die für die Umfrage relevanten philosophischen Fakultäten wurden auf Basis der Mitgliederliste des Philosophischen Fakultätentages ermittelt. Nach einem Abgleich mit der Mitgliederliste der Hochschulrektorenkonferenz wurden einzelne soziologische und sozialwissenschaftliche Fakultäten, die nicht Mitglied des Philosophischen Fakultätentages sind, der Grundgesamtheit hinzugefügt. Für die Wirtschaftswissenschaften wurde die Mitgliederliste des Wirtschafts- und Sozialwissenschaftlichen Fakultätentages (vgl. Wirtschafts- und Sozialwissenschaftlicher Fakultätentag 2010) mit den im aktuellen CHE-Ranking aufgeführten Fakultäten in den Bereichen Volkswirtschaftslehre, Betriebswirtschaftslehre und Wirtschaftswissenschaften abgeglichen (vgl. DIE ZEIT 2011). Die juristischen Fakultäten schließlich wurden aus der Mitgliederliste des Deutschen Juristen-Fakultätentages übernommen (vgl. Deutscher Juristen-Fakultätentag 2010) und ebenfalls mit den im aktuellen CHE-Ranking aufgeführten juristischen Fakultäten abgeglichen (vgl. DIE ZEIT 2011). Da die Fächer Wirtschafts- und Rechtswissenschaften zum Teil in einer Fakultät zusammengefasst werden, wurden die Listen der wirtschaftswissenschaftlichen und juristischen Fakultäten um Dopplungen bereinigt.

Die Grundgesamtheit umfasste 87 Universitäten sowie 246 Fakultäten, davon 129 philosophische, 87 wirtschaftswissenschaftliche (inklusive „Mischfakultäten“ wie z. B. wirtschafts- und rechtswissenschaftliche Fakultäten und wirtschafts- und sozialwissenschaftliche Fakultäten) sowie 30 juristische Fakultäten.

Im zweiten Schritt wurden die Prorektoren/innen für Forschung der 87 Universitäten sowie die Dekane/innen der 246 Fakultäten zum Stichtag 31.05.2011 identifiziert.<sup>9</sup> Diese waren in der Studie Ansprechpartner/innen, da sie als Leiter/innen des Forschungsressorts bzw. als Leiter/innen der Fakultäten als besonders kompetent zur Beantwortung von Fragen hinsichtlich der forschungsbezogenen Leistungsbewertung erachtet wurden.

---

<sup>9</sup> Für die philosophischen Fakultäten war der Stichtag der 1. Oktober 2011.



Im dritten Schritt wurde der Fragebogen zusammen mit einem personalisierten Anschreiben postalisch an die Prorektoren/innen und die Dekane/innen versendet. Um die Rücklaufquote zu erhöhen, konnte der Fragebogen alternativ zur postalischen Rücksendung auch online ausgefüllt werden. Die Befragung der philosophischen Fakultäten wurde online durchgeführt, wobei die Befragten die Möglichkeit hatten, einen Papierfragebogen anzufordern.

Die Befragung der Universitäten sowie der wirtschaftswissenschaftlichen und juristischen Fakultäten lief von Juli bis September 2011, die Befragung der philosophischen Fakultäten fand etwas zeitversetzt, vom 30. Oktober bis 27. November 2011, statt.<sup>10</sup> Zwei juristische Fakultäten, eine wirtschafts- und rechtswissenschaftliche Fakultät und sechs philosophische Fakultäten lehnten explizit eine Teilnahme an der Studie ab. Insgesamt wurden von den Rektoraten 35 und von den Dekanaten 165 Fragebögen zurückgesendet, davon 99 von philosophischen, 47 von wirtschaftswissenschaftlichen (inklusive der „Mischfakultäten“) und 19 von juristischen Fakultäten.<sup>11</sup> Dies entspricht einer Rücklaufquote von durchschnittlich 40,2 Prozent auf Ebene der Rektorate sowie 67,1 Prozent auf Fakultätsebene (philosophische Fakultäten: 76,7 Prozent; wirtschaftswissenschaftliche Fakultäten: 54,0 Prozent; juristische Fakultäten: 63,3 Prozent). Allerdings waren nicht alle Fragebögen vollständig beantwortet, wodurch die Zahl der Beobachtungen zwischen den Fragen leicht variiert.

Der Fragebogen wurde auf Grundlage der Ergebnisse früherer Studien (vgl. Minssen/Wilkesmann 2003; Jaeger et al. 2005; Jaeger 2006) entwickelt und bestand aus drei Teilen. Im ersten Teil wurden Charakteristika der Universität bzw. Fakultät, wie ihre Bezeichnung, ihr Alter und ihre Größe, erhoben. Die Fragen im zweiten Teil zielten auf die Evaluation der Forschungsleistung an der Universität bzw. Fakultät. Hier lag der Fokus beispielsweise auf der Ausgestaltung des Evaluationssystems. Dieser Teil wurde nur von Befragten ausgefüllt, deren Universität bzw. Fakultät bereits ein System zur Evaluation der Forschungsleistung eingeführt hatte. Die übrigen Befragten wurden im dritten Teil des Fragebogens gebeten anzugeben, aus welchen Gründen bislang kein System eingeführt wurde, und ob die Einführung zukünftig geplant ist. Um die Verständlichkeit der Fragen sowie die Vollständigkeit der Antwortmöglichkeiten zu kontrollieren, wurde der Fragebogen vorab dem Referenten/der Referentin eines Prorektors/einer Prorektorin sowie jeweils dem Dekan/der Dekanin einer philosophischen, einer wirtschaftswissenschaftlichen und einer juristischen Fakultät zur Überprüfung vorgelegt.

## 4.2. Ergebnisse der Umfrage

### 4.2.1. Beschreibung der Daten

Der Großteil der teilnehmenden Universitäten (80,0 Prozent) sowie fast alle Fakultäten (94,9 Prozent) befinden sich in staatlicher Trägerschaft, lediglich 20 Prozent der Universitäten haben einen privaten Träger. Bei den Fakultäten sind 0,7 Prozent in privater, 2,2 Prozent in kirchlicher und 2,2 Prozent in sonstiger Trägerschaft. Insgesamt sind 60,0 Prozent der Teilnehmer philosophische, 15,2 Prozent rein wirtschaftswissenschaftliche und 11,5 Prozent rein juristische Fakultäten. Wirtschafts- und sozialwissenschaftliche sowie rechts- und wirtschaftswissenschaftliche Fakultäten sind in den Antworten mit je 4,9 Prozent vertreten.

<sup>10</sup>Die Befragung der philosophischen Fakultäten musste aus organisatorischen Gründen später starten.

<sup>11</sup>Für die philosophischen Fakultäten lässt sich die Rücklaufquote nicht ohne weiteres bestimmen. 92 Teilnehmer haben die erste Seite des Fragebogens aufgerufen und 7 Fragebögen wurden postalisch zurückgeschickt. Für die erste inhaltlich wichtige Frage nach der systematischen Forschungsevaluation liegen gültige Antworten für 87 Fakultäten vor, was einer Rücklaufquote von 67 Prozent entspricht.

Dazu kommen noch Fakultäten, in denen wirtschaftswissenschaftliche Fächer z. B. mit den Naturwissenschaften bzw. der Informatik in einer Fakultät zusammengefasst wurden; sie stellen einen Anteil von 3,6 Prozent. Knapp ein Drittel der Universitäten (31,3 Prozent) ist höchstens 25 Jahre alt, während 37,5 Prozent älter als 50 Jahre sind. Bei den Fakultäten sind 40,4 Prozent höchstens 25 Jahre alt und knapp ein Fünftel der Fakultäten (18,4 Prozent) wurde vor mehr als 50 Jahren gegründet. 54,3 Prozent der Universitäten beschäftigen höchstens 200 Professoren/innen und an 65,7 Prozent sind höchstens 20.000 Studierende eingeschrieben. Demgegenüber liegen an 65,9 Prozent der Fakultäten die Zahl der Professoren/innen bei höchstens 30 und die Anzahl der Studierenden bei 60,7 Prozent der Fakultäten unter 3000 (vgl. Tab. 1). Wie Tabelle 2 zeigt, stammen die Teilnehmenden aus dem gesamten Bundesgebiet, wobei die meisten Antworten auf Nordrhein-Westfalen, Bayern und Niedersachsen entfallen.

Insgesamt verfügen 68,6 Prozent der antwortenden Universitäten und 50,3 Prozent der Fakultäten über ein System zur Evaluation der Forschungsleistung. Mehr als die Hälfte (56,6 Prozent) der philosophischen und 46,8 Prozent der wirtschaftswissenschaftlichen Fakultäten evaluieren regelmäßig die Forschungsleistung ihrer Professorinnen und Professoren, während dies bei knapp drei Vierteln (73,7 Prozent) der rein juristischen Fakultäten nicht der Fall ist. Hinsichtlich des Zeitpunkts der Einführung zeigt sich, dass 14 Universitäten (58,3 Prozent) und 27 Fakultäten (34,2 Prozent) die Forschungsevaluation bereits vor dem Jahr 2005 begonnen haben. Lediglich eine Universität und sechs Fakultäten evaluieren die Forschungsleistung erst seit ein bis zwei Jahren. Insgesamt nimmt die Zahl der Fakultäten, die ein System zur Forschungsevaluation einführen, nur noch degressiv zu.

| Anzahl   | Universität | Phil. Fak. | WiWi | Jura |
|--|-------------|------------|------|------|
|  | 35          | 99         | 47   | 19   |
| <b>Alter in Jahren</b>                           |             |            |      |      |
| bis 26   | 10          | 24         | 19   | 3    |
| 26-50  | 10          | 19         | 20   | 8    |
| mehr als 50                                      | 12          | 7          | 7    | 7    |
| <b>Anzahl Professoren/innen</b>                  |             |            |      |      |
| Universität                                      | Phil. Fak.  | WiWi       | Jura |      |
| bis 50   | bis 20      | bis 20     |      | 6    |
| 51-100   | 21-35       | 21-30      |      | 14   |
| 101-200  | 36-50       | 31-40      |      | 15   |
| 201-300  | 51-65       | 41-50      |      | 11   |
| 301-400  | 66-75       | 51-60      |      | 6    |
| >400   | >75         | >60        |      | 6    |
|  |             |            |      | 29   |
|  |             |            |      | 16   |
|  |             |            |      | 8    |
|  |             |            |      | 3    |
|  |             |            |      | 3    |
|  |             |            |      | 0    |
|  |             |            |      | 0    |
|  |             |            |      | 0    |
| <b>Anzahl Studierende (in Tsd.)</b>              |             |            |      |      |
| Universität                                      | Phil. Fak.  | WiWi       | Jura |      |
| bis 5  |             | bis 1      |      | 10   |
| 5-10   |             | 1-2        |      | 4    |
| 10-20  |             | 2-3        |      | 7    |
| 20-30  |             | 3-4        |      | 11   |
| 30-40  |             | 4-5        |      | 17   |
| >40  |             | >5         |      | 12   |
|  |             |            |      | 7    |
|  |             |            |      | 9    |
|  |             |            |      | 10   |
|  |             |            |      | 8    |
|  |             |            |      | 0    |
|  |             |            |      | 2    |
|  |             |            |      | 2    |
|  |             |            |      | 2    |
| <b>System zur Forschungsevaluation vorhanden</b> |             |            |      |      |
|  |             |            |      | 24   |
|  |             |            |      | 56   |
|  |             |            |      | 22   |
|  |             |            |      | 5    |

Tabelle 1: Charakteristika der Umfrageteilnehmer/innen



| Bundesland             | Anzahl der teilnehmenden Universitäten und Fakultäten |            |           |           | Gesamt     |
|------------------------|---|------------|-----------|-----------|------------|
|                        | Universität   | Phil. Fak. | WiWi      | Jura      |            |
| Baden-Württemberg      | 1   | 5          | 4         | 2         | 12         |
| Bayern                 | 3   | 14         | 8         | 2         | 27         |
| Berlin                 | 3   | 4          | 2         | 1         | 10         |
| Brandenburg            | 2   | 2          | 2         | 2         | 8          |
| Bremen                 | 0   | 1          | 1         | 0         | 2          |
| Hamburg                | 1   | 1          | 2         | 0         | 4          |
| Hessen                 | 1   | 11         | 3         | 2         | 17         |
| Mecklenburg-Vorpommern | 1   | 1          | 0         | 1         | 3          |
| Niedersachsen          | 5   | 6          | 5         | 2         | 18         |
| Nordrhein-Westfalen    | 7   | 7          | 10        | 6         | 30         |
| Rheinland-Pfalz        | 3   | 8          | 2         | 1         | 14         |
| Saarland               | 0   | 1          | 1         | 0         | 2          |
| Sachsen                | 2   | 4          | 2         | 0         | 8          |
| Sachsen-Anhalt         | 2   | 4          | 1         | 0         | 7          |
| Schleswig-Holstein     | 1   | 1          | 1         | 0         | 3          |
| Thüringen              | 1   | 3          | 3         | 0         | 7          |
| keine Angabe           | 2   | 26         | 0         | 0         | 28         |
| <b>Gesamt</b>          | <b>35</b>   | <b>99</b>  | <b>47</b> | <b>19</b> | <b>200</b> |

Tabelle 2: Herkunft der Umfrageteilnehmer/innen

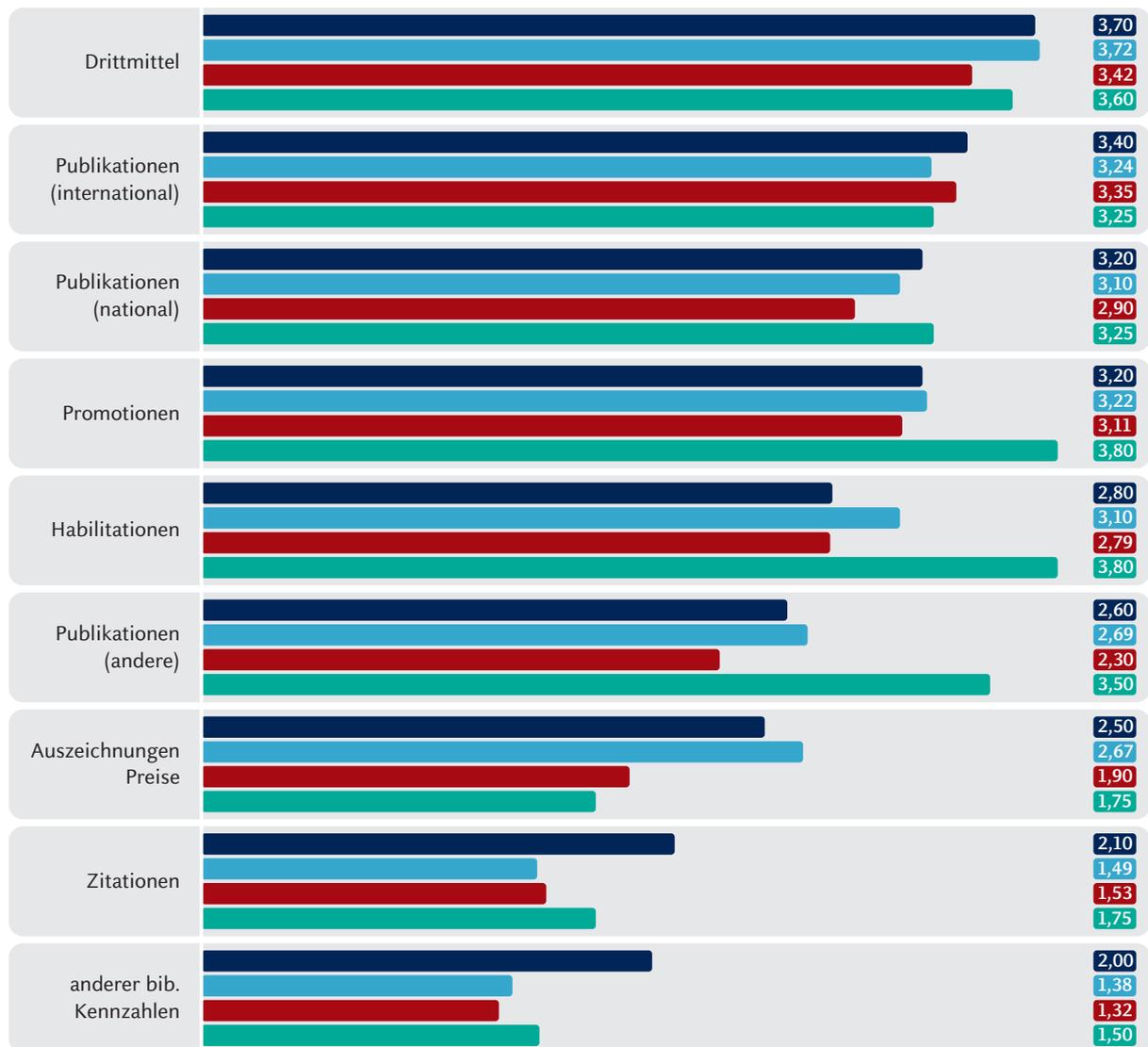
#### 4.2.2. Beschreibung der eingesetzten Kriterien

Während von den Bundesländern insbesondere das Drittmittelvolumen als Indikator für die Forschungsleistung von Universitäten eingesetzt wird, wird sowohl auf universitärer als auch auf fakultärer Ebene eine Vielzahl an Indikatoren verwendet, um die Forschungsleistung der Lehrstühle zu messen (vgl. Tab. 3).

Zwar ist auch auf Ebene der Universitäten die Höhe der eingeworbenen Drittmittel (Mittelwert: 3,7; Skala 1 = keine Bedeutung bis 4 = hohe Bedeutung) das wichtigste Kriterium zur Evaluation der Forschungsleistung, jedoch sind darüber hinaus auch Publikationen in internationalen und nationalen referierten Zeitschriften (Mittelwert: 3,4 bzw. 3,2) sowie die Zahl der Promotionen (Mittelwert: 3,2) wichtige Kriterien. Auffällig ist die geringe Bedeutung von Zitationskennzahlen; weder die Gesamtzahl der Zitationen (Mittelwert: 2,1) noch andere bibliometrische Kennzahlen – wie beispielsweise der h-Index (Mittelwert: 2,0) – sind häufig genutzte Kriterien für die Bestimmung der Forschungsleistung.

Auf Ebene der Fakultäten bietet sich ein ähnliches Bild: An philosophischen Fakultäten stellen Drittmittel, Monografien<sup>12</sup> und Aufsätze in internationalen Zeitschriften die wichtigsten Kriterien dar (Mittelwerte 3, 7, 3, 2 und 3, 2), gefolgt von der Zahl an Promotionen und Habilitationen (Mittelwert 3,2 bzw. 3,1) und Publikationen in nationalen Zeitschriften (Mittelwert 3,1). Die Bedeutung von Zitationen und sonstigen bibliometrischen Kennzahlen ist hier mit Mittelwerten von 1,5 und 1,4 nochmals geringer als auf universitärer Ebene.

<sup>12</sup> Im Fragebogen für die philosophischen Fakultäten wurde explizit nach der Bedeutung von Monografien im Rahmen der Forschungsevaluation gefragt. In den Fragebögen für die wirtschaftswissenschaftlichen und juristischen Fakultäten wurden Monografien über das Kriterium „sonstige Publikationen“ abgedeckt.



Werteskala von 1=keine Bedeutung bis 4=hohe Bedeutung

■ Universitäten     
 ■ Philosophische Fakultäten     
 ■ Wirtschaftswissenschaftliche Fakultäten     
 ■ Juristische Fakultäten

Abbildung 3: Durchschnittliche Bedeutung verschiedener Kriterien der Bewertung der Forschungsleistung

|                               | Universität*  |              | Phil. Fak.*   |              | WiWi*         |              | Jura*         |              |
|-------------------------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|
|                               | keine Bedeut. | hohe Bedeut. |
| Eingeworbene Drittmittel      | 0             | 77           | 0             | 74           | 5             | 58           | 0             | 60           |
| Publikationen (international) | 10            | 62           | 6             | 47           | 10            | 65           | 25            | 75           |
| Publikationen (national)      | 9             | 43           | 6             | 33           | 15            | 40           | 25            | 75           |
| Zahl Promotionen              | 17            | 50           | 2             | 40           | 0             | 42           | 0             | 80           |
| Zahl Habilitationen           | 21            | 29           | 4             | 35           | 16            | 37           | 0             | 80           |
| Publikationen (sonstige)      | 9             | 13           | 6             | 4            | 30            | 15           | 0             | 50           |
| Auszeichnungen/Preise etc.    | 24            | 5            | 13            | 20           | 55            | 10           | 75            | 25           |
| Zitationen gesamt             | 45            | 20           | 60            | 0            | 63            | 5            | 75            | 25           |
| Andere bibl. Kennzahlen       | 43            | 13           | 70            | 0            | 74            | 0            | 75            | 0            |

\* Werte in Prozent

Tabelle 3: Anteil der Teilnehmer, die verschiedenen Kriterien keine oder eine hohe Bedeutung im Rahmen der Forschungsevaluation beimessen



Bei den wirtschaftswissenschaftlichen Fakultäten ergeben sich nur geringe Unterschiede in der Rangfolge der Wichtigkeit. Auch hier stellen Drittmittel und Publikationen in internationalen Zeitschriften die wichtigsten Kriterien dar (Mittelwert je 3,4), allerdings gefolgt von der Zahl der Promotionen (Mittelwert 3,1) sowie Publikationen in nationalen Zeitschriften (Mittelwert 2,9). Zitationen und andere bibliometrische Kennzahlen spielen an wirtschaftswissenschaftlichen Fakultäten ebenfalls kaum eine Rolle (Mittelwert 1,5 bzw. 1,3).

An juristischen Fakultäten bilden die Zahl an Promotionen und die Zahl an Habilitationen die wichtigsten Kriterien zur Forschungsevaluation (Mittelwert je 3,8), gefolgt von Drittmitteln (Mittelwert 3,6), anderen Publikationen (Mittelwert 3,5) und Publikationen in nationalen und internationalen Zeitschriften (Mittelwert je 3,3). Ebenfalls ohne größere Bedeutung sind Zitationen und sonstige bibliometrische Kennzahlen (Mittelwert 1,8 bzw. 1,5).

Insgesamt werden auf universitärer Ebene durchschnittlich 7,0 Kriterien zur Evaluation der Forschungsleistung eingesetzt, von philosophischen Fakultäten 3,7, von wirtschaftswissenschaftlichen Fakultäten 6,0 und von juristischen Fakultäten 6,2.

Da Drittmittel aus unterschiedlichen Quellen stammen können, wurden die Teilnehmer/innen gebeten anzugeben, welche Drittmittelgeber sie im Rahmen ihrer Forschungsevaluation berücksichtigen. Zur Forschungsevaluation auf universitärer Ebene werden am häufigsten Drittmittel der DFG sowie der EU (je 92 Prozent der Universitäten mit System zur Forschungsevaluation) berücksichtigt, gefolgt von Mitteln der Thyssen- bzw. VW-Stiftung sowie vom Bundesministerium für Bildung und Forschung (BMBF) (je 88 Prozent).

Auf Ebene der philosophischen Fakultäten spielen vor allem Mittel der DFG, der EU und des BMBF eine wesentliche Rolle (je 88 Prozent), gefolgt von Mitteln der Thyssen- bzw. VW-Stiftung (85 Prozent). Auch bei wirtschaftswissenschaftlichen Fakultäten werden DFG- und EU- Mittel mit je 90 Prozent am häufigsten verwendet; Mittel des BMBF werden von 85 Prozent der WiWi-Fakultäten herangezogen. Alle juristischen Fakultäten, die über ein System zur Forschungsevaluation verfügen, ziehen Mittel der DFG, der EU sowie von Landesministerien heran. Eine ausführliche Übersicht über die Verbreitung der verschiedenen Drittmittelarten im Rahmen der Forschungsevaluation findet sich in Abbildung 4.

Insgesamt zeigt sich, dass auf Ebene der Universitäten im Mittel 6,9 Drittmittelarten in die Messung der Forschungsleistung einbezogen werden. Von philosophischen Fakultäten werden durchschnittlich 6,1, von wirtschaftswissenschaftlichen Fakultäten 6,4 und von juristischen Fakultäten 7,4 Drittmittelarten verwendet.

Neben der Nichtberücksichtigung von Drittmitteln bestimmter Förderinstitutionen nehmen 32 Prozent der Universitäten, nur 7 Prozent der philosophischen Fakultäten, 25 Prozent der wirtschaftswissenschaftlichen Fakultäten sowie 40 Prozent der juristischen Fakultäten eine qualitative Unterscheidung hinsichtlich der Herkunft von Drittmitteln vor. So werden Drittmittel von öffentlichen Geldgebern oder Stiftungen teilweise höher gewichtet als Industriemittel, womit einer Empfehlung der DFG entsprochen wird (vgl. Deutsche Forschungsgemeinschaft 2004, S. 12). Jedoch zeigen sich auch zwischen den öffentlichen Geldgebern teils deutliche Unterschiede hinsichtlich der Gewichtung der Drittmittel.

Dies kann am Beispiel zweier Fakultäten verdeutlicht werden, die diesbezüglich detaillierte Auskünfte gemacht haben. Die erste Fakultät gewichtet Drittmittel der DFG sowie der VW-Stiftung mit dem Faktor 2,5, sodass ein eingeworbener Euro von diesen Institutionen bei der Forschungsevaluation 2,5 Euro entspricht. Drittmittel der EU, des BMBF sowie von anderen Stiftungen werden mit dem Faktor 1,25 gewichtet, während Industriemittel mit dem Faktor 1 in die Forschungsevaluation eingehen. Auch die zweite Fakultät erkennt

Drittmitteln der DFG einen besonderen Wert zu; sie gehen zu 100 Prozent in die Forschungsevaluation ein, während Drittmittel von allen übrigen öffentlichen und privaten Geldgebern mit lediglich 50 Prozent ihrer Gesamthöhe berücksichtigt werden. Eine andere Fakultät verwendet ein nichtlineares Schema zur Gewichtung der Drittmittel, wonach jeder Lehrstuhl für die ersten 25.000 Euro an eingeworbenen Drittmitteln 215 Euro je 5.000 Euro und für die nächsten 50.000 Euro 125 Euro je 5.000 Euro erhält. Danach verringert sich der Umrechnungsfaktor auf 58 Euro pro 5.000 Euro (für Drittmittel im Bereich von 75.000 Euro bis 375.000 Euro) sowie 12 Euro pro 5.000 Euro (für Drittmittel im Bereich von 375.000 Euro bis 1.000.000 Euro). Ab 1.000.000 Euro eingeworbenen Drittmitteln beträgt der Umrechnungsfaktor nur noch 3 Euro pro 5.000 Euro.

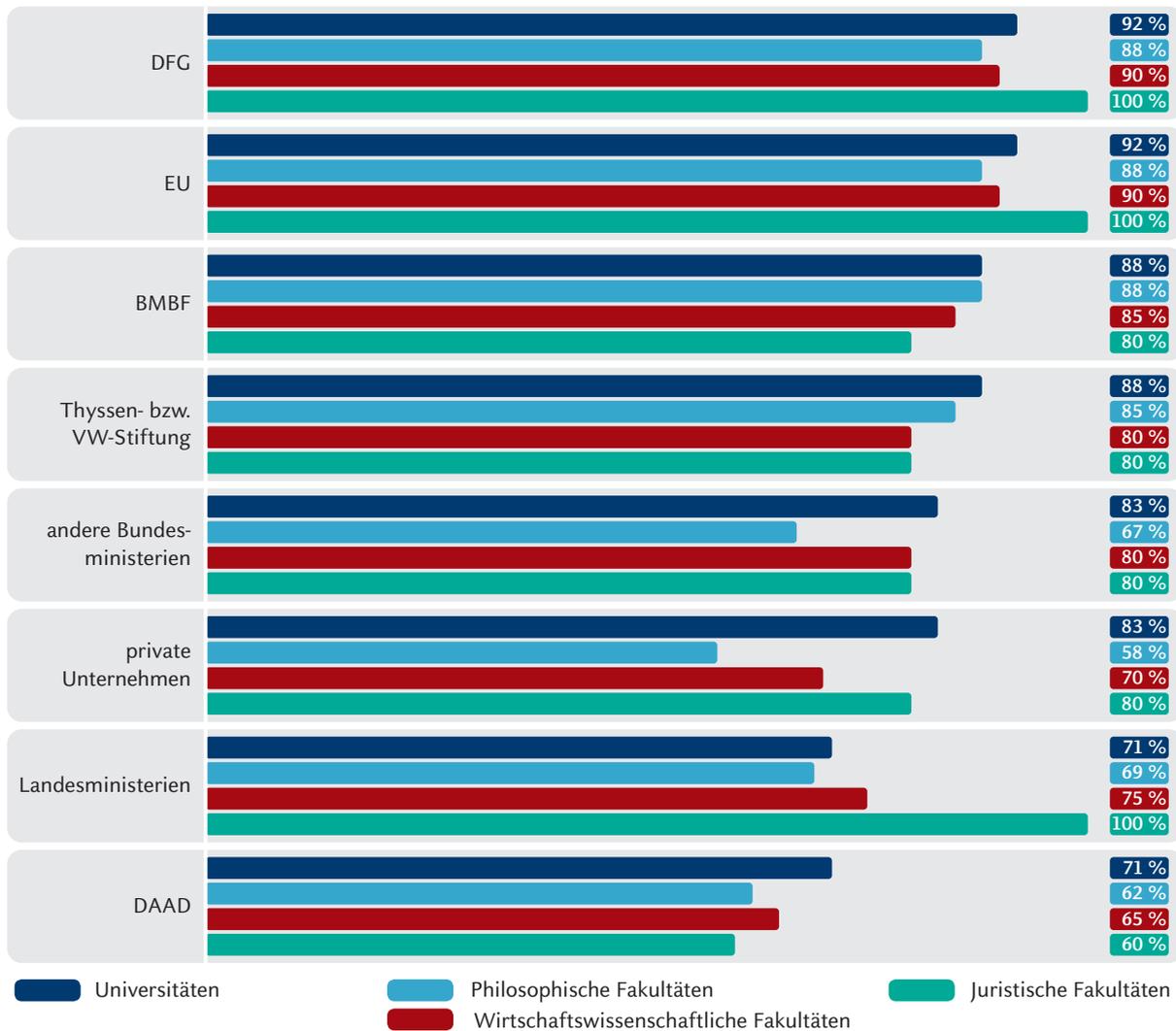


Abbildung 4: Verbreitung verschiedener Drittmittelarten im Rahmen der Forschungsevaluation (in Prozent)

Eine ähnliche Abstufung ist bei Publikationen zu erkennen. Insgesamt verwenden 37,5 Prozent der Universitäten sowie 8,5 Prozent der philosophischen, 57,1 Prozent der wirtschaftswissenschaftlichen und keine der juristischen Fakultäten Rankings, um die Qualität von Zeitschriftenpublikationen zu ermitteln, was ebenfalls einer Empfehlung der DFG entspricht (vgl. Deutsche Forschungsgemeinschaft 2004, S. 15).

Dabei zeigt sich, dass lediglich im Bereich der Wirtschaftswissenschaften allgemein anerkannte, fachspezifische Zeitschriftenrankings existieren. Hierbei handelt es sich zum einen um das vom Verband der Hochschullehrer für Betriebswirtschaftslehre erstellte JOURQUAL-Ranking, das vornehmlich betriebswirtschaftliche Zeitschriften umfasst und von 40,9 Prozent der wirtschaftswissenschaftlichen Fakultäten zur Qualitätsbestimmung von Zeitschriftenaufsätzen herangezogen wird, und zum anderen um das vom „Handelsblatt“



erstellte Zeitschriftenranking, in dem überwiegend volkswirtschaftliche Zeitschriften vertreten sind und auf welches 27,3 Prozent der wirtschaftswissenschaftlichen Fakultäten zurückgreifen.

Die Verwendung von Zeitschriftenrankings hat zur Folge, dass Zeitschriftenartikel bei der Forschungsevaluation mit unterschiedlichen Gewichtungen berücksichtigt werden. Dies wird wiederum am Beispiel zweier Fakultäten deutlich, die das JOURQUAL-Ranking verwenden und hierzu detaillierte Angaben gemacht haben: In der ersten Fakultät werden Artikel, die in einer Zeitschrift der höchsten Kategorie (A<sup>+</sup>) veröffentlicht wurden, mit dem Faktor zehn gewichtet, während Zeitschriften der Kategorie A mit dem Faktor 5, der Kategorie B mit dem Faktor 2,5, der Kategorie C mit dem Faktor 1,5 sowie der Kategorie D mit dem Faktor 0,75 in die Forschungsevaluation eingehen. Bei der zweiten Fakultät finden sich ähnliche Abstufungen; hier werden für Publikationen in A<sup>+</sup> Zeitschriften 100 Punkte vergeben, während Artikel in Zeitschriften der Kategorie A 50 Punkte, der Kategorie B 25 Punkte sowie der Kategorie C und D jeweils 10 Punkte erhalten. Beide Fakultäten differenzieren ferner zwischen Zeitschriftenpublikationen und anderen Publikationen. Während Monografien in der ersten Fakultät mit dem Faktor 4 gewichtet werden und sich dadurch nur unwesentlich von einer Publikation in einer in der Kategorie A gerankten Zeitschrift unterscheiden (Faktor 5), werden sie bei der zweiten Fakultät nicht gewertet. Ähnlich verfahren die Fakultäten mit Beiträgen in Herausgeber/innenbänden; in der ersten Fakultät gehen diese mit dem Faktor 1,5 in die Forschungsevaluation ein, was der Publikation in einer in der Kategorie C gerankten Zeitschrift entspricht, während sie bei der zweiten Fakultät nicht berücksichtigt werden.

Schließlich sollten die Teilnehmer/innen der Umfrage noch angeben, wie die verwendeten Kriterien zur Evaluation der Forschungsleistung miteinander verknüpft werden. Hierbei zeigt sich, dass 39 Prozent der Universitäten die Indikatoren zu einem Forschungsprofil zusammenfassen, während 52 Prozent die Indikatoren zu einem Forschungsscore verdichten.

Weiterhin verwenden 37 Prozent der philosophischen und 6 Prozent der wirtschaftswissenschaftlichen Fakultäten ein Forschungsprofil. Demgegenüber wird ein Forschungsscore von 49 Prozent der philosophischen, 60 Prozent der juristischen und 89 Prozent der wirtschaftswissenschaftlichen Fakultäten eingesetzt.

#### 4.2.3. Verwendung und Auswirkungen der forschungsbezogenen Leistungsevaluation

Wie frühere Studien zeigen, werden Systeme zur forschungsbezogenen Leistungsmessung auf universitärer Ebene insbesondere für die leistungsorientierte Mittelvergabe verwendet (vgl. Heinze 2002, S. 22; Hornbostel 2006, S. 221). Die Ergebnisse der vorliegenden Studie bestätigen dieses Ergebnis, zeigen aber gleichzeitig, dass die Forschungsevaluation an vielen Universitäten ein deutlich breiteres Anwendungsgebiet hat. So verwenden zwar knapp zwei Drittel der Universitäten (67 Prozent) die Forschungsevaluation zur leistungsorientierten Mittelvergabe, ebenso häufig wird sie jedoch auch zur Evaluation von Juniorprofessoren/innen sowie im Rahmen von Bleibeverhandlungen eingesetzt. 67 Prozent der philosophischen Fakultäten verwenden die Forschungsevaluation zur Vergabe interner Boni<sup>13</sup> und 61 Prozent zur leistungsorientierten Mittelvergabe innerhalb der Fakultät. Sowohl bei wirtschaftswissenschaftlichen als auch bei juristischen Fakultäten ist mit 85 Prozent bzw. 100 Prozent die leistungsorientierte Mittelvergabe das Hauptanwendungsgebiet der Forschungsevaluation. Detaillierte Informationen zu den Anwendungsgebieten der Forschungsevaluation auf universitärer und fakultärer Ebene finden sich in Tabelle 4.

<sup>13</sup> Anders als in der Befragung der wirtschaftswissenschaftlichen und juristischen Fakultäten wurde im Fragebogen der philosophischen Fakultäten nach der Verwendung für die Vergabe besonderer Leistungszulagen in der W-Besoldung gefragt, was konzeptuell internen Boni entspricht.

|  | Universität* | Phil. Fak.* | WiWi* | Jura* |
|--|--------------|-------------|-------|-------|
| Vergabe interner Auszeichnungen                      | 17           | 10          | 10    | 0     |
| Vergleich zwischen Fakultäten                        | 38           | -           | -     | -     |
| Vergabe interner Boni                                | 42           | 67          | 10    | 25    |
| Evaluation von Bewerbern/innen                       | 42           | 24          | 30    | 0     |
| Fakultätsinterner Vergleich                          | 42           | 8           | 15    | 0     |
| Berufungsverhandlungen                               | 46           | 31          | 25    | 0     |
| Leistungsorientierte Mittelvergabe (fakultätsintern) | 67           | 61          | 85    | 100   |
| Leistungsorientierte Mittelvergabe (zw. Fakultäten)  | 67           | -           | -     | -     |
| Bleibeverhandlungen                                  | 67           | 45          | 35    | 0     |
| Evaluation von Juniorprofessoren/innen               | 67           | 43          | 45    | 0     |

\* Werte in Prozent

Tabelle 4: Verbreitung verschiedener Einsatzgebiete der Forschungsevaluation

Die Forschungsevaluation wird auf universitärer Ebene für durchschnittlich 5,9 Zwecke eingesetzt und besitzt damit ein deutlich breiteres Anwendungsgebiet als auf fakultärer Ebene. So verwenden philosophische Fakultäten die Forschungsevaluation im Durchschnitt für 2,9 Zwecke, bei wirtschaftswissenschaftlichen und juristischen Fakultäten sind es lediglich 2,6 bzw. 1,3 Zwecke. Allerdings ist dies angesichts des größeren Handlungsspielraums der Rektorate im Vergleich zu den Dekanaten wenig überraschend. Überraschend hingegen ist das Ergebnis, dass lediglich 26 Prozent der Universitäten eine fakultätsspezifisch differenzierte Ausgestaltung des Systems zur Forschungsevaluation vornehmen. Weiterhin geben nur 11 Prozent der philosophischen, 12 Prozent der wirtschaftswissenschaftlichen und 25 Prozent der juristischen Fakultäten an, ihr Konzept zur Forschungsevaluation fakultätsintern spezifisch auf einzelne Bereiche anzupassen.

Hinsichtlich des Volumens, das auf Grundlage der forschungsbezogenen Leistungsmessung an die Lehrstühle verteilt wird, zeigt sich, dass bei 29 Prozent der Universitäten bis zu einer Million Euro und bei 38 Prozent mehr als drei Millionen Euro leistungsorientiert vergeben werden (vgl. Abb. 5).

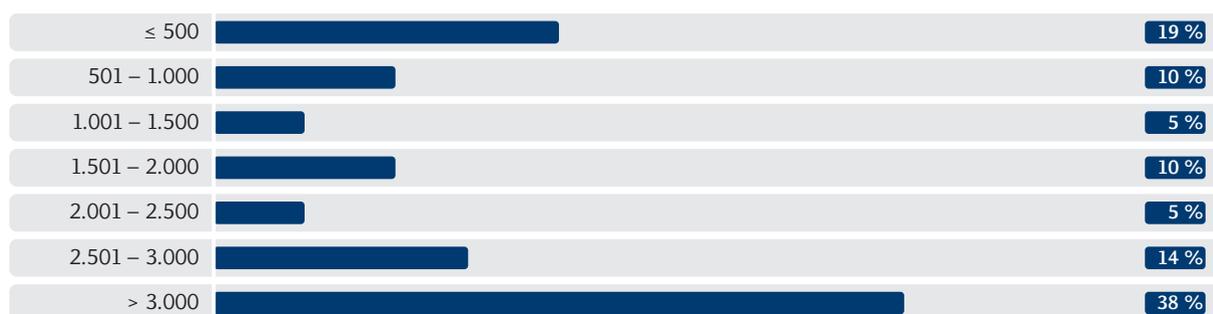


Abbildung 5: Verteilung des Volumens der leistungsorientierten Mittelvergabe auf universitärer Ebene

Bei 42 Prozent der philosophischen Fakultäten liegen leistungsorientiert vergebene Mittel im Bereich bis 50.000 Euro, bei weiteren 42 Prozent liegen sie zwischen 50.000 und 250.000 Euro. Die Grenze von 50.000 Euro wird bei gut einem Drittel (37 Prozent) der wirtschaftswissenschaftlichen Fakultäten mit System zur Forschungsevaluation nicht überschritten, bei 11 Prozent hingegen liegen leistungsorientierten Mittelvergaben bei über einer Million Euro. Bei den juristischen Fakultäten schließlich liegen diese ausschließlich im Bereich zwischen 100.000 und 500.000 Euro (vgl. Abb. 6).



Da die absolute Höhe der verteilten Mittel mit hoher Wahrscheinlichkeit mit der Größe der Universitäten bzw. Fakultäten zusammenhängt, wurde ebenfalls der relative Anteil der leistungsorientiert vergebenen Mittel gemessen am Gesamtbudget erhoben. Bei 68 Prozent der Universitäten beträgt dieser Anteil höchstens 5 Prozent des Gesamtbudgets, während lediglich 5 Prozent der Universitäten mehr als 30 Prozent ihres jährlichen Gesamtbudgets leistungsorientiert vergeben. Demgegenüber vergeben 29 Prozent der philosophischen Fakultäten bis zu 5 Prozent ihres Budgets leistungsorientiert an die Lehrstühle, bei lediglich 5 Prozent sind es mehr als 30 Prozent. Bei etwa einem Fünftel (21 Prozent) der wirtschaftswissenschaftlichen und zwei Dritteln (67 Prozent) der juristischen Fakultäten nehmen leistungsorientiert vergebene Mittel bis zu 5 Prozent des Gesamtbudgets ein. Bei weiteren je 21 Prozent der wirtschaftswissenschaftlichen Fakultäten liegt der Anteil bei 21 bis 30 Prozent bzw. bei über 30 Prozent. Anteile über 10 Prozent werden bei keiner juristischen Fakultät erreicht (vgl. Abb. 7).

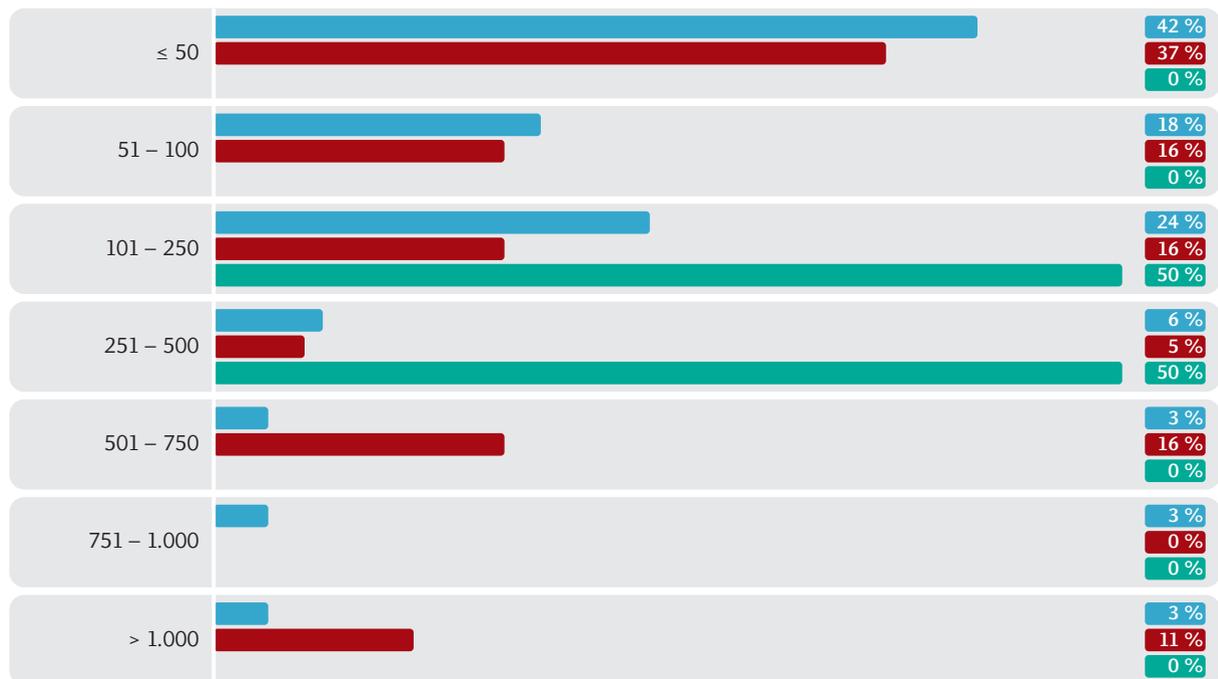


Abbildung 6: Verteilung des Volumens der leistungsorientierten Mittelvergabe auf fakultärer Ebene

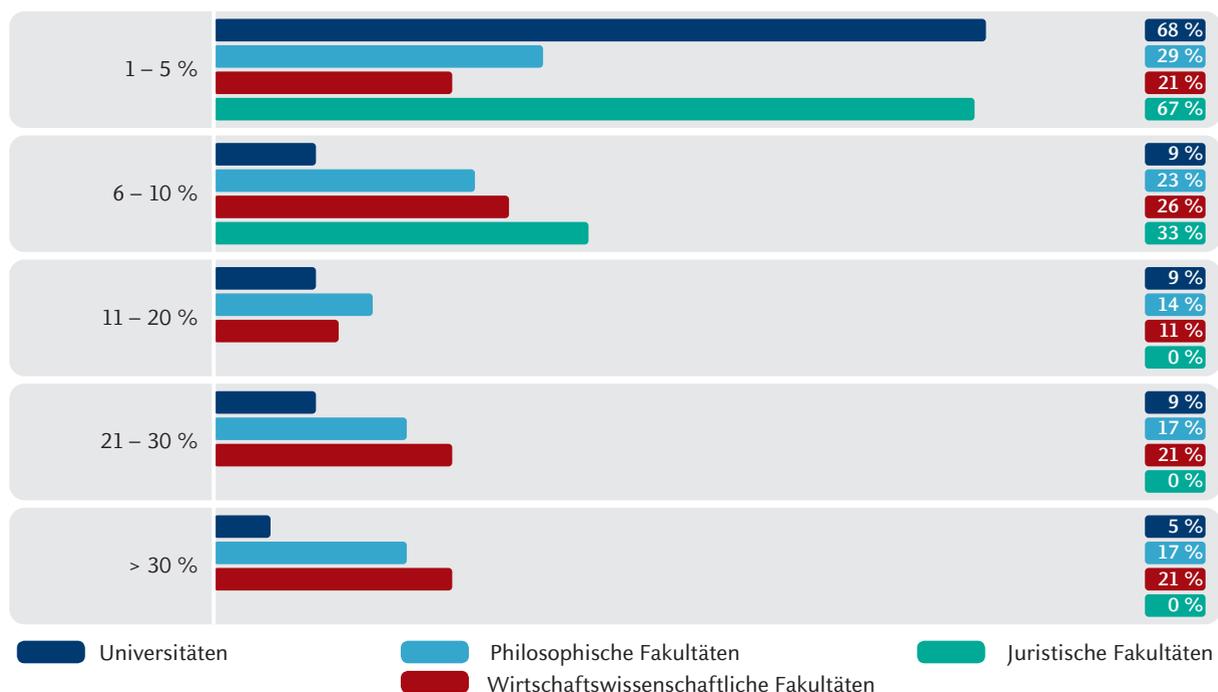


Abbildung 7: Verteilung des Anteils der leistungsorientierten Mittelvergabe am Gesamtbudget

Die Auswirkungen forschungsbezogener Leistungsmessung sehen die befragten Prorektoren/innen und Dekane/innen in der generellen Forschungsorientierung, der Drittmittelinwerbung, den Publikationsaktivitäten und in der Lehre. Den stärksten Einfluss hat die Forschungsevaluation auf die Drittmittellorientierung der Lehrstühle (Mittelwert Universität: 4,0; Mittelwert Fakultäten: 4,0; fünfstufige Skala: 1=deutlich verringert, 5=deutlich erhöht), die generelle Forschungsorientierung (Mittelwert Universität: 3,8; Mittelwert Fakultäten: 3,7) und der Publikationsorientierung (Mittelwert Universität: 3,7; Mittelwert Fakultäten: 3,5). Durchschnittlich keine bedeutenden Auswirkungen hatte die Einführung hingegen auf die Lehrorientierung (Mittelwert jeweils: 3,1); allerdings berichten zwei Universitäten sowie mehrere Fakultäten davon, eine Verschlechterung der Lehrqualität durch die Einführung einer Forschungsevaluation wahrgenommen zu haben. Eine Aufstellung der Antworthäufigkeiten in den einzelnen Kategorien findet sich in Tabelle 5<sup>14</sup>.

| Orientierung  | 2 = verringert |       |      |      | 3 = unverändert |       |      |      | 4 = erhöht |       |      |      | 5 = deutlich erhöht |       |      |      |
|---------------|----------------|-------|------|------|-----------------|-------|------|------|------------|-------|------|------|---------------------|-------|------|------|
|               | Uni            | Phil. | WiWi | Jura | Uni             | Phil. | WiWi | Jura | Uni        | Phil. | WiWi | Jura | Uni                 | Phil. | WiWi | Jura |
| Drittmittel   | 0              | 0     | 0    | 0    | 6               | 11    | 9    | 2    | 11         | 12    | 7    | 2    | 6                   | 21    | 3    | 0    |
| Forschung     | 0              | 1     | 1    | 0    | 7               | 16    | 6    | 4    | 14         | 18    | 9    | 0    | 2                   | 9     | 3    | 0    |
| Publikationen | 0              | 1     | 0    | 0    | 9               | 27    | 7    | 4    | 11         | 13    | 6    | 0    | 3                   | 3     | 6    | 0    |
| Lehre         | 2              | 7     | 1    | 2    | 17              | 26    | 12   | 2    | 4          | 8     | 3    | 0    | 0                   | 1     | 2    | 0    |

Tabelle 5: Auswirkungen der Forschungsevaluation

Schließlich sind die Befragten überwiegend zufrieden mit ihrem System zur forschungsbezogenen Leistungsmessung. Auf einer fünfstufigen Likert-Skala (1=gar nicht zufrieden; 5=sehr zufrieden) wurde von Universitäten durchschnittlich ein Wert von 3,6, von philosophischen Fakultäten 3,3, von wirtschaftswissenschaftlichen Fakultäten 3,5 und von juristischen Fakultäten 2,8 angegeben. Insgesamt 48,0 Prozent der Universitäten, 43 Prozent der philosophischen, 55 Prozent der wirtschaftswissenschaftlichen und lediglich 20 Prozent der juristischen Fakultäten sind „sehr zufrieden“ bzw. „zufrieden“ mit ihrem derzeitigen System.

#### 4.2.4. Hinderungsgründe forschungsbezogener Leistungsmessung

Der Fragebogen richtete sich auch an diejenigen Universitäten und Fakultäten, die derzeit nicht über ein System zur forschungsbezogenen Leistungsmessung verfügen, was insgesamt auf 11 Universitäten und 70 Fakultäten (Phil. Fak.: 31; WiWi: 25; Jura: 14) zutrifft.

Von den Universitäten ohne System zur Forschungsevaluation planen 63,6 Prozent eine Einführung, wobei jedoch bei keiner Universität bereits ein Termin hierfür feststeht. Demgegenüber planen lediglich 48 Prozent der philosophischen, 28 Prozent der wirtschaftswissenschaftlichen und 43 Prozent der juristischen Fakultäten, ein System zur Forschungsevaluation einzuführen. Allerdings steht auch hier in den meisten Fällen noch kein fester Zeitpunkt für die Einführung fest.

<sup>14</sup> Keiner der Befragten gab an, dass sich eine der vier Leistungsdimensionen „deutlich verringert“ habe, weshalb diese Kategorie in Tabelle 5 nicht aufgeführt ist.



Der mit Abstand wichtigste Grund, der bislang die Einführung eines Systems zur Forschungsevaluation verhindert hat, stellt auf universitärer Ebene der Mangel an geeigneten Kriterien zur Messung der Forschungsleistung dar (Mittelwert 3,7; fünfstufige Skala: 1=trifft überhaupt nicht zu, 5=trifft voll und ganz zu). Ein zu hoher mit der Forschungsevaluation verbundener Aufwand (Mittelwert 3,2), die Furcht vor einer zu einseitigen Fokussierung auf die verwendeten Kriterien (Mittelwert 3,0), eine fehlende Notwendigkeit (Mittelwert 2,9) sowie zu starker interner Widerstand (Mittelwert 2,6) scheinen hingegen die Einführung der Forschungsevaluation auf universitärer Ebene kaum zu verhindern.

Als wichtigster Hinderungsgrund auf fakultärer Ebene wird die Befürchtung fehlgeleiteter Anreize durch zu einseitige Fokussierung auf die zur Beurteilung verwendeten Kriterien angeführt (Mittelwert: 4,2), was auch in der Literatur als ein zentraler Kritikpunkt an der Forschungsevaluation diskutiert wird (vgl. Frey 2007, S. 209f.).

Weitere wichtige Hinderungsgründe sind der Mangel an geeigneten Kriterien zur Erfassung der Forschungsleistung (Mittelwert: 3,8), ein zu hoher Aufwand bei der Erfassung der Leistungen (Mittelwert: 3,7) sowie die fehlende Notwendigkeit für ein solches System (Mittelwert: 3,5). Der interne Widerstand innerhalb der Fakultät (Mittelwert: 1,6) hat laut Aussage der Dekane/innen hingegen keinen wesentlichen Einfluss auf die Entscheidung, ein System zur forschungsbezogenen Leistungsmessung nicht einzuführen. In Tabelle 6 sind die Bewertungen der fünf Hinderungsgründe aufgeschlüsselt nach Universitäten und den drei Fakultäten aufgeführt.

|                                    | trifft überhaupt nicht zu* |            |      |      | trifft nicht zu* |            |      |      | teils teils* |            |      |      | trifft zu*  |            |      |      | trifft voll und ganz zu* |            |      |      |
|------------------------------------|----------------------------|------------|------|------|------------------|------------|------|------|--------------|------------|------|------|-------------|------------|------|------|--------------------------|------------|------|------|
|                                    | Universität                | Phil. Fak. | WiWi | Jura | Universität      | Phil. Fak. | WiWi | Jura | Universität  | Phil. Fak. | WiWi | Jura | Universität | Phil. Fak. | WiWi | Jura | Universität              | Phil. Fak. | WiWi | Jura |
| keine geeigneten Kriterien         | 0                          | 7          | 21   | 0    | 0                | 3          | 8    | 7    | 0            | 0          | 8    | 0    | 0           | 5          | 33   | 1    | 0                        | 6          | 29   | 1    |
| zu hoher Aufwand                   | 0                          | 8          | 14   | 0    | 0                | 1          | 14   | 7    | 0            | 0          | 0    | 0    | 0           | 5          | 45   | 7    | 0                        | 6          | 27   | 5    |
| zu einseitige Fokussierung         | 0                          | 4          | 9    | 0    | 4                | 7          | 9    | 0    | 2            | 0          | 5    | 0    | 2           | 8          | 36   | 6    | 1                        | 2          | 41   | 4    |
| keine Notwendigkeit                | 0                          | 2          | 0    | 0    | 0                | 0          | 39   | 8    | 0            | 0          | 4    | 5    | 0           | 2          | 26   | 5    | 0                        | 6          | 30   | 2    |
| am internen Widerstand gescheitert | 0                          | 1          | 71   | 7    | 0                | 7          | 19   | 2    | 0            | 0          | 0    | 1    | 0           | 7          | 10   | 0    | 0                        | 4          | 0    | 0    |

\* Werte in Prozent

Tabelle 6: Übersicht über die Gründe, die die Einführung eines Systems zur Forschungsevaluation verhindern

### 4.3 Zusammenfassung der Ergebnisse und Schlussfolgerungen

Die Befragung hat gezeigt, dass eine Bewertung der Forschungsleistung sowohl auf universitärer Ebene als auch auf fakultärer Ebene vielfach durchgeführt wird. Allerdings können zwischen den Fächern große Unterschiede festgestellt werden. Während in etwa der Hälfte der philosophischen und wirtschaftswissenschaftlichen Fakultäten eine Leistungsmessung durchgeführt wird, ist dies nur für eine Minderheit der rechtswissenschaftlichen Fakultäten der Fall.

Die Ergebnisse zeigen auch, dass die Leistungsbewertung auf Grundlage von mehreren Indikatoren durchgeführt wird. Idealtypisch lässt sich sagen, dass vor allem die eingeworbenen Drittmittel, Publikationen und die Förderung des wissenschaftlichen Nachwuchses als Kriterien zur Messung der Forschungsleistung dienen. Andere Kriterien (insbesondere Zitationsanalysen) spielen in allen Fakultäten sowie auf Universitätsebene hingegen fast keine Rolle.

In der Rangfolge der Kriterien zeigen sich leichte Unterschiede zwischen den Fächern. Während an philosophischen und wirtschaftswissenschaftlichen Fakultäten Drittmittel und Publikationen in internationalen Zeitschriften die beiden wichtigsten Kriterien sind, hat die Zahl der Promotionen und Habilitationen an juristischen Fakultäten die größte Bedeutung für die Forschungsbewertung. Interessant ist auch, dass die Mehrheit der Fakultäten im Rahmen ihrer Forschungsevaluation nicht nur kompetitive Drittmittel berücksichtigt, sondern auch Mittel aus anderen Quellen wie zum Beispiel von privaten Unternehmen.

Eine Gewichtung der verschiedenen Drittmittelarten findet nur an sehr wenigen, überwiegend wirtschaftswissenschaftlichen, Fakultäten statt. Auch findet fast nur an wirtschaftswissenschaftlichen Fakultäten eine Qualitätsgewichtung von Zeitschriftenpublikationen anhand von Zeitschriftenrankings statt. Im Hinblick auf die Tatsache, dass fachbereichsweit anerkannte Zeitschriftenrankings wie das JOURQUAL- oder Handelsblatt-Ranking lediglich in den Wirtschaftswissenschaften existieren, ist dies jedoch nicht überraschend.

Weiterhin zeigen unsere Ergebnisse, dass die wenigsten Fakultäten angeben, ihr System zur Forschungsevaluation spezifisch auf einzelne Teilbereiche der Fakultäten anzupassen. Während dies für relativ homogene Fakultäten wie den Wirtschaftswissenschaften oder den Rechtswissenschaften wenig verwundert, überrascht dies umso mehr im Falle der philosophischen Fakultäten. Wie die Zusammensetzung der Stichprobe zeigt, sind alle philosophischen Fakultäten fachlich stark differenziert (nicht tabellarisch ausgewiesen). Die Einführung der Leistungsbewertung scheitert folglich nicht an der Heterogenität der Fächerkulturen. Dennoch erfolgt die Leistungsbewertung weitgehend undifferenziert, ohne Besonderheiten der Fächer in den Indikatoren abzubilden. Dass dies nicht zu großer Unzufriedenheit mit den vorhandenen Systemen der Leistungsbewertung führt, spricht dafür, dass in der Praxis keine fundamentalen Probleme auftreten, in deren Folge die Leistungsbewertung insgesamt in Frage gestellt wird. Möglicherweise sind die Fakultäten mit den Systemen der Leistungsbewertung auch deshalb nicht unzufrieden, weil das Verteilungsvolumen im Rahmen der leistungsorientierten Mittelvergabe regelmäßig gering ausfällt.

Eingesetzt wird die Bewertung der Forschungsleistung in erster Linie für die leistungsorientierte Mittelvergabe (zwischen den Fakultäten auf universitärer Ebene, innerhalb der Fakultäten auf fakultärer Ebene), wobei gerade auf die Fakultäten bezogen nur ein geringer Anteil der Mittel leistungsbezogen vergeben wird (überwiegend unter 20 Prozent). An den philosophischen Fakultäten werden auch besondere Leistungszulagen auf Grundlage der Leistungsbewertung vergeben.

Insgesamt kann festgestellt werden, dass der überwiegende Teil der Fakultäten mit ihrem System zur forschungsbezogenen Leistungsmessung zufrieden ist. Die wichtigsten Gründe gegen die Einführung eines



solchen Systems sind der große Aufwand, der mit einem solchen System verbunden ist, und das Fehlen geeigneter Indikatoren zur Messung der Forschungsleistung. Aber auch eine einseitige Ausrichtung der wissenschaftlichen Arbeit an den Leistungsindikatoren wird befürchtet, obwohl die Einschätzung der Effekte der Leistungsmessung zeigt, dass zwar die Drittmittelorientierung zugenommen habe, ohne dass aber eine systematische Verringerung der Lehr- oder generellen Forschungsorientierung festzustellen sei.

Aus diesen Ergebnissen können vier zentrale Schlussfolgerungen für die in diesem Papier zu definierenden Indikatoren zur forschungsbezogenen Leistungsmessung an der Heinrich-Heine-Universität gezogen werden:

Erstens bestätigen die Ergebnisse die Überlegungen hinsichtlich einer multikriteriellen Messung der Forschungsleistung. Ein einzelner Indikator ist nicht ausreichend, um Forschungsleistung vollumfänglich abzubilden, da diese ein komplexes Konstrukt ist. Jeder Indikator für sich bildet nur einen Teil dessen ab, was als „Forschungsleistung“ bezeichnet werden kann. Dies lässt den Schluss zu, dass nur der gleichzeitige Einsatz mehrerer Indikatoren der Komplexität des Konstrukts „Forschungsleistung“ zumindest annähernd gerecht werden kann, was im Übrigen auch verfassungsrechtlich geboten scheint (vgl. Ausführungen in Kapitel 3).

Zweitens sollte die Forschungsleistung im Wesentlichen anhand der drei Indikatoren *Nachwuchsförderung*, *eingeworbene Drittmittel* sowie *Publikationen* evaluiert werden. Allerdings sollte berücksichtigt werden, dass sich die relative Bedeutung dieser drei Kriterien zwischen den philosophischen, wirtschaftswissenschaftlichen und juristischen Fakultäten zum Teil deutlich unterscheidet. Diese Kriterien sollten somit fachspezifisch implementiert werden.

Forschungsleistung sollte durch drei Indikatoren gemessen werden



Abbildung 8: Übersicht über die Indikatoren, die zur Forschungsevaluation an der HHU eingesetzt werden sollten

Drittens zeigt sich, dass die Mehrheit der Fakultäten nicht nur Drittmittel der DFG und der EU, sondern auch Drittmittel anderer Quellen wie zum Beispiel von privaten Unternehmen oder vom DAAD im Rahmen der Forschungsevaluation berücksichtigt, woraus sich schließen lässt, dass im Rahmen der Forschungsevaluation an der HHU auch eine Vielzahl unterschiedlicher Drittmittelgeber berücksichtigt werden sollte.

Viertens sollte bei der Verwendung von Publikationen zur Messung der Forschungsleistung beachtet werden, dass die Bedeutung unterschiedlicher Publikationsformen in den drei Fakultäten sehr unterschiedlich ist. Diesen Unterschieden sollte im Rahmen der Forschungsevaluation Rechnung getragen werden.

## 5. INDIKATOREN ZUR MESSUNG DER FORSCHUNGSLEISTUNG

Im Rahmen der Forschungsevaluation wird überwiegend auf einen Verbund vier zentraler Kriterien zurückgegriffen. (vgl. Dyckhoff et al. 2005; Fiedler et al. 2006; Hornbostel/Heise 2006; Jansen et al. 2007; Rassenhövel 2010; Wissenschaftsrat 2007, 2008). Hierbei handelt es sich um (1) die Forschungsevaluation anhand von Dissertationen und Habilitationen, (2) Forschungsevaluation anhand von Drittmittelinwerbungen, (3) Forschungsevaluation anhand von Publikationen sowie (4) die Forschungsevaluation anhand eines *Informed-Peer-Reviews*. Eine detaillierte Beschreibung dieser vier Indikatoren erfolgt in den nachfolgenden Kapiteln.

### 5.1. Forschungsevaluation anhand von Dissertationen und Habilitationen

Ein häufig genutztes Kriterium zur Messung der Forschungsleistung ist die Anzahl der Dissertationen und Habilitationen (vgl. Jaeger/In der Smitten 2009, S. 13; Jaeger/In der Smitten 2010, S. 10), was sich sowohl auf interne als auch auf externe Dissertationen und Habilitationen bezieht. Durch dieses Kriterium werden insbesondere die Forschungsleistungen des wissenschaftlichen Nachwuchses gemessen. Da der wissenschaftliche Nachwuchs häufig keine Drittmittel einwirbt und zum Teil neben der Dissertation keine eigenen Publikationen in Form von z.B. Zeitschriftenartikeln veröffentlicht, wird dessen Forschungsleistung durch Kennzahlen wie beispielsweise Publikationen und Drittmittelinwerbungen nur bedingt abgebildet. Dies könnte sich zwar durch die zunehmende Verbreitung der kumulativen Dissertation bzw. Habilitation ändern, jedoch sind auch hier fachspezifische Unterschiede zu konstatieren.

Ein spezifischer Vorteil der Nutzung dieses Indikators ist, dass dadurch die forschungsbezogene Betreuungsleistung von Professoren/innen abgebildet werden kann. Da nicht in allen Fällen die betreuenden Professoren/innen auch als Koautoren/innen der Publikationen auftreten (dies gilt insbesondere bei monografischen Dissertationen und Habilitationen), wird die Betreuungsleistung über die Publikationsaktivität eines/r Professors/Professorin nur unzureichend erfasst. Promotionen und Habilitationen bieten hier die Möglichkeit, die abstrakte Betreuungsleistung zumindest näherungsweise zu messen; dies ist eine Erklärung, weshalb dieser Indikator relativ häufig zur Leistungsevaluation herangezogen wird (vgl. Kapitel 4).

Für jeden Indikator sind verschiedene Varianten zur Umsetzung denkbar



Abbildung 9: Übersicht über Verfahren zur Verwendung der Nachwuchsförderung als Indikator der Forschungsleistung



Grundsätzlich kann die Nachwuchsförderung auf zwei Arten für die Messung der Forschungsleistung eines Professors/einer Professorin, Instituts oder einer Fakultät berücksichtigt werden. Zum einen kann schlicht die Anzahl der Promotionen bzw. Habilitationen pro Professor/in gezählt werden, zum anderen kann die Anzahl der Promotionen bzw. Habilitationen pro Kopf relativ zum fachspezifischen Bundesdurchschnitt (Promotions- bzw. Habilitationsquote) herangezogen werden.

#### 5.1.1. Anzahl der Promotionen und Habilitationen pro Kopf

Ein einfacher Indikator ist die Anzahl der Promotionen und Habilitationen pro Kopf. Dazu reicht es, in den zuständigen Prüfungsämtern die Daten bezüglich erfolgreicher Promotionen und Habilitationen zu erfassen. Soll die Leistung eines einzelnen Professors/einer einzelnen Professorin beurteilt werden, so werden lediglich die von ihm betreuten Promotionen/Habilitationen berücksichtigt. Handelt es sich bei der Bewertungseinheit um ein Institut oder eine Fakultät, so wird die Zahl der Promotionen/Habilitationen durch die Zahl der dort arbeitenden Professoren/innen geteilt.

Jedoch muss darauf hingewiesen werden, dass eine Habilitation keine notwendige Voraussetzung für die Berufung auf eine Professur darstellt. So ist beispielsweise eine erfolgreiche Evaluation eines Junior-Professors/einer Junior-Professorin eine habilitationsäquivalente Leistung. Darüber hinaus können beispielsweise Post-Docs auch ohne eine formale Habilitation auf eine Professur berufen werden, sofern sie ebenfalls habilitationsäquivalente Leistungen, zum Beispiel in Form von Publikationen, nachweisen. Hier ist ein Ruf auf eine W2- bzw. W3-Professur als habilitationsäquivalent zu bezeichnen und sollte somit auch berücksichtigt werden.

Allerdings hat der Indikator auch spezifische Nachteile. Beispielsweise können keine Aussagen über die Qualität der Dissertation bzw. der Habilitation getroffen werden. Eine Berücksichtigung der Note der Abschlussarbeit könnte Fehlanreize auf Seiten der Betreuer/innen setzen, so dass diese Möglichkeit nicht genutzt werden sollte. Darüber hinaus ist insbesondere bei Habilitationen die Frage, wie groß die Betreuungsleistung in diesem Fall ist. Dies müsste fachspezifisch ermittelt werden, da auch hier teils signifikante Unterschiede auftreten können. Ein weiterer Nachteil ist, dass Promotionen und Habilitationen in verschiedenen Fächern unterschiedliche Bedeutung haben. Insbesondere die durchschnittliche Anzahl von Qualifikationsarbeiten unterscheidet sich nach Fächern erheblich. Das schränkt die Vergleichbarkeit des Indikators über Fächergrenzen hinweg ein.

#### 5.1.2. Anzahl der Promotionen und Habilitationen pro Kopf relativ zum Bundesdurchschnitt

Neben der Möglichkeit, die Anzahl an Promotionen und Habilitationen pro Kopf zu messen, kann auch eine Relativierung durch den Vergleich zum Bundesdurchschnitt vorgenommen werden. Dieses Vorgehen hat den Vorteil, dass die Vergleichbarkeit des Leistungsindikators erhöht wird, weil die Leistung eines Forschers/einer Forscherin mit der anderer Forscher/innen des gleichen Fachbereichs verglichen wird. Dadurch können Verzerrungen, die durch unterschiedliche Fachkulturen verursacht werden könnten, vermieden werden.

Hierzu kann die vom Statistischen Bundesamt als Standardindikator der Hochschulstatistik berichtete sogenannte „Promotionsquote“ herangezogen werden, d.h. die Anzahl von Promotionen pro Professor/in pro Jahr, aufgeschlüsselt nach Fächergruppen der Fächersystematik des Statistischen Bundesamtes (Statistisches Bundesamt, 2011, S. 24). Analog kann aus der Habilitationsstatistik des Statistischen Bundesamtes die Habilitationsquote für die einzelnen Fächer berechnet werden.

Die Anzahl der von einem Hochschullehrer/einer Hochschullehrerin betreuten Promotionen im Jahr kann durch die Promotionsquote geteilt werden. Der Indikator zeigt die Leistung in der Nachwuchsförderung im Verhältnis zum Fachgruppendurchschnitt an. Ein Wert größer 1 würde bedeuten, dass ein Professor/eine Professorin mehr Dissertationen bzw. Habilitationen erfolgreich betreut hat als ein durchschnittlicher Hochschullehrer/eine durchschnittliche Hochschullehrerin seines/ihres Fachbereichs, während ein Wert kleiner 1 eine unterdurchschnittliche Quote implizieren würde.

## 5.2. Forschungsevaluation anhand von Drittmittleinwerbungen

Drittmittleinwerbungen stellen einen weiteren Indikator zur Evaluation der Forschungsleistung dar. Nach dem Gesetz über die Hochschulen des Landes Nordrhein-Westfalen (§ 71 Abs. 1 S. 1 HG NRW; vgl. insofern auch § 25 Abs. 1 S. 1 HRG) kann von Drittmitteln gesprochen werden, wenn Forschungsvorhaben „nicht aus den der Hochschule zur Verfügung stehenden Haushaltsmitteln, sondern aus Mitteln Dritter finanziert werden.“ Sie stammen beispielsweise von der DFG, den Bundes- und Landesministerien, Stiftungen oder aus der Industrie. Dabei sind Drittmittel, anders als Publikationen und Zitationen, eine Input- und keine Outputgröße (vgl. Rassenhövel/Dyckhoff 2006, S. 95; Jansen et al. 2007, S. 125). Jedoch werden Drittmittleinwerbungen zum Teil auch als Ergebnis der Forschungsleistung aufgefasst, da Institutionen bei der Vergabe von Drittmitteln projektbezogene Vorarbeiten berücksichtigen bzw. Vorarbeiten einen großen Einfluss auf den Erfolg eines Projektantrags nehmen (vgl. Rassenhövel/Dyckhoff 2006, S. 96).

Bundesländer nutzen die von einer Hochschule eingeworbenen Drittmittelvolumina als Indikator für die Forschungsleistung (vgl. Jaeger 2008, S. 92), um auf dieser Grundlage leistungsorientierte Mittel an die Hochschulen zu vergeben (vgl. Jaeger/In der Smitten 2010). Auch in die Hochschulrankings des CHE und des FOCUS sowie in das Förderranking der DFG gehen die eingeworbenen Drittmittel als ein wichtiger Indikator zur Ermittlung der Forschungsleistung ein (vgl. Rassenhövel/Dyckhoff 2006, S. 93).

Jedoch unterliegt die Nutzung von Drittmitteln zur Evaluation der Forschungsleistung Grenzen. So fördert eine Messung der Leistung am Input (eingeworbene Drittmittel) tendenziell ineffizientes Arbeiten. Beispielsweise sind Forscher/innen bzw. Institute, die besonders effizient arbeiten und deshalb bei ihrer Arbeit kaum Drittmittel benötigen, bei der Nutzung von Drittmittelindikatoren benachteiligt (vgl. Kieser 1998, S. 216; Rassenhövel 2010, S. 136). Gleiches gilt für Disziplinen (insbesondere die Geistes- und Sozialwissenschaften), in denen die Forschung kaum auf (kostspielige) technische Geräte angewiesen ist (vgl. Kieser 1998, S. 216; Wissenschaftsrat 2010).<sup>15</sup> Ergebnisse empirischer Studien stellen vor diesem Hintergrund die Nutzung von Drittmitteln zur Messung der Forschungsleistung generell in Frage.

Jansen et al. (2007) finden in einer Untersuchung von drei Forschungsdisziplinen (Mikroökonomie, Astrophysik und Nanotechnologie) einen umgekehrt U-förmigen Zusammenhang zwischen Drittmitteln und Publikationen: Mit zunehmenden Drittmitteln steigt anfangs die Zahl der Publikationen; ab einem kritischen Schwellenwert reduzieren zusätzliche Drittmittel jedoch den Forschungsoutput (vgl. Jansen et al. 2007, S. 144). Die Autoren vermuten, dass dieser Effekt auf ein überoptimales Wachstum der Arbeitsgruppen zurückzuführen ist, was zu steigenden intraorganisationalen Transaktionskosten führt und infolgedessen den Publikationsoutput verringert (vgl. Jansen et al. 2007, S. 137).

<sup>15</sup> In der Terminologie der Expertenkommission des Wissenschaftsrates handelt es sich um eine Form der „Forschungsermöglichung“ bzw. ein „Aktivitätsmaß“.



Jansen et al. (2007) können jedoch auch zeigen, dass der prognostizierte Scheitelpunkt für alle drei untersuchten Fächer über dem durchschnittlichen Drittmittelaufkommen des jeweiligen Faches liegt. Verallgemeinert man diesen Befund, ist ein negativer Effekt der Einwerbung von Drittmitteln erst bei deutlich überdurchschnittlichem Drittmittelaufkommen zu erwarten. Um für den Fall vorzubeugen, dass der Anteil Drittmittel erheblich zunimmt und damit die Gefahr einer Effizienzreduktion eintritt, kann eine Kappung bei der Gratifikation von Drittmittelinwerbungen vorgesehen werden.

Trotz der Einwände kommt Drittmittelinwerbungen zur Evaluation der Forschungsleistung in der aktuellen Evaluationspraxis eine hohe Bedeutung zu. Sie sind sowohl in einflussreichen Rankings als auch bei der leistungsorientierten Mittelvergabe auf Länderebene ein wesentliches Kriterium zur Bestimmung der Forschungsleistung (vgl. Jaeger/In der Smitten 2010, S. 11; Borgwardt 2011; Hornbostel 1997, S. 1997; Meier/Schimank 2004, S. 110), was auch aus den Ergebnissen der Befragung hervorgeht (siehe Kapitel 4).

Für jeden Indikator sind verschiedene Varianten zur Umsetzung denkbar

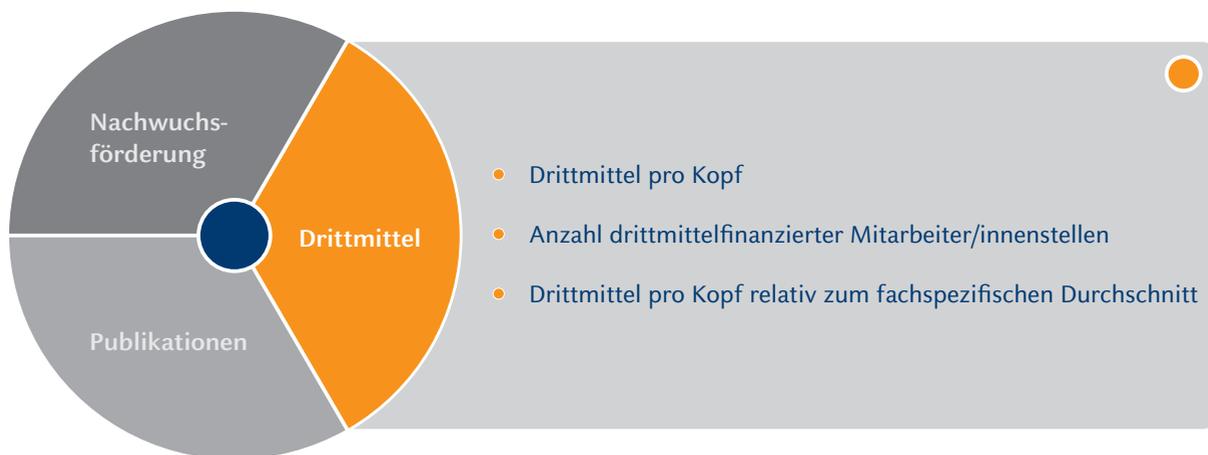


Abbildung 10: Übersicht über Verfahren zur Verwendung von Drittmitteln als Indikator der Forschungsleistung

Grundsätzlich lassen sich drei unterschiedliche Vorgehensweisen zur Nutzung von Drittmittelinwerbungen zur Forschungsevaluation unterscheiden. Erstens können die eingeworbenen Drittmittel pro Kopf zur Evaluation der Forschungsleistung herangezogen werden. Zweitens können die mit den eingeworbenen Drittmitteln finanzierten vollzeitäquivalenten Mitarbeiter/innenstellen betrachtet werden und drittens können die eingeworbenen Drittmittel pro Kopf relativ zum fachspezifischen Bundesdurchschnitt einfließen.

### 5.2.1. Drittmittel pro Kopf

Der technisch einfachste Weg, Drittmittel als Indikator der Forschungsleistung zu verwenden, ist, das absolute Aufkommen von Drittmitteln pro Kopf zu bestimmen. Bei diesem Verfahren stellt sich lediglich die Frage, ob die Drittmittel pro Professor/in bestimmt werden oder auch die Anzahl der wissenschaftlichen Mitarbeiter/innen in der Größenrelativierung berücksichtigt wird.

Werden die Drittmittel pro Professor/in zur Leistungsmessung verwendet, werden nur solche Drittmittel in die Leistungsevaluation einbezogen, an deren Antragstellung der/die Professor/in beteiligt ist.

Sollen darüber hinaus auch Drittmittel berücksichtigt werden, an deren Antragstellung der/die Professor/in nicht beteiligt war, so ist es ggf. angemessener, eine Größenrelativierung anhand der Zahl der Wissenschaftler/innen durchzuführen. Dies gilt umso mehr, wenn sich die Anzahl der wissenschaftlichen Mitarbeiter/innen in den Forschungseinheiten, die einzelnen Professoren/innen zugeordnet werden, stark unterscheidet. Auf diese Weise können auch Drittmittel, die Wissenschaftler/innen ohne Mitwirkung des Lehrstuhlinhabers/der Lehrstuhlinhaberin eingeworben haben, in der Leistungsevaluation berücksichtigt werden.

Der Einsatz der Drittmittel pro Kopf als Indikator zur Messung der Forschungsleistung ist jedoch problematisch, wenn Unterschiede im Drittmittelaufkommen nicht von der Leistung abhängen, sondern auf strukturelle Unterschiede zwischen Fächern oder Forschungsfeldern zurückgehen. Ein Beispiel ist der Bedarf an Drittmitteln für exzellente Forschung, der in den verschiedenen Fächern sehr unterschiedlich ausfällt. So unterscheiden sich die durchschnittlichen Drittmittelaufkommen pro Kopf je nach Fächern zum Teil erheblich (siehe Anhang 1). Insbesondere in den Geistes- und Rechtswissenschaften (aber auch Mathematik und theoretischer Physik) lässt sich exzellente Forschung auch ohne Drittmittel bewerkstelligen (Wissenschaftsrat 2010), so dass diese Forscher für ihre Arbeit zum Teil nicht auf Drittmittel angewiesen sind.

### 5.2.2. Drittmittelfinanzierte Mitarbeiter/innenstellen

Wie im vorangegangenen Abschnitt beschrieben, bestehen zwischen einzelnen Fächern zum Teil erhebliche Unterschiede in den eingeworbenen Drittmittelsummen. Ein Grund hierfür ist, dass die Aufwendungen für Sachmittel und Infrastruktur in technischen und naturwissenschaftlichen Fächern deutlich höher ausfallen als in den Geisteswissenschaften, während in den Sozialwissenschaften teilweise erhebliche Kosten für große Datenerhebungsprojekte (wie zum Beispiel repräsentative Bevölkerungsumfragen) anfallen. Um Unterschiede im Sachmittelbedarf zu korrigieren, kann das Drittmittelaufkommen in Mitarbeiter/innenstellen ausgedrückt werden. So wird aus dem Gesamtaufkommen nur der Anteil der Drittmittel zur Beurteilung der Forschungsleistung herangezogen, der effektiv für die Bearbeitung der Projekte (Auswertung von Daten, Publikation etc.) verwendet wird. Als Maßzahl werden vollzeitäquivalente Mitarbeiter/innenstellen (TV-L 13 oder höher) verwendet.

### 5.2.3. Drittmittel pro Kopf relativ zum fachspezifischen Bundesdurchschnitt

Wenn Leistungsindikatoren auch den Vergleich über Fächergrenzen hinweg erlauben sollen, muss die fachspezifische Bedeutung von Drittmitteln für den Forschungsprozess in die Umsetzung der Indikatoren einfließen, was in der Erhebung über drittmittelfinanzierte Mitarbeiter/innenstellen nur ansatzweise geschieht. In diesem Fall ist es nicht sinnvoll, das Drittmittelaufkommen pro Kopf zu verwenden (Meier/Schimank 2004), sondern es sollte eine fachspezifische Einbettung der eingeworbenen Drittmittelvolumen vorgenommen werden. Dazu müssen sinnvolle Verfahren der Standardisierung des Drittmittelaufkommens über Fächer hinweg angewendet werden. Diese Standardisierung des Drittmittelaufkommens kann anhand eines fachspezifischen größenrelativierten Drittmittelaufkommens erfolgen. Bei diesem Verfahren werden die Drittmittel, die ein/e Professor/in (und ggf. seine Mitarbeiter/innen) einwirbt, durch das durchschnittliche Drittmittelaufkommen in dem betreffenden Fachgebiet geteilt. Daraus ergibt sich eine Kennzahl, die angibt, um wie viel Prozent der Drittmittelerfolg vom Durchschnittswert im jeweiligen Fachbereich abweicht. Werte größer 1 zeigen überdurchschnittlichen Erfolg an, Werte kleiner 1 demgegenüber unterdurchschnittlichen Erfolg.

Die Schwierigkeit dieses Verfahrens liegt darin, eine adäquate Basis für die Größenrelativierung zu bestimmen. Daten zum durchschnittlichen Drittmittelaufkommen von Professoren/innen werden vom statistischen



Bundesamt (DESTATIS) bereitgestellt (siehe die tabellarische Übersicht in Anhang 1). DESTATIS erfasst alle Formen von Drittmitteln (DFG, andere öffentliche Forschungsförderer, private Stiftungen, Unternehmen). Die Daten sind auf Ebene der Fächer verfügbar. Nur Fächer mit weniger als 100 Professoren/innen werden in der Aufstellung nicht berücksichtigt. An der HHU betrifft dies in erster Linie das Institut für Modernes Japan und das Institut für Jüdische Studien an der Philosophischen Fakultät, für die kein eigener Fächerdurchschnitt ermittelt werden kann. In diesen Fällen kann eine Orientierung an verwandten Fächern erfolgen oder die einzelnen Professoren/innen der genannten Fächer ordnen sich den Fächern zu, denen ihre Forschung entspricht.<sup>16</sup> Sowohl für die übrigen Institute der Philosophischen Fakultät als auch für die Wirtschaftswissenschaftliche und die Juristische Fakultät stellt DESTATIS die benötigten Daten mit dem notwendigen Aggregationsniveau bereit. Eine weitere Schwierigkeit ist, dass DESTATIS derzeit lediglich Daten über den fächerspezifischen Mittelwert bereitstellt, woraus keine Schlüsse auf die Verteilung der Daten um den Mittelwert gezogen werden können. So könnte es beispielsweise sein, dass der Mittelwert aufgrund einzelner extrem hoher Werte verzerrt ist. Daten zur Varianz könnten hier wichtige Informationen liefern, jedoch werden diese Daten aktuell nicht von DESTATIS zur Verfügung gestellt. Aus diesem Grund ist auch eine Standardisierung der Daten, beispielsweise durch eine z-Transformation, nicht möglich.

Ein Nachteil des Verfahrens liegt darin, dass keine Unterscheidung zwischen kompetitiven Drittmitteln (DFG, EU-Rahmenprogramme, gemeinnützige Stiftungen etc.) und Drittmitteln aus anderen Quellen (v.a. aus Unternehmen) vorgenommen wird. Dem Problem könnte begegnet werden, indem auf die größenrelativierten Bewilligungssummen der DFG zurückgegriffen wird. Bis dato werden diese nur auf der Ebene der Wissenschaftsbereiche publiziert (Deutsche Forschungsgemeinschaft 2009), sodass sämtliche geisteswissenschaftlichen Fächer (mit Sozial-, Wirtschafts- und Rechtswissenschaften) gemeinsam betrachtet werden. Notwendige Differenzierungen zwischen den Fächern können auf dieser Basis nicht abgebildet werden. Unter Umständen kann die DFG mit der Publikation des Förderrankings 2012 durchschnittliche Bewilligungssummen auf der Ebene der Fachkollegien bereitstellen. So kann eine differenziertere Abbildung von Fächerunterschieden im Drittmittelaufkommen erfolgen, als dies mit den Daten des Statistischen Bundesamtes möglich ist. Zum Beispiel ist eine Unterscheidung von Literatur und Sprachwissenschaften leichter durchführbar und verschiedene Bereiche von Physik und Chemie können differenziert werden. Auch die Zuordnung der kleinen Fächer ist in der Systematik der Fachkollegien leichter umsetzbar. Die Bewilligungssummen der DFG nach Fachkollegien eignen sich als Benchmark zur Beurteilung des Erfolges in der Einwerbung von DFG-Drittmitteln. Es ist aber nicht sicher, dass diese Daten als Standardisierungsbasis zeitnah zur Verfügung stehen.

Wird der Drittmittelerfolg über die größenrelativierten Bewilligungssummen der DFG gemessen, stellt sich darüber hinaus die Frage, wie mit Drittmitteln von anderen Förderorganisationen umgegangen wird, die ebenfalls eine kompetitive Vergabe im offenen Antragsverfahren durchführen, d.h. in erster Linie gemeinnützige Stiftungen wie z. B. der Volkswagen- oder Thyssen-Stiftung.<sup>17</sup> Insgesamt fällt das Fördervolumen der gemeinnützigen Stiftungen deutlich geringer aus als die Förderung durch die DFG. Beispielsweise hat die Volkswagen-Stiftung im Jahr 2010 geistes- und sozialwissenschaftliche Projekte mit etwa 23 Mio. Euro gefördert, die Thyssen-Stiftung mit ca. 10 Mio. Euro und die Gerda-Henkel-Stiftung mit ca. 6 Mio. Euro. Dem steht in den Geistes- und Sozialwissenschaften ein Fördervolumen der DFG von 850 Mio. Euro gegenüber. Da der Aufwand für die Berechnung größenrelativierter Bewilligungssummen für die gemeinnützigen Stiftungen hoch ist, ist es sinnvoll die DFG-Durchschnitte auch als Referenz für die Förderung durch Stiftungen und andere kompetitive Drittmittelgeber zu verwenden.

<sup>16</sup> Weil die durchschnittlichen Drittmittelausgaben pro Professor/in in den Daten von DESTATIS überraschend stark schwanken, sollte in der Praxis die Standardisierung für ein bestimmtes Referenzjahr anhand eines durchschnittlichen Durchschnittes über mehrere Jahre stattfinden.

<sup>17</sup> Wie aus dem DFG-Förderranking hervorgeht, spielt die Forschungsförderung durch den Bund und die Rahmenprogramme der EU insbesondere in den Geistes- und Sozialwissenschaften – gegenüber der Förderung durch die DFG – nur eine untergeordnete Rolle (Forschungsgemeinschaft, 2009, p. 77f).

### 5.3. Forschungsevaluation anhand von Publikationen

Publikationen sind als Medien wissenschaftlicher Kommunikation ein häufig genannter Indikator zur Messung der Forschungsleistung (vgl. u. a. Wolf et al. 2005; Fiedler et al. 2008). Forscher/innen publizieren wissenschaftliche Forschungsergebnisse im Wesentlichen aus zwei Gründen: Zum einen sind sie an der Verbreitung ihrer Ergebnisse und der Teilnahme an wissenschaftlichen Diskursen interessiert (vgl. Huff 1999, S. 5). Zum anderen tragen Publikationen – in Abhängigkeit vom Publikationsorgan in unterschiedlich starkem Maße – entscheidend zur Reputation von Wissenschaftlern/innen bei (vgl. Franck 1998; Haucap et al. 2005; Ellison 2010). Dazu existieren verschiedene Publikationsmedien wie zum Beispiel Monografien, (wissenschaftliche) Zeitschriften oder Herausgeber/innenbände (vgl. Stock 2000, S. 247f.).

Der Stellenwert der einzelnen Publikationsformen unterscheidet sich nach Fächern erheblich. Während in der Medizin sowie den Natur-, Bio- und Geowissenschaften überwiegend in Fachzeitschriften mit doppelblindem Begutachtungsverfahren veröffentlicht wird, spielen in den Geistes- und Rechtswissenschaften Monografien eine herausragende Rolle. Auch Beiträge in Herausgeber/innenbänden sind in den Geisteswissenschaften von großer Bedeutung. In den Rechtswissenschaften spielen zudem Kommentare von Gesetzen sowie Urteilsbesprechungen eine wichtige Rolle. Die Wirtschafts- und Sozialwissenschaften wiederum befinden sich in einer Übergangsphase, in der die Publikation von Aufsätzen in nationalen und internationalen Fachzeitschriften kontinuierlich an Bedeutung gegenüber Buchpublikationen gewinnt (vgl. Fiedler et al. 2006, S. 474; Fiedler et al. 2008, S. 486). Betrachtet man das Wissenschaftssystem insgesamt, dann kann festgestellt werden, dass Fachzeitschriften das wichtigste Publikationsmedium sind.

Die Verwendung von Publikationen zur Messung der Forschungsleistung bietet den Vorteil, dass diese als das eigentliche Produkt wissenschaftlicher Arbeit betrachtet werden können. Während Drittmittel zum Teil nur als Input-Faktor für Forschung gesehen werden, sind Publikationen der Output wissenschaftlicher Forschung. Jedoch kann die Nutzung von Publikationen als Indikator zur Messung der Forschungsleistung auch nicht kritiklos gesehen werden, da das Produkt wissenschaftlicher Forschung strenggenommen nicht die Publikation an sich ist, sondern die darin enthaltenen Ideen und Gedanken des Forschers/der Forscherin. Diese Gedanken sollen durch die Publikation anderen Forschenden zugänglich gemacht werden und einen Beitrag zum wissenschaftlichen Diskurs leisten. Die bloße Tatsache, dass die Gedanken in Form einer Publikation niedergeschrieben sind, lässt jedoch noch keinen Schluss über die wissenschaftliche Qualität der darin enthaltenen Ideen zu. Darüber hinaus wird kritisiert, dass die Fokussierung auf Publikationen Anreize zur Mehrfachverwertung von Ideen und Ergebnissen bzw. zur Aufteilung von Ergebnissen in „least publishable units“ bieten kann (Broad 1981).

Trotz der Kritik an der Nutzung von Publikationen als Indikator zur Evaluation der Forschungsleistung werden sie in der Praxis häufig verwendet. Dabei scheinen Publikationen in der hochschulpolitischen Diskussion jedoch eine geringere Bedeutung zu haben als in der wissenschaftlichen. Ein Indiz dafür ist, dass Publikationen im Rahmen der leistungsorientierten Mittelvergabe auf Länderebene nicht berücksichtigt werden (vgl. Jaeger/In der Smitten 2009, S. 12; 2010, S. 11), während sie beispielsweise auf Berufungen großen Einfluss haben (vgl. Müller-Camen/Salzgeber 2005, S. 277).

Zur Evaluation der Forschungsleistung anhand von Publikationen stehen verschiedene Herangehensweisen zur Verfügung. So kann zunächst die Zahl der Publikationen pro Professor/in oder alternativ pro vollzeitäquivalenter Mitarbeiter/innenstelle als Maß der Forschungsleistung herangezogen werden. Dieses Verfahren ist jedoch als nicht wissenschaftsadäquat zu verwerfen, wie in Abschnitt 5.3.1. näher ausgeführt wird. Weiterhin kann eine Gewichtung der Publikationen mittels einfacher Qualitätsranglisten (vgl. Abschnitt 5.3.2.) oder anhand von Journal-Impact-Faktoren (vgl. Abschnitt 5.3.3.) erfolgen.



Alternativ hierzu kann auch die Zahl der Zitationen, welche die Arbeiten eines Professors/einer Professorin erhalten, zur Messung der Forschungsleistung herangezogen werden, wie in Abschnitt 5.3.4. näher ausgeführt wird. Da sich diese drei Varianten im Wesentlichen auf die von einem Professor/einer Professorin veröffentlichten Zeitschriftenaufsätze beziehen, zeigt Abschnitt 5.3.5. auf, wie auch Monografien, welche insbesondere in den geistes- und rechtswissenschaftlichen Fächern einen hohen Stellenwert besitzen, in die Messung der Forschungsleistung einbezogen werden können.

Für jeden Indikator sind verschiedene Varianten zur Umsetzung denkbar



Abbildung 11: Übersicht über Verfahren zur Verwendung von Publikationen als Indikator der Forschungsleistung

### 5.3.1. Anzahl der Publikationen pro Kopf als nicht wissenschaftsadäquater Indikator zur Messung der Forschungsleistung

Die erste Möglichkeit zur Messung der Forschungsleistung durch Publikationen ist die Erfassung der Publikationen pro Kopf. Dabei wird beispielsweise auf Fakultätsebene die Gesamtzahl der Publikationen durch die Anzahl der Professoren/innen geteilt. Eine Differenzierung nach Art der Publikation (z. B. Artikel in peer-reviewed-Zeitschriften, in Sammelbänden etc.) würde nicht vorgenommen werden. Bei diesem Indikator würde eine höhere Anzahl an Publikationen als eine bessere Forschungsleistung interpretiert werden.

Ein unbestreitbarer Vorteil dieses Verfahrens ist die Einfachheit. Es müsste lediglich die Gesamtzahl der Publikationen sowie die Anzahl der Professoren/innen vorliegen, anschließend könnte der Indikator einfach berechnet werden.

Allerdings weist dieses Vorgehen gravierende Nachteile auf, die im Ergebnis dazu führen, dass der Indikator als nicht wissenschaftsadäquat eingestuft werden muss. Ein erster Nachteil ist, dass unterschiedliche Publikationsformen nicht berücksichtigt werden. Beispielsweise würde eine Monografie gleich gewichtet werden wie eine vierseitige Rezension. Zweitens ist eine Vergleichbarkeit über Fächergrenzen hinweg fraglich, da die Publikationskulturen sich sehr stark voneinander unterscheiden. In Teilen der Philosophischen Fakultät erfolgt die Kommunikation innerhalb der Scientific Community noch überwiegend über Monografien, während in anderen Bereichen, wie zum Beispiel den Naturwissenschaften, die Diskussion in Zeitschriften verläuft. Da Monografien sehr umfangreiche Werke sind, deren Erstellung einen entsprechenden zeitlichen Aufwand in Anspruch nimmt, wäre eine Vergleichbarkeit nicht gewährleistet.

Der Hauptkritikpunkt ist jedoch, dass die Qualität der Publikationen nicht berücksichtigt wird. Denn selbst wenn man verschiedene Publikationen quantitativ differenziert erfasst, indem man etwa die Seitenzahlen der Publikationen zählt und als Vergleichsgrundlage heranzieht, lässt dies keinerlei Rückschlüsse auf die Qualität der Publikation zu. Im Extremfall zählt beispielsweise ein richtungsweisender Aufsatz in einer Zeitschrift wie beispielweise „nature“ oder „Science“ ebenso viel wie ein Beitrag in einem nicht begutachteten Sammelband oder eine Rezension. Die Nutzung des Indikators könnte somit einen Anreiz setzen, anstelle von qualitativ hochwertigen Artikeln nur noch solche mit geringem Qualitätsniveau zu schreiben und dafür die Quantität zu erhöhen. Die DFG hat dieses Vorgehen scharf kritisiert und spricht in diesem Zusammenhang von „Qualität statt Quantität“ (vgl. Kleiner 2010). Aus diesen Gründen wird der Indikator Publikationen pro Kopf im Rahmen dieses Berichts nicht weiter verfolgt.

### 5.3.2. Qualitätsgewichtung anhand einfacher Ranglisten

Eine Möglichkeit, neben der Anzahl auch die Qualität der Publikationen im Rahmen der Forschungsevaluation zu berücksichtigen, ist die Verwendung einfacher Qualitätsranglisten. Dabei wird jedem Publikationstyp ein bestimmter Punktwert (meist zwischen 0 und 1) zugeordnet. Die Publikationsleistung ergibt sich dann, indem die Punktwerte für die einzelnen Publikationen aufsummiert werden.

Ein vor allem im Bereich der Wirtschaftswissenschaften weit verbreitetes Beispiel für solche Ranglisten sind Zeitschriftenranglisten oder Zeitschriftenrankings, deren Zahl in den letzten Jahren massiv zugenommen hat. Vom Verband der Hochschullehrer für Betriebswirtschaft wird beispielsweise das VHB-JOURQUAL-Ranking erstellt, das die Qualität nationaler und internationaler betriebswirtschaftlicher Zeitschriften abbilden soll (vgl. Henning-Thurau et al. 2004; Schrader/Henning-Thurau 2009). Trotz Kritik an dem Ansatz und der Methodik des Rankings (vgl. Dilger 2011) konnte es sich in der deutschen Betriebswirtschaftslehre als wichtigstes Zeitschriftenranking etablieren (vgl. Voeth et al. 2011, S. 440). Ein vergleichbares Instrument stellen die Rankings des „Handelsblatts“ für die Volks- und Betriebswirtschaftslehre dar.<sup>18</sup> Diese Rankings kategorisieren Zeitschriften anhand ihrer Qualität als A, B, C oder D-Journals, wobei die Kategorie A den 10 international besten Zeitschriften vorbehalten ist. Zeitschriftenrankings in dieser Form sind daher fachspezifisch konstruiert.

In den Wirtschaftswissenschaften werden diese Rankings regelmäßig dazu verwendet, um die Publikationsleistung von Wissenschaftlern/innen zu analysieren. So nutzen beispielsweise Fiedler et al. (2008) das VHB-JOURQUAL Ranking, um die Publikationsproduktivität des wissenschaftlichen Nachwuchses im Bereich der Betriebswirtschaftslehre zu analysieren. Ein ähnliches Verfahren wird auch im Handelsblatt-Ranking BWL genutzt, um die (Publikations-)Leistung betriebswirtschaftlicher Forscher/innen zu bewerten, wobei die Zeitschriftenbewertungen aus drei Rankings (VHB-JOURQUAL, Erasmus Research Institute of Management Journal List sowie (Social) Science Citation Index) zusammengefasst und dementsprechend den Publikationen Punkte zugewiesen werden (vgl. Müller 2010, S. 156f.). Auch die Forschungsleistung von Fakultäten und Universitäten wird stellenweise auf Basis von Publikationen ermittelt (vgl. z. B. Wolf et al. 2005). Ähnlich geht die University of Texas at Dallas (UTD) bei dem „UTD Top 100 Business School Research Ranking“ vor; hier werden Publikationen in 24 als international führend eingestuft wirtschaftswissenschaftlichen Zeitschriften gewertet, um Business Schools hinsichtlich ihrer Forschungsleistung zu vergleichen (University of Texas 2011).

<sup>18</sup> Zur Methodik siehe:

<http://www.handelsblatt.com/politik/oekonomie/vwl-ranking/methodik-so-funktioniert-das-vwl-ranking/4575334.html>



Ein Hauptkritikpunkt an diesem Vorgehen ist, dass die Anzahl an Publikationen – auch nach einer Gewichtung mit der Reputation der Zeitschrift, in der sie veröffentlicht wurden – nur eine eingeschränkte Aussagekraft über die Qualität der Forschungsleistung hat (vgl. Dilger 2000, S. 475; Dilger 2010). So zeigt Schläpfer (2010), dass zwischen der Position von Volkswirten im Handelsblatt-Ranking des Jahres 2008 und den Zitationsdaten aus dem „Web of Science“ nur ein sehr schwacher statistischer Zusammenhang besteht. Ähnliche Kritik wird von Müller (2010) sowie Voeth et al. (2011) formuliert. Bräuning et al. (2011) bemerken, dass selbst in hochkarätigen Zeitschriften publizierte Aufsätze teilweise gar nicht zitiert werden, sondern die Verteilung der Zitationen über die Artikel einer Zeitschrift sehr „schief“ ist, das heißt, dass sich viele Zitationen auf wenige Aufsätze verteilen. Anders ausgedrückt ist der durchschnittliche Impact-Faktor für die in einer Zeitschrift publizierten Aufsätze deutlich höher als der Median. Zudem variiert die Schiefe der Verteilung zwischen Fachzeitschriften, und schließlich werden viele Aufsätze in hochkarätigen Journalen weniger häufig zitiert als ein Aufsatz in manchen weniger renommierten Zeitschriften (vgl. Oswald 2007; Wall 2009).

Ein Nachteil von Ranglisten ist, dass nur für VWL und BWL bundesweit anerkannte Zeitschriftenlisten vorliegen. Für andere Fächer liegen keine Ranglisten vor oder sie werden lokal erstellt und eingesetzt (z. B. an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln). Hinzu kommt, dass Buchpublikationen in den Zeitschriftenrankings nicht berücksichtigt werden. Diese müssen jedoch in das Bewertungsschema integriert werden, damit eine vollständige Leistungsbewertung anhand von Publikationen umgesetzt werden kann. Ranglisten für Publikationen müssten daher fachspezifisch von den Fakultäten ausgearbeitet werden. Das führt zu dem Anschlussproblem, wie die fächerübergreifende Vergleichbarkeit der Ranglisten sichergestellt werden kann.

### 5.3.3. Qualitätsgewichtung anhand von Journal-Impact-Faktoren

International weiter verbreitet als fachspezifische Zeitschriftenrankings ist die Beurteilung von Fachzeitschriften anhand von Journal-Impact-Faktoren. Impact-Faktoren beschreiben die durchschnittliche Anzahl von Zitationen pro publiziertem Artikel in einer Zeitschrift in den zwei bzw. fünf Jahren nach Erscheinen. Der zweijährige Impact-Faktor einer Zeitschrift für 2010 wird folglich über die Anzahl aller Zitate im Jahr 2010 von Artikeln, die in den Jahren 2009 und 2008 erschienen sind, geteilt durch die Anzahl der in den Jahren 2009 und 2008 in dieser Zeitschrift erschienen Artikel, bestimmt. Impact-Faktoren messen somit die Sichtbarkeit einer Publikationsquelle. Der Verwendung von Impact-Faktoren als Maß für die Qualität einer Zeitschrift liegt die Annahme zu Grunde, dass Fachzeitschriften mit einem hohen Impact-Faktor im Begutachtungsprozess besonders hohe Ansprüche an die Qualität von Aufsätzen stellen und diese auch durchsetzen können. Deshalb wird der Impact-Faktor als approximatives Maß für die Qualität der Aufsätze verwendet, zum Beispiel als Gewichtungsfaktor für Publikationen bei der Bildung eines Publikationsindex. Impact-Faktoren werden insbesondere von den Anbietern wissenschaftlicher Literaturdatenbanken „ThomsonReuter“ (Science-Citation-Index (SCI), Social-Science-Citation-Index (SSCI)) und „Elsevier“ (Scopus) angeboten.<sup>19</sup> Die Bedingung für die Aufnahme einer Zeitschrift in den SCI, SSCI oder Scopus ist deshalb, dass Aufsätze vor der Publikation ein Peer-Review-Verfahren durchlaufen, das heißt im Doppelblindverfahren begutachtet werden.

Der wichtigste Unterschied zwischen Zeitschriftenrankings und Impact-Faktoren liegt darin, dass bei der Erstellung von Zeitschriftenrankings häufig neben quantitativen Größen, wie z. B. Zitirraten, auch qualitative Aspekte, wie zum Beispiel Einschätzungen von Experten und ähnliches, in die Bewertung mit einfließen (siehe z. B. das VHB-JOURQUAL Ranking). Ein Problem besteht darin, dass die Forschungsleistung der Experten

<sup>19</sup> Für den Arts & Humanities Citation Index (A&HCI) werden keine bibliometrischen Kennzahlen berichtet.

häufig auch auf Grundlage des Rankings bewertet wird, so dass der Anreiz besteht, die Zeitschriften besser zu bewerten, in denen die Experten selbst veröffentlicht haben. Die Verwendung von Journal-Impact-Faktoren kann somit als objektiver eingestuft werden, da dabei ausschließlich quantitative Kriterien berücksichtigt werden, die transparent und nachvollziehbar sind.

Ein großer Nachteil „einfacher“ Journal-Impact-Faktoren gegenüber Zeitschriftenrankings ist die Tatsache, dass Impact-Faktoren in der Regel absolut interpretiert werden. Allerdings variieren die Journal-Impact-Faktoren zwischen den Fächern zum Teil erheblich. So erreicht beispielsweise das „Journal of Economic Literature“, die meistzitierte volkswirtschaftliche Zeitschrift, einen Impact-Faktor von 7,4, wohingegen das „New England Journal of Medicine“, das zu den wichtigsten medizinischen Fachzeitschriften gehört, einen Impact-Faktor von 53,5 aufweist. Eine Interpretation über Fächergrenzen hinweg ist daher nicht möglich, da hierbei bestimmte Fächergruppen systematisch benachteiligt würden. Um eine fächerübergreifende Vergleichbarkeit zu gewährleisten, muss zunächst eine fachspezifische Einbettung der Impact-Faktoren erfolgen.

Der einfachste Weg der fachspezifischen Einbettung beruht darauf, die jeweiligen Impact-Faktoren („Citations Per Publication“ – CPP) anhand eines fächerspezifischen, durchschnittlichen Impact-Faktors zu gewichten („Mean Field Citation Score – FCS). Dabei wird der Impact-Faktor für ein Set von 1... n Publikationen (z.B. Zeitschriften) mit  $c_i$  der Anzahl der Zitationen der Publikation  $i$  und  $e_i$ , dem Durchschnitt der Zitationen für Publikationen in einem Fachgebiet durch den durchschnittlichen Impact-Faktor, geteilt (Waltman et al. 2011, S. 469):

$$\frac{CPP}{FCS} = \frac{\sum_{i=1}^n \frac{c_i}{n}}{\sum_{i=1}^n \frac{e_i}{n}} = \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^n e_i}$$

Das Ergebnis dieser Ratio zeigt an, ob und wie stark die Sichtbarkeit einer Zeitschrift unter oder über dem Durchschnitt der Zeitschriften eines Faches liegt. Beispielsweise bedeutet ein Wert von 1,5, dass auf Artikel in dieser Zeitschrift 50 Prozent mehr Zitationen entfallen als im Durchschnitt des Faches. Der CPP/FCS, wie er vom CWTS der Universität Leiden entwickelt wurde, klassifiziert die Zeitschriften anhand der Fächerklassifikationen (subject areas) des „Web of Science“ (WoS).

Allerdings stellt sich die Frage, ob die Fächerklassifikation des WoS ausreichend differenziert ist, um systematische Differenzen zu korrigieren. Deshalb empfiehlt die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) eine weitere Differenzierung der durchschnittlichen Impact-Faktoren für die Subdisziplinen der Medizin, weil klinische Studien sehr viel häufiger zitiert werden als medizinische Grundlagenforschung. Im Ergebnis haben die klinischen Zeitschriften deshalb einen im Mittel sehr viel höheren Impact-Faktor, ohne dass Rückschlüsse auf die Qualität der Zeitschriften möglich sind. Die AWMF ordnet alle medizinischen Fachzeitschriften bestimmten Teilgebieten der Medizin zu und berechnet den durchschnittlichen Impact-Faktor der Teilgebiete.<sup>20</sup> Diese Liste wird von der AWMF jährlich aktualisiert und publiziert. Die Publikationen eines Autors können dann anhand der mittelwertkorrigierten Impact-Faktoren gewichtet und zu einem Publikationsscore summiert werden. Für Publikationen in deutschsprachigen Zeitschriften empfiehlt die AWMF den gewichteten Impact-Faktor zu verdoppeln.<sup>21</sup>

<sup>20</sup> Siehe: <http://www.awmf.org/forschung-lehre/stellungnahmen/forschung-lehre/leistungsorientierte-mittelvergabe-an-med-fakultaeten.html> und <http://www.awmf.org/index.php?id=94>.

<sup>21</sup> Die AWMF gibt auch Empfehlungen für den Umgang mit Publikationen, für die kein gewichteter Impact-Faktor bestimmt werden kann. Für diesen Fall wird eine untere Grenze für die Gewichtung eingeführt (0,2), mit der alle Publikationen mindestens in den Publikations-Score einfließen. Auch Beiträge zu Herausgeber/innenbänden und Monografien fließen mit einem festen Impact-Faktor (sog. Ersatzimpactfaktorpunkten) in den Publikations-Score ein, der nach dem Umfang der Publikation gestaffelt wird: 5-20 Seiten = 0,5 Punkte; 21-50 Seiten = 1,0 Punkte; 51-200 Seiten = 2,0 Punkte, > 200 Seiten = 4,0 Punkte. Um gewertet zu werden, müssen die Beiträge in einem wissenschaftlichen Fachverlag erschienen sein.



Der Vorteil dieses Verfahrens liegt in der transparenten Logik und der verhältnismäßig einfachen praktischen Umsetzbarkeit (sofern eine ausreichende Anzahl an Zeitschriften eines Fachgebietes in den einschlägigen Datenbanken erfasst ist). Ein Nachteil dieses Verfahrens liegt in der Abgrenzung der einzelnen Fächer, für die durchschnittliche Impact-Faktoren bestimmt werden. Insbesondere wenn eine Zeitschrift mehreren (Teil-)Fächern zugeordnet werden kann, ist unklar, wie der durchschnittliche Impact-Faktor zur Gewichtung bestimmt werden kann. Darüber hinaus wird der CPP/FCS von den kommerziellen Datenbankanbietern nicht standardmäßig bereitgestellt.

Um Probleme bei der *a priori*-Klassifikation bei der Bestimmung der durchschnittlichen Impact-Faktoren zu umgehen, kann die Zuordnung einer Zeitschrift zu einem Fachgebiet und damit zu dessen „Publikationskultur“ auch empirisch erfolgen. In der Szientometrie wird in diesem Zusammenhang von „field normalization“ gesprochen (eine Übersicht über verschiedene Normalisierungsverfahren gibt (Waltman et al. 2011, p. 302)). Der Datenbankanbieter „Elsevier“ bietet mit dem „Source Normalized Impact per Paper“ (SNIP) (Moed, 2009; D. Torres-Salinas & Jimenez-Contreras, 2010) standardmäßig einen normalisierten Impact-Faktor an. Eine ausführliche Beschreibung des SNIP ist in Anhang 2 zu finden.

In der Praxis stößt die Evaluation von Publikationen anhand von Impact-Faktoren dann an ihre Grenzen, wenn ein bedeutender Teil der Publikationen eines Faches nicht in Fachzeitschriften mit Begutachtungsverfahren erscheint. Dies ist insbesondere für die Rechtswissenschaften der Fall, deren Publikationen von keinem Datenbankanbieter systematisch erfasst werden. Alle Publikationen, die nicht in den Datenbanken des „Web of Science“ oder von „Elsevier“ erfasst sind, könnten dann nicht in die systematische Leistungsevaluation einbezogen werden. Dann können keine Impact-Faktoren zur Beurteilung der Publikationen herangezogen werden.

#### 5.3.4. Zitationsanalysen

Zitationskennzahlen sind ein weiterer, viel diskutierter Ansatz zur Evaluation der Forschungsleistung. Ihrer Nutzung liegt die Annahme zugrunde, dass die Qualität eines Artikels nur unzureichend durch die Reputation des Publikationsmediums bewertet werden kann (vgl. Dilger 2010; Müller 2010; Schläpfer/Schneider 2010), da trotz des Peer-Review-Verfahrens Qualitätsunterschiede zwischen den in einer Zeitschrift veröffentlichten Artikeln bestehen (vgl. Starbuck 2005). Wie bereits erläutert werden selbst in hochkarätigen Zeitschriften viele Aufsätze nur selten oder sogar überhaupt nicht zitiert, sodass die Verwendung durchschnittlicher Zitieraten in Form von Impact-Faktoren die Qualität einzelner Aufsätze teilweise nur sehr unzureichend beschreibt.

Um diesem Problem zu begegnen, kann daher die tatsächliche Zitierate – d.h. die Gesamtzahl der Zitierungen, die eine Publikation erhalten hat – erhoben werden, da davon auszugehen ist, dass „gute“ Publikationen häufiger zitiert werden als „weniger gute“ (vgl. Dilger 2000, S. 475; Dyckhoff et al. 2005, S. 351; Dyckhoff/Schmitz 2007). In diesem Fall ergibt sich die Leistung eines Forschers/einer Forscherin, einer Fakultät oder einer Universität anhand der Gesamtzahl der Zitationen, die auf alle Publikationen des Forschers/der Forscherin bzw. der Fakultät oder Universität entfallen.

Neben der Messung der Gesamtzahl der Zitationen existieren noch weitere Zitationskennzahlen, beispielsweise der h-Index (vgl. Hirsch 2005), der g-Index (Egghe 2006) oder auch die nach dem Alter eines Artikels gewichtete Erfassung der Zitationen (BiHui et al. 2007, S. 859), die jeweils eine möglichst exakte Bestimmung der Forschungsleistung zum Ziel haben. Grundlage der Erhebung von Zitationsdaten sind Datenbanken wie beispielsweise das „Web of Science“, „Scopus“ und „Google Scholar“ (vgl. Kulkarni et al. 2009; Meho/Yang 2007).

In der wissenschaftlichen Praxis variiert der Einfluss von Zitationskennzahlen bei der Evaluation der Forschungsleistung zum Teil erheblich. In Deutschland finden sich hierzu nur wenige Ansätze. Litzenberger und Sternberg (2005) evaluieren beispielsweise die Forschungsleistung von soziologischen Instituten an zehn deutschen Universitäten mittels Zitationsdaten aus dem „Social Science Citation Index“ (SSCI). Dyckhoff et al. (2005) wiederum nutzen Zitationsdaten, um die Forschungsleistung auf individueller Ebene – in diesem Fall von Hochschullehrern/innen für Betriebswirtschaftslehre – zu messen. In etablierten Forschungsrankings wie beispielsweise dem Handelsblatt-Ranking oder dem CHE-Ranking in den Sozial- und Wirtschaftswissenschaften haben Zitationsdaten hingegen kein Gewicht. Auch im Rahmen der leistungsbezogenen Mittelvergabe zwischen Ländern und Universitäten werden Zitationskennzahlen bislang nicht berücksichtigt (vgl. Jaeger/In der Smitten 2009, S. 13; 2010, S. 9). Ein anderes Bild zeigt sich in internationalen Wissenschaftsrankings (vgl. Enserink 2007): So gehen Zitationskennzahlen mit einem Gewicht von 20 Prozent sowohl in das „Academic Ranking of World Universities“ der Shanghai University als auch in das „Times Higher Education Supplement“ ein, während das Ranking des Centre for Science and Technology Studies der Leiden University die Forschungsleistung europäischer Universitäten sogar ausschließlich nach Zitationskennzahlen bewertet.

Jedoch existieren auch Grenzen von Zitationskennzahlen zur Evaluation der Forschungsleistung, da Zitationskennzahlen durch Selbstzitate, Zitationskartelle, strategisches Zitieren oder auch durch Unterlassungen von Zitaten manipuliert werden können (vgl. Kieser 1998, S. 217; Dilger 2000, S. 476). So zeigt Hudson (2007) in seiner empirischen Untersuchung der Artikel aus American „Economic Review“ und „Economic Journal“, dass die Zahl der Zitationen, die auf einen Artikel entfallen, auch davon abhängt, wie häufig andere Artikel in der gleichen Ausgabe der Zeitschrift zitiert werden. Darüber hinaus erfassen die Kennzahlen nicht, ob eine Arbeit zitiert wird, weil sie als besonders gut oder besonders schlecht eingestuft wird (Frey/Rost 2010, S. 6f.), so dass eine hohe Zitierrete nicht automatisch auch eine hohe Qualität implizieren muss.

Zusätzlich können Moden dazu führen, dass Arbeiten nur zitiert werden, weil sie vorher bereits in anderen Artikeln zitiert wurden (vgl. Frey/Rost 2010, S. 7). Der sogenannte „Matthäus-Effekt“ verstärkt dies (vgl. Merton 1968): Er besagt, dass Arbeiten, unabhängig von ihrer Qualität, häufiger zitiert werden, wenn sie von Forschern/innen verfasst wurden, deren Arbeiten in der Vergangenheit bereits häufig zitiert wurden (vgl. Merton 1968, S. 60).

Obwohl Zitationskennzahlen in der Literatur diskutiert sowie in internationalen Rankings als Indikator zur Evaluation der Forschungsleistung genutzt werden, ist ihr direkter Einfluss auf die Forschungsevaluation in Deutschland gering. Sie werden derzeit weder bei einflussreichen Rankings noch im Rahmen der leistungsorientierten Mittelvergabe auf Länderebene direkt berücksichtigt (vgl. Jaeger/In der Smitten 2010, S. 11).

### 5.3.5. Behandlung von Buchpublikationen im Rahmen der Forschungsevaluation

Ein Desiderat in der bibliometrischen Forschung ist der Umgang mit Buchpublikationen (Monografien, Herausgeber/innenbände und Gesetzeskommentare) in der Leistungsevaluation. Die Notwendigkeit, Buchpublikationen in die Bewertung wissenschaftlicher Leistungen einzubeziehen, ist mehrfach nachgewiesen worden (Gläser 2006; Linmans 2010; Zuccala/van Leeuwen 2011).<sup>22</sup> Nederhof et al. (2010) zum Beispiel zeigen in einer Analyse von Literaturverzeichnissen in Fachzeitschriften, dass in Psychologie, Politikwissenschaften und Wirtschaftswissenschaften von den Zitaten, die auf nicht im WoS erfasste Quellen verweisen, die Hälfte auf Buchpublikationen entfallen (Lindholm-Romantschuk/Warner 1996; Nederhof, et al. 2010, S. 368).

<sup>22</sup> Anders argumentieren hingegen Litzenberger & Sternberg, 2005.



Dadurch, dass Zitate auf und aus Buchpublikationen in Zitationsanalysen nicht erfasst werden, sind die Ergebnisse für die Geistes- und Sozialwissenschaften nicht brauchbar (Kousha & Thelwall 2009).

Dennoch spielen quantitative Indikatoren zur Bewertung der Qualität von Monografien in der Wissenschaftsevaluation bis dato kaum eine Rolle, da in den Naturwissenschaften, der Medizin und den Wirtschaftswissenschaften Monografien in erster Linie als Hand- und Lehrbücher eine Rolle spielen. Gerade in den Naturwissenschaften verweisen nur drei bis fünf Prozent der Referenzen nicht auf Artikel in Fachzeitschriften, die im WoS erfasst sind.

Ein Grund für die Vernachlässigung von Buchpublikationen in der szientometrischen Literatur ist das Fehlen von Datenbanken, die systematisch Buchpublikationen und deren Referenzen erfassen.<sup>23</sup> Seit kurzer Zeit steht mit „Google-Books“ eine erste Datenbank zur Verfügung, die es erlaubt, Zitationen aus Büchern und auf Bücher zu analysieren (Kousha & Thelwall 2009). Allerdings handelt es sich nicht um eine systematische Erfassung wissenschaftlicher Werke. Mit „Google-Scholar“ steht eine Datenbank zur Verfügung, die alle über das Internet verfügbaren wissenschaftlichen Publikationen umfasst.<sup>24</sup> Der Nachteil vor allem von „Google-Scholar“ ist jedoch, dass auch Zitate aus Handouts, Präsentationen und graue Literatur gezählt werden, die in eine Zitationsanalyse in der Regel nicht eingehen sollen. Die Abgrenzung der Publikationsformen gelingt in „Google Scholar“ noch nicht zufriedenstellend.

Seit Oktober 2011 wird von ThomsonReuter ein „Book-Citation-Index“ bereitgestellt (BCI), der vorerst jedoch auf englischsprachige Bücher mit Peer-Review beschränkt bleibt und damit für die Wissenschaftsevaluation in Deutschland nur sehr eingeschränkt verwendet werden kann, weil gerade in den Geistes- und Sozialwissenschaften neben englischsprachigen insbesondere auch deutsch-, französisch-, spanisch- und italienischsprachige Bücher erscheinen.

Torres-Salinas und Moed (2009, S. 10) schlagen vor, elektronische Bibliothekskataloge („Online Public Access Catalogs“ – OPACs) für die Analyse der Verbreitung von Buchpublikationen zu verwenden. Die Autoren argumentieren, dass das Vorhandensein eines Buches in einer wissenschaftlichen Bibliothek ein Indikator für den „Konsum“ dieses Titels ist, aus dem auf dessen Nutzen für die wissenschaftliche Arbeit geschlossen wird. Indem eine ausreichend große Grundgesamtheit von wissenschaftlichen Bibliotheken identifiziert wird, kann ein reliables Bild des Konsums einer Buchpublikation nachgezeichnet werden, der dann den Nutzen des Buches für die Arbeit anderer Wissenschaftler/innen abbildet.

Dem liegt die Annahme zugrunde, dass mit der Qualität einer Buchpublikation die Wahrscheinlichkeit zunimmt, dass diese von wissenschaftlichen Bibliotheken angeschafft wird (z.B. aufgrund eines Anschaffungsvorschlages von Wissenschaftlern/innen). Das Maß der Bibliotheksverbreitung steht dann für eine Synthese formaler (z.B. das Renommee des Verlages, Bekanntheitsgrad des Autors/der Autorin bzw. Herausgebers/Herausgeberin etc.) und inhaltlicher (z.B. Relevanz des Themas, Qualität der Bearbeitung etc.) Qualitätskriterien für Bücher. Die Bibliotheksverbreitung eines Buches kann daher als funktionales Äquivalent eines Impact-Faktors angesehen werden.<sup>25</sup>

<sup>23</sup> Für die Sozialwissenschaften werden deutschsprachige Buchpublikationen durch die GESIS erfasst ([www.sowiport.de](http://www.sowiport.de)). Englischsprachige Bücher werden in den „Cambridge Sociological Abstracts“ erfasst. Allerdings werden Zitate in beiden Datenbanken nicht ausgewiesen.

<sup>24</sup> Eine systematische Analyse ist mit dem Programm „Publish or Perish“ möglich.

<sup>25</sup> Konzeptuell ist die Buchverbreitung als Indikator folglich zwischen der direkten Untersuchung von Zitationen auf Artikel und dem Impact-Faktor einer Zeitschrift angesiedelt, da er sich zwar auf den Buchtitel bezieht (und nicht etwa auf das Prestige des Verlages), nicht jedoch auf den Zitationen auf das Buch beruht. Dies gilt insbesondere für die Bibliotheksverbreitung von Herausgeber/innenbänden, da nicht jeder Beitrag einzeln von den Bibliotheken angeschafft wird.

Das einfachste Maß für die Bibliotheksverbreitung ist die Diffusionsrate eines Titels (Daniel Torres-Salinas & Moed, 2009, S. 13), d.h. der Anteil der Katalogeinträge eines Buches von der maximal möglichen Zahl von Katalogeinträgen, wobei  $CI_s$  für die Katalogeinträge eines Titels oder eine Gruppe von Titeln steht und  $CI_m$  für die maximal mögliche Zahl von Katalogeinträgen:

$$DR = \left( \frac{CI_s}{CI_m} \right)$$

Die Grundgesamtheit der Bibliothekskataloge, die für die Berechnung der Diffusionsrate herangezogen wird, setzt sich ausschließlich aus wissenschaftlichen Bibliothekskatalogen zusammen, d.h. in erster Linie Universitätskatalogen.

Ein Nachteil der Diffusionsrate ist – genauso wie bei Brutto-Impact-Faktoren – dass Unterschiede in den Publikationskulturen der Fächer a priori nicht berücksichtigt werden. Wenn alle Fächer anhand der gleichen Bibliothekskataloge evaluiert werden, werden solche Fächer einseitig benachteiligt, die nicht an allen Universitäten vertreten sind (in erster Linie also die kleinen, geisteswissenschaftlichen Fächer). Die Berücksichtigung fachspezifischer Publikationskulturen kann nur über eine fachspezifische Auswahl der relevanten wissenschaftlichen Bibliothekskataloge erfolgen ( $CI_m$ ). Ein Bezug zu fachspezifischen Durchschnitten wird in der Diffusionsrate nicht abgebildet.

Eine Alternative zur Diffusionsrate, die diesen Bezug herstellt, ist die „Relative Catalog Inclusion Rate“ (RCIR). Dieses Maß bestimmt für eine bestimmte Grundgesamtheit von Titeln die durchschnittliche Anzahl Katalogeinträge. Die Eintragsrate eines bestimmten Titels bzw. einer Gruppe von Titeln wird dann durch die durchschnittliche Inklusionsrate geteilt, wodurch angezeigt wird, ob ein Titel bzw. eine Gruppe von Titeln über- oder unterdurchschnittlich viele Katalogeinträge erhält. Um dieses Maß umzusetzen, muss die durchschnittliche Inklusionsrate ermittelt werden, was relativ aufwendig ist, da es die fachspezifische Einbettung des Indikators erforderlich macht, fachgebietsspezifische durchschnittliche Inklusionsraten zu berechnen. Dabei stellt sich das praktische Problem, wie sinnvolle Abgrenzungen von Fächern vorgenommen werden können.

Ein generelles Problem bei der Anwendung der Bibliotheksverbreitungsrate zur Bewertung von Buchpublikationen ist, dass die Universitätsbibliotheken von den Verlagen E-Books im Rahmen größerer Pakete abonnieren. Ist ein Titel im Bibliothekskatalog vorhanden, liegt dann nicht mehr zwingend eine spezifische Kaufentscheidung für den Titel zugrunde, was die konzeptuelle Logik dieses Indikators in Frage stellt.

Kennzahlen zur Bibliotheksverbreitung sind derzeit nicht vorhanden und müssten deshalb für die Publikationen an der HHU ermittelt werden, was eine erhebliche Investition in informationswissenschaftliche Infrastruktur bedeutet. Die Verwendung bibliometrischer Maße für die Bewertung von Buchpublikationen ist folglich mit erheblichen praktischen Problemen verbunden.

#### 5.4. Forschungsevaluation mittels Informed-Peer-Review

Der Wissenschaftsrat empfiehlt für die Bewertung von Forschungsleistungen nicht nur auf standardisierte Maße für Publikationsqualität (z.B. Impact-Faktoren, wie bereits beschrieben) oder Kennzahlen zum Drittmittelerfolg zurückzugreifen. Im Evaluationsverfahren des Wissenschaftsrates werden stattdessen ausgewählte Publikationen jedes/r Wissenschaftlers/in von anderen Fachwissenschaftlern/innen begutachtet, wobei die „Originalität der Forschung“ (Wissenschaftsrat 2005a) das wichtigste Bewertungskriterium ist.



Dieses Verfahren wurde beim Forschungsrating für Soziologie und Chemie eingesetzt (Wissenschaftsrat 2008) und für das (ausstehende) Forschungsrating eines geisteswissenschaftlichen Faches empfohlen (Wissenschaftsrat 2010).<sup>26</sup> In den Pilotstudien des Wissenschaftsrates wird die qualitative Bewertung durch weitergehende quantitative (insbesondere Anzahl der Publikationen nach Publikationstypen, Höhe der eingeworbenen Drittmittel etc.) und qualitative (Übersicht über Drittmittelprojekte, Selbstdarstellung von Stärken und Schwächen der Institute etc.) Indikatoren gestützt, die mit in die Gesamtbewertung des Gutachters/der Gutachterin einfließen.<sup>27</sup>

Der Vorteil eines Peer-Review-Verfahrens liegt darin, dass die Qualität der Forschungsleistung unabhängig vom Ort der Publikation erfasst wird. Auf diese Weise werden auch Leistungen gewürdigt, die vom Thema oder Format her nicht in das Schema von Fachzeitschriften passen, was insbesondere die Bewertung geisteswissenschaftlicher Fächer erleichtert, weil auch die Qualität von Buchpublikationen beurteilt wird, für die – wie bereits erläutert – standardisierte Maße weitgehend fehlen.<sup>28</sup> Ein weiterer Vorteil von Peer-Reviews ist, dass die fachspezifische Einbettung der Bewertung in der Verfahrenslogik angelegt ist (es sollen schließlich keine fachfremden Gutachter/innen eingesetzt werden).

Allerdings liegen auch die Nachteile des Begutachtungsverfahrens auf der Hand: Erstens ist die Organisation eines Peer-Review-Verfahrens sehr aufwendig, weil Gutachter/innen für alle eingereichten Publikationen gefunden und zur Mitarbeit bewegt werden müssen. Daher müssen ggf. Honorare für die Erstellung der Gutachten vorgesehen werden. Auch kann der Fall eintreten, dass aufgrund der Knappheit geeigneter Gutachter/innen kein unabhängiges Gutachten eingeholt werden kann (Schmoch 2009). Zudem muss die Befangenheit von Gutachtern/innen ausgeschlossen werden, so dass gerade in den kleinen Fächern (vor allem in den Geisteswissenschaften) ein unabhängiges Peer-Review-Verfahren möglicherweise nur eingeschränkt durchgeführt werden kann. Darüber hinaus sind Schwierigkeiten bei der Rekrutierung von Gutachtern/innen wahrscheinlich, da trotz einer stagnierenden Anzahl an Gutachtern/innen der Bedarf steigt (z. B. für den Reviewprozess in Fachzeitschriften, die Vergabe von Drittmitteln, die Evaluation von Forschungsinstituten etc.).

Zweitens liegt in der Begutachtung von Publikationen durch Peers zwangsläufig ein subjektives Element, da die Kriterien für die Bewertung der Publikationen zwar vorgegeben werden können, deren Anwendung in der Bewertungsskala allerdings der Einschätzung des Gutachters/der Gutachterin unterliegt. Folglich sind die Einschätzungen von Gutachtern/innen nur eingeschränkt vergleichbar. Gelöst werden kann dieses Problem, indem jede Publikation von mehreren Gutachtern/innen bewertet wird, wodurch der Aufwand für die Durchführung des Verfahrens allerdings noch weiter erhöht wird. Ein Peer-Review-Verfahren zur Bewertung der Forschungsqualität kann deshalb nur in relativ großen Intervallen von drei bis fünf Jahren stattfinden.<sup>29</sup>

Die Einschätzung der Brauchbarkeit einer Informed-Peer-Review zur Bewertung der Wissenschaftler/innen der HHU muss deshalb zurückhaltend erfolgen. Zwar ist das Verfahren gut geeignet, Unterschiede in den Fächerkulturen zu berücksichtigen, allerdings ist der Aufwand für die Durchführung sehr groß und das subjektive Element in der Bewertung kann die Akzeptanz der Ergebnisse beeinträchtigen.

---

<sup>26</sup> Das Rating erfolgt für die Anglistik. Die Ergebnisse werden Ende 2012 publiziert.

<sup>27</sup> Für eine kritische Auseinandersetzung mit dem Verfahren der Informed-Peer-Review vgl. Riordan et al. 2011.

<sup>28</sup> Der Wissenschaftsrat definiert in diesem Zusammenhang sehr allgemeine Bewertungskriterien: „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.“ (Wissenschaftsrat 2005b, S. 203).

<sup>29</sup> Die Leibniz-Gemeinschaft evaluiert ihre Mitgliedsinstitute regelmäßig in einem Zyklus von sieben Jahren.

## 6. EMPFEHLUNGEN BEZÜGLICH GEEIGNETER KRITERIEN ZUR MESSUNG DER FORSCHUNGSLEISTUNG IN DER PHILOSOPHISCHEN, DER WIRTSCHAFTSWISSENSCHAFTLICHEN UND DER JURISTISCHEN FAKULTÄT

### 6.1 Anforderungen an geeignete Leistungsparameter

Das zu entwickelnde System zur Messung der Forschungsleistung muss verschiedenen Ansprüchen genügen, um eine aussagekräftige wissenschaftsadäquate Leistungsbewertung zu ermöglichen. Ein System, das diesen Ansprüchen gerecht wird, kann mit großer Akzeptanz in der Professoren/innenschaft rechnen.

*Erstens* muss eine Leistungsbewertung multikriteriell erfolgen, um sicherzustellen, dass alle relevanten Dimensionen der Forschungstätigkeit von den Leistungsparametern abgedeckt werden. Eine Einengung auf einen Indikator wird den normativen Ansprüchen an die Leistungsbewertung nicht gerecht (vgl. BVerfGE 111, 333, 359 und Kapitel 3 dieses Gutachtens).

*Zweitens* muss die Leistungsbewertung die fachspezifischen Wissenschaftskulturen berücksichtigen (vgl. BVerfGE 111, 333, 359 und Kapitel 3 dieses Gutachtens). Das bedeutet, dass die Forschungsleistung über Fächergrenzen hinweg nicht direkt miteinander verglichen werden dürfen, weil die strukturellen Bedingungen für die Leistungserbringung und die Definitionen exzellenter wissenschaftlicher Leistungen sich von Fach zu Fach unterscheiden.

*Drittens* sollten die Indikatoren so weit wie möglich einen Vergleich der Fächer und Fakultäten zulassen. Folglich müssen die Indikatoren zwar die Besonderheiten der Fächer abbilden, aber konzeptuell vergleichbar aufgebaut sein.

Indikatoren genügen den Ansprüchen aus zweitens und drittens, wenn die gleichen Indikatoren im Verhältnis zu fachspezifischen Durchschnittsleistungen interpretiert werden. Auf diese Weise findet der Vergleich der Leistungen innerhalb der Fächer statt, indem festgestellt wird, wie die Einheiten relativ zum relevanten Durchschnitt stehen. Gleichzeitig können die Fächer verglichen werden, weil die Indikatoren für alle Fächer gleich sind.

*Viertens* sollte die Leistungsbewertung so effizient wie möglich erfolgen. Daher müssen die Indikatoren so gewählt werden, dass der Aufwand für Datenerhebung und Datenmanagement so gering wie möglich ausfällt. Dazu sollten vorhandene Datenbestände erschlossen und für die Leistungsbewertung verfügbar gemacht werden.

### 6.2 Bewertung der Leistungsindikatoren

Die in Kapitel 5 diskutierten Indikatoren können vor dem Hintergrund der im vorangegangenen Abschnitt formulierten Ansprüche an wissenschaftsadäquate Leistungsparameter bewertet werden:

#### *Nachwuchsförderung*

Als Indikatoren für die Nachwuchsförderung wurden die Anzahl der Promotionen und Habilitationen pro Kopf sowie die Anzahl der Promotionen bzw. Habilitationen relativ zur Promotions- bzw. Habilitationsquote diskutiert.



Findet die Leistungsbewertung innerhalb von relativ homogenen Fakultäten statt, können die Promotionen/Habilitationen pro Professor/in als Indikator dienen. Bei großen systematischen Unterschieden zwischen den Fächern ist die Standardisierung am Fachdurchschnitt zu empfehlen, damit der Indikator vergleichbar ist, d.h. wenn die Leistung unterschiedlicher Fakultäten verglichen wird.

### *Drittmittel*

Drei verschiedene Verfahren zur Verwendung von Drittmitteln als Leistungsindikator wurden diskutiert:

- (1) das Drittmittelaufkommen pro Kopf,
- (2) drittmittelfinanzierte Mitarbeiter/innenstellen und
- (3) Drittmittel pro Kopf relativ zum fachspezifischen Bundesdurchschnitt.

Welcher Indikator am besten geeignet ist, hängt vom Verwendungszweck ab. Werden nur Leistungen innerhalb von relativ homogenen Fakultäten (d.h. mit einer geringen Fächerdifferenzierung) verglichen, kann das absolute Drittmittelaufkommen verwendet werden.

Dieser Indikator hat, wie oben ausgeführt, allerdings den Nachteil große Disparitäten im Drittmittelaufkommen zwischen den Fächern nicht auszugleichen. Ein Weg dies zu tun, ist die Verwendung der Anzahl vollzeit-äquivalenter Mitarbeiter/innenstellen, die aus Drittmitteln finanziert sind. Auf diese Weise können Ausgaben für Geräte und Infrastruktur aus dem Drittmittelaufkommen herausgerechnet werden.

Allerdings ist die fachspezifische Einbettung über dieses Verfahren nur ansatzweise gewährleistet, weil die Notwendigkeit von Drittmitteln für die Durchführung von Projekten nach Fächern unterschiedlich ausfallen kann, was auch bei einer Korrektur um den Sachmittelbedarf nicht berücksichtigt wird.

Für eine fächervergleichende Interpretation des Drittmittelindikators wird deshalb empfohlen, die eingeworbenen Drittmittel pro Kopf relativ zum fachspezifischen Bundesdurchschnitt zu verwenden. Dazu stehen Daten von DESTATIS zur Verfügung, die für die betroffenen Fakultäten ausreichend differenziert sind. Der Vorteil dieses Indikators liegt in der fachspezifischen Einbettung auf einem niedrigen Aggregationsniveau. In einzelnen Fällen sind möglicherweise Korrekturen notwendig, wenn zum Beispiel eine Einordnung in die Fächersystematik von DESTATIS unklar ist. Werden die Daten relativ zum Fächerdurchschnitt von DESTATIS interpretiert, dann sollten alle Drittmittel in die Leistungsevaluation einbezogen werden. Eine Differenzierung nach Drittmittelgebern entspräche nicht der Standardisierungsbasis, weil DESTATIS alle Arten von Drittmitteln erfasst.

### *Publikationen*

Die Bewertung von Forschungsleistungen anhand von Publikationen ist komplexer als dies für die Nachwuchsförderung und Drittmittel der Fall ist. Zum einen werden in der Praxis zahlreiche unterschiedliche Verfahren angewendet, zum anderen bilden die einzelnen Verfahren noch stärker die jeweiligen Publikationskulturen der Fächer ab.

Das einfache Auszählen von Publikationen ist – wie dargestellt wurde – nicht-wissenschaftsadäquat und deshalb als Verfahren nicht geeignet, auch wenn kein Vergleich über Fächergrenzen hinweg stattfindet. Werden Publikationen in der Leistungsbewertung berücksichtigt, muss folglich eine Qualitätsgewichtung der einzelnen Publikationen stattfinden.

Drei Verfahren zur Qualitätsgewichtung von Publikationen wurden diskutiert:

- (1) die Bildung von Zeitschriftenlisten,
- (2) die Verwendung von Journal-Impact-Faktoren und
- (3) die Bibliotheksverbreitungsrate für die Bewertung von Buchpublikationen.

Letztere kann aufgrund des großen Aufwandes nicht adäquat erhoben werden und wird deshalb nicht empfohlen.

Gegen die Verwendung von Journal-Impact-Faktoren (ob absolut oder um den Fächermittelwert korrigiert, ist dafür unerheblich) spricht, dass in den Geistes-, Sozial- und Rechtswissenschaften eine Mehrzahl der Publikationen nicht in Fachzeitschriften erscheint. Eine Bewertung anhand von Impact-Faktoren würde daher eine große Zahl von Publikationen nicht erfassen können. Darüber hinaus existieren entsprechende Zeitschriften-datenbanken auch nicht für alle Fächer (vgl. Abschnitt 5.3.3).

Eine Qualitätsgewichtung von Publikationen kann deshalb nur listenbasiert erfolgen. Für die Wirtschaftswissenschaften bestehen Ranglisten für Fachzeitschriften (VHB-JOURQUAL, Handelsblatt-Ranking), die zu diesem Zweck eingesetzt werden können. Für die Geistes-, Sozial- und Rechtswissenschaften müssten fachspezifische Qualitätslisten erst entwickelt werden.

Die Leistungsevaluation anhand von Zitationen ist zwar direkter als die Qualitätsgewichtung der Publikationen, sie ist aus unterschiedlichen Gründen bis dato noch nicht praktikabel. Erstens liegen für Zitationsanalysen keine Datenbanken vor, die für alle betroffenen Fächer auch Buchpublikationen abdecken. Dies ist für die Geistes- und Sozialwissenschaften jedoch unverzichtbar, da die Mehrzahl der Publikationen nicht in Zeitschriftendatenbanken erfasst ist. Ähnlich ist dieser Aspekt auch für die Rechtswissenschaften zu bewerten, wobei hier auch hinsichtlich der Zeitschriftenpublikationen keine Datenbanken existieren. Zweitens stellt sich auch für Zitationsanalysen das Problem der fachspezifischen Einbettung der Indikatoren. Je größer ein Forschungsfeld ist, umso höher liegen die durchschnittlichen Zitationsraten. Die Zitationszahlen sagen daher zunächst mehr über die Größe des Forschungsfeldes als über die Qualität der wissenschaftlichen Arbeit. Eine Standardisierung der Zitationszahlen anhand von fachspezifischen Zitationsdurchschnitten ist aufgrund fehlender Daten nicht zu bewerkstelligen. Daher kann die Leistungsbewertung anhand von Zitationen nicht empfohlen werden.

Anders als für die Nachwuchsförderung und Drittmittel kann für Publikationen kein Indikator empfohlen werden, der einen Bezug zum fachspezifischen Publikationsdurchschnitt herstellt. Damit ist der Vergleich der Publikationsleistung zwischen den Fächern noch immer problematisch. Eine Normierung listenbasierter Qualitätsgewichte ist nicht umsetzbar.

### 6.3 Fakultätsspezifische Ausgestaltung der Forschungsevaluation

Ob die Empfehlungen für die Leistungsindikatoren in allen Fakultäten gleichermaßen umgesetzt werden können, hängt in erster Linie von praktischen Hindernissen ab. Insbesondere muss sichergestellt werden, dass alle Indikatoren in den Fakultäten aussagekräftig sind (besondere Begebenheiten also nicht zu Verzerrungen in der Leistungsevaluation führen) und die Datenlage eine angemessene Umsetzung der Indikatoren für alle Fächer zulässt.

Diese Fragen werden in den anschließenden Abschnitten für jede der drei Fakultäten einzeln diskutiert.



### 6.3.1 Philosophische Fakultät

Eine Befragung der Professoren/innen an der Philosophischen Fakultät hatte gezeigt, dass Forschungsevaluation den folgenden Ansprüchen genügen muss, um akzeptiert zu werden:<sup>30</sup>

- (1) Die Leistungsbewertung muss in erster Linie auf Grundlage der Publikationsaktivität erfolgen. Eine Einschränkung auf Fachzeitschriften mit Peer-Review-Verfahren, wie sie für die Naturwissenschaften und zum Teil für die Wirtschaftswissenschaften üblich ist, kann für die Fächer der Philosophischen Fakultät keine valide Messung der Leistung ergeben. Neben Fachzeitschriften müssen deshalb mindestens Monografien in der Evaluation berücksichtigt werden. Soweit sinnvoll umsetzbar, sollten auch weitere Publikationsformen in der Leistungsbewertung Berücksichtigung finden, wobei diskutiert werden muss, in welchem Verhältnis verschiedene Publikationsformen gewichtet werden.
- (2) Drittmittel sind in der Fakultät nur partiell als Maß wissenschaftlicher Leistungen anerkannt, was sich mit der kritischen Einschätzung von Drittmitteln als Indikator deckt, die zum Beispiel der Wissenschaftsrat abgibt. Werden Drittmittel dennoch zur Leistungsmessung verwendet, sind es vor allem Drittmittel der DFG und gemeinnütziger Stiftungen, die in die Bewertung einfließen sollten, da diese in einem kompetitiven, offenen Antragsverfahren vergeben werden und folglich keine Auftragsforschung darstellen.
- (3) Es wird kein zusammenfassender Summenscore über die verschiedenen Leistungsdimensionen (Publikationen, Drittmittel, Lehre) gewünscht. Vielmehr sollten die Dimensionen in Form eines wissenschaftlichen Tätigkeitsprofils getrennt ausgewiesen werden, wodurch u.a. den Besonderheiten einzelner Fächer (insbesondere der kleinen Fächer) in der Leistungsbewertung Rechnung getragen werden kann.

Insgesamt muss ein System der Leistungsbewertung folglich die verschiedenen Dimensionen wissenschaftlicher Arbeit abbilden, ohne dabei arbiträre Gewichtungen der einzelnen Dimensionen vorzunehmen.

Der *Nachwuchsförderung* wird als Indikator von den Professoren/innen der Philosophischen Fakultät mittlere Bedeutung zugeschrieben. Zwei mögliche Indikatoren wurden diskutiert: Erstens die Anzahl der Promotionen und Habilitationen pro Professor/in und zweitens die Anzahl der Promotionen und Habilitationen pro Professor/in relativ zur fachspezifischen Promotionsquote. Aufgrund der fachlichen Heterogenität in der Philosophischen Fakultät kann der erste Indikator nicht sinnvoll verwendet werden. Wird die Nachwuchsförderung als Leistungsmessungsparameter berücksichtigt, sollte deshalb eine Relativierung anhand der Promotions- bzw. Habilitationsquote vorgenommen werden.

Die Befragung der Philosophischen Fakultäten hat ergeben, dass *Drittmittel* als Indikator der Forschungsevaluation am weitesten verbreitet sind. Drei Möglichkeiten, Drittmittel als Indikator zu verwenden, wurden diskutiert: Drittmittel pro Professor/in, Anzahl der drittmittelfinanzierten Mitarbeiter/innenstellen pro Professor/in sowie Drittmittel pro Professor/in relativ zum fachspezifischen Durchschnitt.

Allerdings zeigen sich große Unterschiede in den durchschnittlichen Drittmittelaufkommen der Fächer, die an der Philosophischen Fakultät vertreten sind. Während in den allgemeinen Sprach- und Kulturwissenschaften bundesweit im Durchschnitt der Jahre 2008 und 2009 ca. 163.000 Euro Drittmittel pro Professor/in verausgabt wurden, ist dieser Betrag für die Anglistik und Romanistik mit ca. 25.000 Euro nur etwa ein Siebtel so hoch (siehe Tab. 7). Deshalb ist das Drittmittelaufkommen pro Kopf als Indikator nicht geeignet.

<sup>30</sup> Der Abschlussbericht zur Befragung ist diesem Bericht als Anlage beigefügt.

| Lehr- und Forschungsbereich                                    | Drittmittel je Professor/in (in 1.000 Euro) |        |                          |                           |
|--|---|--------|--------------------------|---------------------------|
|  | 2008  | 2009   | Durchschnitt (2008/2009) | Min/Max Rate <sup>a</sup> |
| Sprach- und Kulturwissenschaften allgemein                     | 135,91                                      | 190,34 | 163,12                   | 1,00                      |
| Philosophie  | 50,22                                       | 47,92  | 49,07                    | 0,30                      |
| Geschichte   | 85,14                                       | 96,70  | 90,92                    | 0,56                      |
| Allgemeine und vergleichende Literatur- und Sprachwissenschaft | 112,65                                      | 82,57  | 97,61                    | 0,60                      |
| Altphilologie (Klassische Philologie)                          | 45,19                                       | 46,12  | 45,65                    | 0,28                      |
| Germanistik (Deutsch, germanische Sprachen ohne Anglistik)     | 40,12                                       | 44,99  | 42,55                    | 0,26                      |
| Anglistik, Amerikanistik                                       | 23,66                                       | 22,58  | 23,12                    | 0,14                      |
| Romanistik   | 21,51                                       | 29,72  | 25,62                    | 0,16                      |
| Sonstige/ Außereuropäische Sprach- und Kulturwissenschaften    | 83,85                                       | 96,45  | 90,15                    | 0,55                      |
| Sozialwissenschaften   | 67,79                                       | 72,54  | 70,17                    | 0,43                      |

Anmerkungen: Quelle: DESTATIS 2010. Ausgewiesen sind nur Fächer, die an der Philosophischen Fakultät der HHU vertreten sind.

<sup>a</sup> Verhältnis des Fachdurchschnitts zum besten Fachdurchschnitt.

**Tabelle 7: Durchschnittliches Drittmittelaufkommen pro Professor/in für die Fächer der Philosophischen Fakultät**

Werden Drittmittel als Indikator für die Forschungsleistung verwendet, ist eine Gewichtung am bundesweiten Fächerdurchschnitt notwendig, wie sie in Abschnitt 5.3.5 vorgeschlagen wird. Allerdings erscheint eine Beschränkung auf kompetitive Drittmittel (DFG, Stiftungen) geboten, um die Qualität der geförderten Forschung sicherzustellen.

Die großen Disparitäten in den Durchschnittssätzen der Fächer können zu erheblichen Verzerrungen bei der Bewertung führen, weil es für bestimmte Fächer vergleichsweise einfach ist, ein überdurchschnittliches Drittmittelaufkommen zu erreichen. Um dem entgegenzuwirken, wird empfohlen, bei der Berechnung der Drittmittelquoten einen Mindestwert für die Fachdurchschnitte zu verwenden. Die Höhe dieser Grenze kann sich am Bedarf eines idealtypischen Sachmittelprojektes der DFG orientieren. Dabei werden in den Geistes- und Sozialwissenschaften im Regelfall Mittel für die Beschäftigung eines Doktoranden/innen (in Höhe von 65 Prozent einer Wissenschaftler/innenstelle der Tarifgruppe TV-L 13, d.h. 36.500 Euro) und Sachmittel (hier pauschal mit 15.000 Euro beziffert) beantragt. Die Untergrenze für die Bestimmung der Drittmittelquote würde in diesem Verfahren auf ca. 50.000 Euro festgelegt (und damit die Min/Max-Rate von drittmittelstärksten zum drittmittelschwächsten Fach auf etwa ein Drittel (0,31) festgelegt.

Ein zweites praktisches Problem ergibt sich aus der Zuordnung der kleinen Fächer der Fakultät (d.h. Jüdische Studien und Modernes Japan) zu den Fächerdurchschnitten. Dieses Problem kann gegebenenfalls anhand der Zuordnung zu Fachkollegien der DFG gelöst werden.



Publikationen sollen aus Sicht der Professoren/innen der Philosophischen Fakultät der wichtigste Indikator für die Forschungsevaluation sein. Eine Gewichtung an fächerspezifischen Durchschnittswerten für die Bundesrepublik wäre auch für Publikationen wünschenswert, ist aber angesichts der schwierigen Datenlagen nicht realisierbar. Der Einsatz gängiger bibliometrischer Kennzahlen wie des Journal-Impact-Faktor (JIF) und des „Source Normalized Impact per Paper“ (SNIP) ist jedoch kaum möglich, weil Buchpublikationen, insbesondere Monografien, neben Zeitschriftenaufsätzen eine herausragende Bedeutung für die Publikationskulturen der geistes- und sozialwissenschaftlichen Fächer haben.

Auch Zeitschriftenaufsätze können nur eingeschränkt anhand dieser Maße bewertet werden. Von den Zeitschriften, die in der Befragung der Professoren/innen als besonders wichtig für das jeweilige Fach genannt wurden sind, nur 70 Prozent in Scopus und 62 Prozent im Web of Science erfasst (vgl. Tab. 8). Je nach Fach unterscheiden sich diese Quoten erheblich. Für die Jüdischen Studien und das Moderne Japan erreicht die Abdeckung in der Scopus Datenbank gerade einmal 31 und 27 Prozent. Daher ist eine Qualitätsgewichtung von Zeitschriftenartikeln auf Grundlage von Impact-Faktoren nicht systematisch umsetzbar.<sup>31</sup>

| Institut                         | Scopus      | WoS <sup>a</sup> | Anzahl Zeitschriftentitel |
|----------------------------------|-------------|------------------|---------------------------|
| Anglistik und Amerikanistik      | 74 %        | 70 %             | 27                        |
| Germanistik                      | 82 %        | 65 %             | 17                        |
| Geschichtswissenschaften         | 76 %        | 66 %             | 29                        |
| Jüdische Studien                 | 31 %        | 54 %             | 13                        |
| Klassische Philologie            | 81 %        | 64 %             | 11                        |
| Kultur- und Medienwissenschaften | 43 %        | 29 %             | 14                        |
| Kunstgeschichte                  | 53 %        | 37 %             | 19                        |
| Modernes Japan                   | 27 %        | 27 %             | 11                        |
| Philosophie                      | 88 %        | 77 %             | 26                        |
| Romanistik                       | 57 %        | 71 %             | 19                        |
| Sozialwissenschaften             | 77 %        | 34 %             | 47                        |
| Sprache und Information          | 88 %        | 63 %             | 24                        |
| <b>Gesamt</b>                    | <b>70 %</b> | <b>62 %</b>      | <b>256</b>                |

<sup>a</sup> Web of Science (Science Citation Index, Social Science Citation Index und Arts & Humanities Citation Index)

**Tabelle 8: Datenbankreichweite für Zeitschriften nach Instituten der Philosophischen Fakultät**

<sup>31</sup> Denkbar ist lediglich, dass einzelne Institute Impact-Faktoren für eine fachspezifische Gewichtung der Publikationsqualität verwenden.

Eine Qualitätsgewichtung von Buchpublikationen z.B. anhand der absoluten oder relativen Bibliotheksverbreitungsrate ist bis dato noch nicht standardmäßig möglich und wird mittelfristig von den wichtigsten Datenbank Anbietern nicht bereitgestellt.<sup>32</sup> Der im „Web of Science“ eingeführte „Book Citation Index“ (BCI) erfasst nur ca. 30.000 englischsprachige Bände und ist daher für die Forschungsevaluation an deutschen Universitäten auch mittelfristig nicht brauchbar (und zudem mit hohen Anschaffungskosten verbunden).

Der einzig praktikable Weg, eine Qualitätsgewichtung für Publikationen umzusetzen, ist, Veröffentlichungen nach fachspezifischen Kriterien zu kategorisieren, wie es in Abschnitt 5.3.4 dargestellt wird.

| Beschreibung der Kategorien |   |
|-----------------------------|---|
| Kategorie A+                | Artikel in den besten zehn Zeitschriften eines Fachs (zwingend mit wissenschaftlicher Begutachtung und international herausragender Sichtbarkeit)   |
| Kategorie A                 | Beiträge von herausragender wissenschaftlicher Bedeutung: (begutachtete) Monografien mit mehr als 200 Seiten (ohne Lehrbücher und Einführungen); Aufsätze in führenden internationalen Fachzeitschriften                                      |
| Kategorie B                 | Beiträge von besonderer wissenschaftlicher Bedeutung: (begutachtete) Monografien mit 200 Seiten oder weniger, Lehrbücher, Einführungswerke; Aufsätze in führenden nationalen Fachzeitschriften  |
| Kategorie C                 | Beiträge von mittlerer wissenschaftlicher Bedeutung; nicht begutachtete Monografien, Aufsätze in sonstigen Fachzeitschriften mit Begutachtungsverfahren; Herausgeberschaft von Sammelbänden, begutachtete Beiträge in Herausgeber/innenbänden |
| Kategorie D                 | Sonstige Publikationen: Aufsätze in Zeitschriften ohne Begutachtungsverfahren, Beiträge zu Herausgeber/innenbänden ohne Begutachtung, Beiträge zur Popularisierung von Wissenschaft etc.  |

Tabelle 9: Vorschlag für Kategorien zur Bewertung von Publikationen an der Philosophischen Fakultät

Die Kategorien (A+, A, B, C, D) können fakultätsweit definiert werden. Ein Beispiel findet sich in Tabelle 9. Konkretisiert werden die Ansprüche an die Kategorien den fachspezifischen Gegebenheiten entsprechend durch die Institute. Eine akademische Begutachtung ist für Publikationen in der Kategorie A+ zwingend notwendig und in den Kategorien A und B ein wichtiges Kriterium. Die Ausgestaltung der Kategorisierung für Zeitschriftenartikel kann anhand einer Prestigerangliste erfolgen, in der jedes Institut die relevanten Fachzeitschriften jeweils einer Kategorie zuordnet. Nicht-wissenschaftliche Publikationen (Zeitungsartikel, Zeitschriftenartikel) und Beiträge zu Lexika werden für die Leistungsbewertung nicht berücksichtigt.

Anders als in den Naturwissenschaften und der Medizin gibt es in den Geistes- und Sozialwissenschaften keine „Ehrenautorschaften“. Findet die Forschungsevaluation auf Ebene der Professoren/innen statt, stellt sich die Frage, ob die Leistungen der wissenschaftlichen Mitarbeiter/innen in die Bewertung einbezogen werden müssen, um Verzerrungen beim Vergleich von Fächern zu vermeiden. Findet die Evaluation auf der Ebene der Professoren/innen/Institute statt, ist neben einer Korrektur um die Anzahl der wissenschaftlichen Mitarbeiter/innen auch eine Korrektur um tatsächlich geleistete Lehrveranstaltungen sinnvoll.

<sup>32</sup> Dr. Matthias Winterhager, an der Universität Bielefeld verantwortlich für den Aufbau einer deutschlandweiten Datenbank wissenschaftlicher Publikationen, verweist darauf, dass eine vollständige Erfassung von Monografien und Herausgeber/innenbänden in absehbarer Zeit nicht geleistet werden kann. Der Grund dafür ist unter anderem, dass die Universitäten und Technischen Hochschulen ihre Publikationen nicht systematisch erfassen, so dass ein Retrieval aus dezentralen Quellen nicht möglich ist.



### 6.3.2 Wirtschaftswissenschaftliche Fakultät

Die Bewertung der Anwendbarkeit der dargelegten Indikatoren zur Messung der Forschungsleistung auf Ebene der Wirtschaftswissenschaftlichen Fakultät hängt mit der Frage zusammen, ob diese Kriterien zur Bewertung der Forschungsleistung innerhalb der Fakultät (intrafakultär) oder zwischen verschiedenen Fakultäten (interfakultär) erfolgen soll. Da es hier zum Teil zu signifikanten Unterschieden kommen kann, wird im Folgenden jeweils zwischen der Anwendbarkeit auf intra- und interfakultärer Ebene unterschieden.

Hinsichtlich des Indikators *Nachwuchsförderung* wurden zwei mögliche Messansätze diskutiert: Anzahl der Promotionen und Habilitationen pro Professor/in sowie Anzahl der Promotionen und Habilitationen pro Professor/in relativ zur fachspezifischen Promotionsquote.

Da nicht davon auszugehen ist, dass zwischen der Volkswirtschafts- und der Betriebswirtschaftslehre signifikante Unterschiede hinsichtlich der Zahl der Promotionen und Habilitationen festzustellen sind, können die zwei aufgeführten Indikatoren aus Sicht der Wirtschaftswissenschaftlichen Fakultät zumindest für die intrafakultäre Messung der Forschungsleistung eingesetzt werden. Die interfakultäre Anwendbarkeit des Indikators „Anzahl der Promotionen und Habilitationen pro Professor/in“ wird jedoch angezweifelt, da die Daten des Statistischen Bundesamts darauf hindeuten, dass zum Teil erhebliche fächerspezifische Unterschiede zwischen den Fakultäten feststellbar sind. Vor diesem Hintergrund wird der Einsatz einer relativen Promotions- und Habilitationsquote empfohlen. Insbesondere bei der Habilitationsquote sollte jedoch berücksichtigt werden, dass auch – wie in Abschnitt 5.1 beschrieben – habilitationsäquivalente Leistungen wie zum Beispiel eine erfolgreiche Evaluation eines Junior-Professors/einer Junior-Professorin berücksichtigt werden.

Hinsichtlich des Indikators *Drittmittel* wurden drei mögliche Messansätze diskutiert: Drittmittel pro Professor/in, Anzahl drittmittelfinanzierte Mitarbeiter/innenstellen pro Professor/in sowie Drittmittel pro Professor/in relativ zum fachspezifischen Durchschnitt.

Aus Sicht der Wirtschaftswissenschaftlichen Fakultät können diese drei Ansätze für die intrafakultäre Messung der Forschungsleistung eingesetzt werden, da wiederum von keinen signifikanten Unterschieden zwischen der Volkswirtschaft und der Betriebswirtschaft ausgegangen wird. Problematischer ist es hinsichtlich eines interfakultären Vergleichs der Forschungsleistung. Daten des Statistischen Bundesamts zeigen, dass insbesondere naturwissenschaftliche Fächer im Durchschnitt signifikant höhere Drittmittelinwerbungen aufweisen als die Wirtschaftswissenschaften, was zum Teil auch durch die Anschaffung kostenintensiver Geräte und Labormaterialien erklärt werden kann. Dies hat zur Folge, dass Drittmittel pro Professor/in kein geeigneter Indikator für einen interfakultären Vergleich ist. Auch die Anzahl der drittmittelfinanzierten Mitarbeiter/innenstellen pro Professor/in ist aus Sicht der Wirtschaftswissenschaftlichen Fakultät kein geeigneter Indikator für einen interfakultären Vergleich. Zwar werden in diesem Fall keine Geräte bzw. Labormaterialien in den Drittmitteln erfasst, jedoch können dennoch signifikante Fächerunterschiede festgestellt werden.

Die Empfehlung der Wirtschaftswissenschaftlichen Fakultät zur interfakultären Messung der Forschungsleistung ist eine Gewichtung der eingeworbenen Drittmittel pro Professor/in am fachspezifischen Bundesdurchschnitt. Durch die Nutzung dieses Berechnungsansatzes werden Professoren/innen unmittelbar mit ihren Peers verglichen und ins Verhältnis gesetzt, was ihre Leistung aus Sicht der Wirtschaftswissenschaftlichen Fakultät besser abbildet.

Zuletzt kann hinsichtlich des Indikators *Publikationen* festgestellt werden, dass die drei vorgeschlagenen Messansätze an der Wirtschaftswissenschaftlichen Fakultät nur zum Teil für die Messung der intrafakultären Forschungsleistung eingesetzt werden können.

Eine listenbasierte Qualitätsgewichtung von Zeitschriftenpublikationen ist in beiden wirtschaftswissenschaftlichen Fächern möglich; so existiert mit dem Handelsblatt-Zeitschriftenranking eine etablierte Zeitschriftenliste für die Volks- und Betriebswirtschaft, die zur Messung der intrafakultären Forschungsleistung genutzt werden kann. Darüber hinaus könnte in der Betriebswirtschaftslehre auch auf die VHB-JOURQUAL Zeitschriftenliste des Verbands der Hochschullehrer für Betriebswirtschaft zurückgegriffen werden. Hierbei ist jedoch zu berücksichtigen, dass Fächer wie beispielsweise die betriebliche Steuerlehre eine abweichende Publikationskultur aufweisen, die berücksichtigt werden muss.

Demgegenüber ist eine listenbasierte Qualitätsgewichtung von Monografien mit größeren Schwierigkeiten verbunden. Da jedoch auch Monografien nur qualitätsgewichtet zur Messung der Forschungsleistung herangezogen werden sollten, empfiehlt die Arbeitsgruppe, ein Monografienranking basierend auf dem Verlag, in welchem die Monografie erschienen ist, zu erstellen. Weiterhin sollte bei der Qualitätsbeurteilung zwischen Fachbüchern einerseits und Lehrbüchern und Einführungswerken andererseits differenziert werden.

Ein interfakultärer Vergleich der Publikationsleistung auf Grundlage von Zeitschriftenlisten ist schwierig durchzuführen. Ein Ansatz könnte es sein, Publikationen in verschiedene Qualitätsgruppen (z. B. A+, A, B usw.) einzuteilen. Zeitschriftenlisten könnten eine Grundlage für diese Einteilung sein, jedoch sollten auch in dem Fall Monografien berücksichtigt und deren Qualität zumindest näherungsweise bestimmt werden.

Während eine listenbasierte Qualitätsgewichtung von Publikationen in den Wirtschaftswissenschaften möglich ist, können die beiden anderen Verfahren nicht empfohlen werden. Die Gewichtung von Publikationen mit Journal-Impact-Faktoren kann nicht empfohlen werden, da insbesondere deutsche Zeitschriften sowohl in Scopus als auch im SSCI nur unsystematisch erfasst werden. In der Betriebswirtschaftslehre sind jedoch Publikationen in deutschsprachigen Zeitschriften anerkannt, weshalb diese nicht ignoriert werden können. Darüber hinaus sind auch Monografien in weiten Teilen der Betriebswirtschaftslehre verbreitet und für die Forschung von hoher Relevanz. Diese könnten jedoch bei einer Gewichtung mit Journal-Impact-Faktoren nicht weiter berücksichtigt werden. Aufgrund dieser Problematik, die den Einsatz des Indikators auf intrafakultärer Ebene verhindert, kann auch eine Nutzung für interfakultäre Vergleiche nicht empfohlen werden.

Zitationsanalysen weisen ähnliche Probleme auf wie die Gewichtung von Publikationen mit Journal-Impact-Faktoren. Da für Artikel in deutschen Zeitschriften keine systematischen Zitationsdaten erhoben werden, würden diese in der Analyse nicht berücksichtigt werden, was vor dem Hintergrund der Qualität der Publikationen in den Zeitschriften nicht vertretbar ist. Auch Monografien würden bei Zitationsanalysen nicht berücksichtigt werden können, obwohl auch sie ein Ausdruck der betriebs- und volkswirtschaftlichen Forschung sind. Vor diesem Hintergrund muss von diesem Indikator Abstand genommen werden. Dies bezieht dabei nicht nur auf eine Nutzung für intrafakultäre Vergleiche, sondern auch für interfakultäre Vergleiche. Zu letzteren kommt verstärkend hinzu, dass Zitationsquoten stark fachabhängig sind und sich zwischen Medizin, Naturwissenschaften und Wirtschaftswissenschaften zum Teil erheblich unterscheiden.

### 6.3.3 Juristische Fakultät

Die vorgestellten Indikatoren müssen hinsichtlich ihrer Anwendbarkeit als Kriterien zur Forschungsevaluation in der Juristischen Fakultät mehrdimensional differenziert betrachtet werden. Die Geeignetheit der kriterienspezifischen Umsetzungsmodelle hängt dabei einerseits von fachkulturellen Gegebenheiten, andererseits von der bezogenen Vergleichsgruppe und mithin von den konkreten Evaluationszielen ab. Dies erfordert eine Bewertung der Modelle sowohl in Bezug auf ihre intrafakultäre als auch hinsichtlich ihrer interfakultären Anwendbarkeit.



Das Kriterium der *Nachwuchsförderung* kann in zwei Modellen umgesetzt werden (vgl. Abschnitt 5.1.): Zum einen können abgeschlossene Promotionen und Habilitationen je Professor/in absolut gezählt werden, zum anderen kann dieser Wert in Relation zur fachspezifischen Promotions- bzw. Habilitationsquote gesetzt werden. Beide Modelle sollten sowohl interne als auch externe Promovenden/innen bzw. Habilitanden/innen umfassen. Im Rahmen der Habilitationsquote sind auch Habilitationsäquivalente – wie beispielsweise die erfolgreiche Evaluation eines Juniorprofessors/einer Juniorprofessorin (vgl. Abschnitt 5.1) – zu berücksichtigen. Für die intrafakultäre Vergleichbarkeit von Forschungsleistungen anhand der Nachwuchsförderung kommt eine Anwendbarkeit beider Modelle in Betracht. Zwischen den verschiedenen Rechtsgebieten (Zivil-, Straf- und Öffentliches Recht) bestehen insofern keine signifikanten Unterschiede, die zu einer Inkomparabilität der absoluten Werte führen würden.

Hinsichtlich der interfakultären Vergleichsebene bestehen hingegen Bedenken. Die Daten des Statistischen Bundesamts zeigen, dass aufgrund divergierender Fächerkulturen im Bereich der Nachwuchsförderung erhebliche Abweichungen der absoluten Werte auftreten. Die Kriterienanwendbarkeit beschränkt sich daher für diese Vergleichsebene auf das Modell einer fachspezifisch relativierten Evaluation.

Die Höhe der eingeworbenen *Drittmittel* kann als Kriterium zur Forschungsevaluation in drei Modellen angewendet werden (vgl. Abschnitt 5.2): Erstens können Drittmittel pro Professor/in absolut gewertet werden. Zweitens kann eine Relativierung durch die Reduzierung auf drittmittelfinanzierte Mitarbeiter/innenstellen pro Professor/in erfolgen. Drittens kann der absolute Wert in Relation zum fachspezifischen Durchschnitt gesetzt werden.

Für die Evaluation der Forschungsleistung innerhalb der Juristischen Fakultät ist die Anwendbarkeit aller Modelle gegeben. Auch in Bezug auf dieses Kriterium führen die zwischen den Rechtsgebieten auftretenden minimalen Divergenzen nicht zu Inkomparabilitäten.

Anders ist auch hier die Kriterienanwendbarkeit auf interfakultärer Ebene zu bewerten. Den Daten des Statistischen Bundesamts ist zu entnehmen, dass medizinische und naturwissenschaftliche Fächergruppen aufgrund der um ein Vielfaches höheren Materialkosten im absoluten Vergleich deutlich oberhalb der von juristischen Fakultäten eingeworbenen Drittmittel liegen. Auch eine Anwendung des durch die Beschränkung auf drittmittelfinanzierte Mitarbeiter/innenstellen pro Professor/in relativierten Wertes scheidet aus Sicht der Juristischen Fakultät aus. Zwar ist er bei diesem Modell um die materialkostenbedingte Diskrepanz bereinigt, doch zeigen sich infolge der fächerspezifischen Forschungskulturen nach wie vor erhebliche Abweichungen. Für den interfakultären Vergleich anhand des Kriteriums eingeworbener Drittmittel ist daher im Ergebnis allein das Modell einer fachspezifischen Gewichtung der je Professor/in eingeworbenen Drittmittel geeignet. Nur die Relativierung innerhalb der fachlichen Referenzgruppe gewährleistet hier eine echte Komparabilität.

Vergleichbar mit den Gegebenheiten in der Philosophischen Fakultät sind *Publikationen* auch in den Rechtswissenschaften der wichtigste Indikator für die Forschungsleistung. Die Forschungsevaluation anhand von Publikationen kann nach den Ausführungen in Abschnitt 5.3 in drei verschiedenen Modellen erfolgen: Erstens können Publikationen anhand fächerspezifisch erstellter Ranglisten bewertet werden. Zweitens lassen sich Aufsätze über Journal-Impact-Faktoren bewerten. Drittens kann die Forschungsleistung im Bereich der Publikationen über Zitationsanalysen erfolgen.

Dieses Kriterium ist auf rechtswissenschaftliche Publikationen derzeit nicht anwendbar und daher weder für intra- noch für interfakultäre Vergleiche gegeben. Alle Modelle setzen entsprechende Ranglisten oder Datenbanken für die Bewertung von Publikationen voraus. Ranglisten, wie sie etwa die Betriebswirtschaft

in Form des VHB-JOURQUAL-Rankings kennt, existieren für die Rechtswissenschaften nicht. Ebenso gibt es keine Datenbanken, aus denen sich Journal-Impact-Faktoren oder Zitationsdaten für rechtswissenschaftliche Zeitschriften entnehmen lassen. Dies hat folgende Gründe: Trotz des zunehmenden Einflusses europäischer Rechtssetzungsprozesse auf das nationale Recht der EU-Mitgliedstaaten, sind das Recht und mit ihm auch die Rechtswissenschaften nach wie vor in erheblichem Umfang national geprägt. Die jeweilige Wissenschaftsgemeinde führt ihren Diskurs nicht nur fast ausschließlich im nationalen Rahmen, sie führt ihn auch in einem nach Rechtsgebieten, Teilrechtsgebieten und Untergebieten aufgefächerten Literatur- und Zeitschriftenspektrum, in dem abhängig von Themenschwerpunkten und Adressaten/innenkreisen selektiv publiziert wird. Für die Analyse juristischer Fachzeitschriften zur Ermittlung eines Journal-Impact-Faktors fehlt es somit schon an einer einheitlichen Publikationskultur.

Hinzu kommt, dass es in den Rechtswissenschaften keine Zeitschriften gibt, die echte Peer-Review-Verfahren vorschalten. Mithin fehlt es auch an einem Qualitätssurrogat als Kriterium für die Zeitschriftenauswahl. Diese Ergebnisse sind uns im persönlichen Gespräch von Dr. Henk Moed (Elsevier) bestätigt worden, der an der Entwicklung der Scopus-Datenbank (vgl. Abschnitt 5.3.2) beteiligt ist. Zu den gleichen Resultaten kam der Bibliometrieforscher Dr. Matthias Winterhager vom Institut für Technik- und Wissenschaftsforschung (Universität Bielefeld) im Rahmen des Workshops am 13. März 2012.

Neben der differenzierten Publikationskultur weisen die Rechtswissenschaften zudem besondere Publikationsarten auf, deren wissenschaftliche Qualität mit quantitativen Verfahren nicht adäquat berücksichtigt werden kann. Nach wie vor kommt hier den Monografien, sei es als Dissertations- oder Habilitationsschriften oder als themenspezifische Veröffentlichungen etwa in einer bestimmten Verlagsreihe, ein hoher Stellenwert zu. Der für die Wirtschaftswissenschaften in Abschnitt 5.3 beschriebene Rückgang der Bedeutung von Monografien kann in den Rechtswissenschaften nicht festgestellt werden. Im Zusammenhang mit den Monografien sind freilich auch weitere Arten der Buchpublikationen, wie beispielweise Lehrbücher, Festschriften und Einzelkommentare zu berücksichtigen. Insbesondere die Kommentarliteratur ist ein rechtswissenschaftliches Publikationsspezifikum von enormer wissenschaftlicher Bedeutung. Die Mitarbeit an einem angesehenen Gesetzes- oder Vertragskommentar kann für die Reputation eines Rechtswissenschaftlers/einer Rechtswissenschaftlerin gewichtiger sein als eine ganze Reihe verschiedener Aufsätze in Fachzeitschriften. Das Bedeutungsverhältnis von Monografien und Kommentaren, aber auch von Kommentaren und Zeitschriften bedürfte daher einer wissenschaftsadäquaten Erfassung durch Kennzahlen, die bisher nicht erfolgt. Darüber hinaus werden in den Fachzeitschriften Rezensionen und Urteilsbesprechungen publiziert, die ebenfalls keiner pauschalen Gewichtung zugänglich sind. Auch in diesem Bereich können einzelne Beiträge den wissenschaftlichen Erkenntniswert beispielsweise eines Aufsatzes übersteigen.

Die Evaluation rechtswissenschaftlicher Forschungsleistungen anhand von Publikationen ist daher mit den derzeit zur Verfügung stehenden Mitteln nicht in einer wissenschaftsadäquaten Weise umsetzbar. Gleichwohl ist eine systematische Erfassung der Publikationsleistung erstrebenswert, da sich gerade darin die fachspezifische Forschungsleistung widerspiegelt.

Solange dies nicht möglich ist, sollte eine an der Forschungsleistung orientierte Mittelverteilung anhand der anderen Kriterien (Nachwuchsförderung/Drittmittel) nur in dem begrenzten Umfang erfolgen, der durch diese Kriterien im Verhältnis zum Hauptkriterium der Publikationsleistung abgebildet werden kann.

Eine Übersicht über die fakultätsspezifische Anwendbarkeit der verschiedenen Verfahrensvorschläge für die drei Indikatoren *Nachwuchsförderung*, *Drittmittel* und *Publikationen* findet sich in Tabelle 10.



|   | Phil. Fak. | WiWi | Jura |
|---|------------|------|------|
| <b>Nachwuchsförderung</b>   |            |      |      |
| Anzahl der Promotionen und Habilitationen pro Kopf  | /          | ✓    | ✓    |
| Anzahl der Promotionen und Habilitationen pro Kopf relativ zur fachspezifischen Promotionsquote | ✓          | ✓    | ✓    |
| <b>Drittmittel</b>  |            |      |      |
| Drittmittel pro Kopf  | /          | ✓    | ✓    |
| Anzahl der drittmittelfinanzierten Mitarbeiter/innenstellen                                     | ✓          | ✓    | ✓    |
| Drittmittel pro Kopf relativ zur fachspezifischen Promotionsquote                               | ✓          | ✓    | ✓    |
| <b>Publikationen</b>  |            |      |      |
| Listenbasierte Qualitätsgewichtung der Publikationen  | ✓          | ✓    | /    |
| Gewichtung mit Journal-Impact-Faktoren  | /          | /    | /    |
| Zitationsanalysen   | /          | /    | /    |

Tabelle 10: Anwendbarkeit der Verfahrensvorschläge in der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät

## 7. VORAUSSETZUNGEN ZUR EINFÜHRUNG EINER FORSCHUNGSEVALUATION AN DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

Im vorangegangenen Kapitel wurde detailliert erläutert, mit welchen Indikatoren die Forschungsleistung in der Philosophischen, Wirtschaftswissenschaftlichen und Juristischen Fakultät adäquat erfasst und bewertet werden kann. Bevor diese Indikatoren jedoch konkret zur Forschungsevaluation verwendet werden können, sind aus Sicht der Arbeitsgruppe noch vier zentrale Voraussetzungen zu erfüllen. *Erstens* sollte vor dem Aufbau eines Systems zur Forschungsevaluation zunächst definiert werden, zu welchen Zwecken ein solches System zukünftig genutzt werden soll. Nach der Festlegung der Anwendungsgebiete muss *zweitens* entschieden werden, ob die verwendeten Kriterien im Rahmen eines Systems zur Forschungsevaluation zu einem Forschungsprofil zusammengefasst oder zu einem Forschungsscore verdichtet werden sollen. *Drittens* muss vor der eigentlichen Einführung des Systems zur Forschungsevaluation die bestehende Infrastruktur an der HHU erweitert werden, damit die erforderlichen Daten in geeigneter Weise bereit gestellt werden können. Schließlich ist es *viertens* aus Sicht der Arbeitsgruppe dringend erforderlich, nach der Einführung eines Systems zur regelmäßigen Messung der Forschungsleistung, dieses im Rahmen eines regelmäßig stattfindenden Audits auf seine Effektivität und Effizienz hin zu überprüfen. In den nachfolgenden Abschnitten werden diese vier Anforderungen detailliert dargestellt.

### 7.1 Definition der Anwendungsgebiete der Forschungsevaluation

Bevor eine Entscheidung bezüglich der zur Forschungsevaluation maßgeblichen Indikatoren getroffen wird, sollte zunächst entschieden werden, für welche Zwecke die Messung der Forschungsleistung eingesetzt werden soll. Dies erscheint aus zwei Gründen sinnvoll. Zum einen führt die vorherige Definition der Anwendungsgebiete zu einer erheblichen Reduktion der mit der Forschungsevaluation verbundenen Unsicherheit. Es ist davon auszugehen, dass bei der Einführung eines Systems zur regelmäßigen Beurteilung der Forschungsleistung mit nicht unerheblichem intrafakultärem Widerstand zu rechnen ist. Dieser dürfte um ein Vielfaches höher sein, wenn ex ante nicht klar ist, wozu das System zur Forschungsevaluation letztlich genutzt werden wird, über das die Fakultätsmitglieder gegebenenfalls zu entscheiden haben. Wenn jedoch die Anwendungsgebiete vorab definiert und transparent kommuniziert werden, so dürfte dies die Widerstände gegen ein solches System erheblich reduzieren. Zum anderen hängt die Frage, welche Kriterien zur Messung der Forschungsleistung geeignet sind, auch davon ab, wozu die Forschungsevaluation genutzt werden soll. Ist es beispielsweise das Ziel, die Forschungsevaluation zur leistungsorientierten Mittelvergabe zwischen Fakultäten zu nutzen, so können zur Messung der Forschungsleistung nur solche Kriterien herangezogen werden, die eine fachbereichsübergreifende Vergleichbarkeit erlauben. Soll die Forschungsevaluation hingegen zur Mittelvergabe innerhalb einer Fakultät verwendet werden, so ist es nicht erforderlich, eine fakultätsübergreifende Vergleichbarkeit zu gewährleisten.

Wie sich im Rahmen unserer Befragung der Universitäten und Fakultäten in Deutschland gezeigt hat, kann die Forschungsevaluation für eine Reihe von Zwecken genutzt werden, die sich hinsichtlich der Aggregationsstufe der zu bewertenden Beobachtungseinheit unterscheiden.

Auf der niedrigsten Aggregationsstufe kann die Forschungsevaluation dazu verwendet werden, die Forschungsleistung einzelner Personen zu beurteilen. Beispielhafte Anwendungsgebiete wären hier die Bewertung von Bewerbern und Bewerberinnen im Rahmen von Berufungsverfahren, der Einsatz im Rahmen von Berufungsverhandlungen, die Bewertung von Juniorprofessoren/innen, die Bewertung von Professoren/innen im Rahmen von Bleibeverhandlungen, die Bewertung von Professoren/innen im Rahmen der Vergabe besonderer Leistungszulagen der W-Besoldung sowie die Vergabe interner Preise und Auszeichnungen.



Auf der nächst höheren Aggregationsebene könnte die Forschungsleistung von Lehrstühlen oder Instituten beurteilt werden. Beispielhafte Anwendungsfelder wären hier die leistungsorientierte Mittelvergabe innerhalb einer Fakultät oder ein Benchmarking der verschiedenen Institute/Lehrstühle einer Fakultät. Auf der obersten Stufe könnte die Forschungsleistung einer gesamten Fakultät evaluiert werden. Zu den Anwendungsgebieten auf dieser Ebene zählt die leistungsorientierte Mittelvergabe zwischen den verschiedenen Fakultäten einer Universität. Weiterhin könnte die Forschungsevaluation bei hochschulpolitischen Entscheidungen eine Rolle spielen, beispielsweise wenn es um die Frage geht, welche Fachbereiche einer Fakultät zukünftig stärker ausgebaut werden sollen. Unabhängig von der Aggregationsstufe kann die Forschungsevaluation auch zur Außendarstellung der Universität verwendet werden, etwa wenn es darum geht, die Politik oder die interessierte Öffentlichkeit über die Attraktivität einzelner Forschungscluster sowie die Forschungsleistung der Universität insgesamt zu informieren oder potentielle Kooperationspartner aus der Wirtschaft für eine Zusammenarbeit zu gewinnen.

## 7.2 Zusammenfassung der Indikatoren zur Messung der Forschungsleistung: Forschungsprofil oder Forschungsscore

Nachdem die Anwendungsgebiete der Forschungsevaluation festgelegt und geeignete Kriterien zur Erfassung der Forschungsleistung definiert sind, stellt sich die Frage, wie die Kriterien miteinander verknüpft werden. Wie die Ergebnisse unserer Befragung zeigen, gibt es hierzu im Wesentlichen zwei Möglichkeiten. Die Indikatoren können entweder in einem Forschungsprofil zusammengefasst oder zu einem Forschungsscore verdichtet werden.

Bei der Verwendung eines Forschungsprofils stellt jeder der verwendeten Indikatoren eine eigene Dimension der Forschungsleistung dar. Hierbei ist es nicht erforderlich, Leistung in jeder Dimension durch einen exakten Punktwert zu beschreiben (kardiale Skalierung), sondern es ist ausreichend, die Forschungsleistung ordinal zu beschreiben, beispielsweise in den Kategorien „gering“, „mittel“ und „hoch“. Weiterhin erfolgt keine unterschiedliche Gewichtung der einzelnen Dimensionen, sondern sie stehen vielmehr gleichberechtigt nebeneinander.

Die Verwendung eines Forschungsprofils bietet eine Reihe von Vorteilen. Da die Leistungen in den verschiedenen Dimensionen des komplexen Konstrukts „Forschungsleistung“ explizit ausgewiesen werden, sind die spezifischen Stärken und Schwächen einer Fakultät klar erkennbar. Weiterhin ist es möglich, spezifische Anreize zur Leistungssteigerung in einzelnen Dimensionen der Forschungsleistung zu setzen. So könnte beispielsweise im Rahmen der leistungsorientierten Mittelvergabe zwischen den Fakultäten der Vergaberahmen auf die einzelnen Dimensionen des Forschungsprofils aufgeteilt werden. Sollen nun Anreize geschaffen werden, die Leistung im Bereich *Publikationen* zu erhöhen, so könnte einfach das Volumen der in dieser Dimension zu vergebenden Mittel erhöht werden, während die Mittel in den übrigen Dimensionen verringert werden. Auf der individuellen Ebene könnten entsprechende Anreize beispielsweise über Zielvereinbarungen bezüglich der zu erbringenden Leistung in den verschiedenen Dimensionen gesetzt werden. Schließlich liegt ein weiterer Vorteil der Verwendung eines Forschungsprofils darin, dass eine relative Gewichtung der verwendeten Kriterien untereinander – und damit eine Entscheidung darüber, ob die Einwerbung von Drittmitteln wichtiger (und falls ja, wie viel wichtiger) ist als die Veröffentlichung von Büchern und Aufsätzen – nicht erforderlich ist. Da aus theoretisch-normativer Sicht keine Entscheidung für eine bestimmte Gewichtung der Indikatoren abgeleitet werden kann, ist die Festlegung der relativen Bedeutung der Indikatoren der Forschungsleistung eine rein hochschulpolitische Entscheidung. Folglich können die Gewichte der Indikatoren nur im Rahmen eines fakultätsinternen Verhandlungsprozesses bestimmt werden, wobei zu erwarten ist, dass eine solche Entscheidung sowohl sehr kontrovers als auch überaus zeitaufwendig ist.

Eine Alternative zum Forschungsprofil stellt der Forschungs-Score dar. Hierbei wird die Leistung für jeden einzelnen Indikator mit einem Punktwert versehen (kardinale Skalierung) und jedem Indikator ein Gewicht zwischen 0 und 1 zugewiesen, wobei die Summe der Indikatoren Gewichte 1 ergibt. Der Forschungs-Score ergibt sich dann durch Multiplikation der Punkte für jeden Indikator mit dessen Gewicht und anschließendem Aufsummieren.

Der Vorteil dieses Verfahrens liegt darin, dass die Forschungsleistung anhand einer konkreten Zahl dargestellt werden kann, wodurch auf einen Blick ein Vergleich der Forschungsleistung zwischen verschiedenen Beobachtungseinheiten möglich ist. Dem steht jedoch der Nachteil gegenüber, dass eine Gewichtung der verschiedenen Indikatoren erfolgen muss, was – wie bereits erläutert – in der praktischen Umsetzung mit erheblichen Problemen verbunden sein dürfte.

Auch lässt sich argumentieren, dass durch die Existenz einer einzigen Kennzahl der Forschungsleistung eine Vergleichbarkeit lediglich suggeriert wird, die bei näherer Betrachtung allerdings nicht gegeben ist. So kann es beispielsweise sein, dass eine Fakultät in zwei von drei Dimensionen eine überdurchschnittlich hohe Leistung bringt, in der dritten Dimension jedoch eine stark unterdurchschnittliche, wodurch sich möglicherweise nur ein Forschungs-Score in durchschnittlicher Höhe ergibt. Die Forschungsleistung dieser Fakultät würde somit genauso bewertet, wie die einer Fakultät, die in allen drei Kriterien lediglich durchschnittliche Leistungen erbringt und somit lediglich in einer Dimension bessere, in den beiden übrigen Dimensionen hingegen schlechtere Leistungen erbringt. Andererseits kann zur Vermeidung solcher Diskrepanzen gerade das Modell des Forschungs-Scores gezielt als Steuerungsinstrument eingesetzt werden. So können die in den einzelnen Dimensionen erreichbaren Scores gedeckelt werden. Werden beispielsweise drei Dimensionen bewertet, so kann der jeweils erreichbare Maximal-Score auf 25 festgesetzt werden. Somit kann ein Gesamtwert von 75 nur über eine gleichstarke Performance in allen Bereichen erzielt werden. Mit dieser Modifikation gehen Wertdeckelungen innerhalb der einzelnen Dimensionen einher. Ist beispielsweise eine Dimension durch die Höhe der eingeworbenen Drittmittel geprägt, so wird hier ein score-relevanter Höchstwert von beispielsweise 500.000 Euro absolut oder ein fachspezifisch-relativierter Drittmittelwert eingesetzt. Somit bleibt der maximal erreichbare Score für diese Dimension ab diesem eingeworbenen Betrag bzw. ab dem erreichten Wert gleich. Es wird daher für die Gesamtperformance interessanter, sich auch den anderen Dimensionen zu widmen.

### 7.3 Erweiterung der bestehenden Infrastruktur

Eine dritte Voraussetzung zur erfolgreichen Einführung eines Systems zur Forschungsevaluation ist die Erweiterung der derzeit bestehenden Infrastruktur, um zu gewährleisten, dass die benötigten Daten in geeigneter Form vorliegen. Die bestehende Infrastruktur sollte dergestalt erweitert werden, dass alle für die Leistungsparameter relevanten Informationen in einem System vorgehalten und ausgewertet werden können.

*Erstens* sollte das bestehende Forschungsinformationssystem erweitert und universitätsweit eingeführt werden, um eine einheitliche Erfassung aller *Publikationen* von Angehörigen der HHU zu ermöglichen. Allerdings ist eine zentrale Eintragung aller Publikationen durch einen/eine hierfür zuständigen Mitarbeiter/zuständige Mitarbeiterin zu fehleranfällig und aufwendig. Daher erscheint es sinnvoll, dass die Forscher/innen ihre Publikationen selbst in das Informationssystem eintragen. Um die Eintragung zu erleichtern und die Gefahr von Eingabefehlern zu minimieren, sollte hierzu eine web-basierte Eingabemaske bereitgestellt werden, die jeder Mitarbeiter/jede Mitarbeiterin über seinen/ihren Arbeitsplatz aufrufen kann. Weiterhin sollte jeder Mitarbeiter/jede Mitarbeiterin bei der Eintragung einer Publikation diese mittels einer Auswahlliste einer Publikationsart zuordnen. Dabei sollten folgende Kategorien zur Auswahl stehen: Monografien, Lehrbücher,



Artikel in Peer-Reviewed-Zeitschriften, Artikel in Zeitschriften ohne Peer-Review, Artikel in Sammelbänden und Festschriften, Urteilsbesprechungen, (Gesetzes-)Kommentare sowie Konferenzbeiträge. Im Idealfall findet unmittelbar eine Zuordnung zu den Kategorien statt, anhand derer eine listenbasierte Qualitätsgewichtung durchgeführt wird (siehe Abschnitt 6.2).

Weiterhin sollte geprüft werden, inwiefern das Forschungsinformationssystem Schnittpunkte zu Typo3 aufweist, damit möglicherweise die Daten direkt vom Forschungsinformationssystem in Typo3 übertragen werden können. Da viele Lehrstühle Publikationen auf ihrer Homepage veröffentlichen, könnte ein Datenaustausch zwischen dem Forschungsinformationssystem und Typo3 zu einer Erleichterung führen, was wiederum die Akzeptanz der Eintragung erhöhen könnte. Darüber hinaus sollte explizit kommuniziert werden, dass nur Publikationen, die im Forschungsinformationssystem eingetragen wurden, auch in der leistungsorientierten Mittelvergabe berücksichtigt werden. Dies könnte einen disziplinierenden Effekt haben und den Anreiz zur Eintragung der Publikationen in das Forschungsinformationssystem erhöhen.

*Zweitens* sollten eingeworbene *Drittmittel* im Forschungsinformationssystem erfasst werden. Dabei ist darauf zu achten, neben der Höhe der eingeworbenen Mittel auch Mittelgeber und die Projektlaufzeit/Förderzeitraum zu erfassen, um die Drittmittel über den Förderzeitraum verteilen und hierdurch die tatsächlich eingeworbenen Drittmittel pro Jahr bestimmen zu können. Da Projekte zum Teil nur wenige Monate laufen können, wird empfohlen, die Dauer in Monaten zu erfassen. Die dadurch erhobenen „Drittmittel pro Monat“ können anschließend problemlos zu „Drittmittel pro Jahr“ aufsummiert werden.<sup>33</sup>

*Drittens* sollten die Dekanate alle Angaben zur *Nachwuchsförderung* erheben und über das Forschungsinformationssystem verfügbar machen. Das gilt für Promotionen und Habilitationen genauso wie für die habilitationsäquivalenten Leistungen, insbesondere W2-/W3-Rufe und positive Evaluationen einer Juniorprofessur.

Langfristig ist es anzustreben, neben den Leistungsparametern der Forschungsevaluation auch Indikatoren für weitere Leistungen im Forschungsinformationssystem vorzuhalten. Dabei sind insbesondere Parameter für Leistungen in der Lehre (Anzahl betreuter Abschlussarbeiten, Ergebnisse der Lehrevaluation) und die Mitarbeit in der akademischen Selbstverwaltung (Mitarbeit in Ausschüssen und Berufungskommissionen, Tätigkeit als Dekan/in, Studiendekan/in, Vorsitzender/e des Prüfungsamts, Gleichstellungsbeauftragte etc.) von Belang. Liegen diese Daten vor, kann evaluiert werden, ob die Einführung einer Leistungsevaluation im Bereich der Forschung sich zu Lasten anderer Tätigkeiten von Hochschullehrern/innen auswirkt (siehe dazu auch Abschnitt 7.4).

## 7.4 Einführung eines regelmäßigen Audits des Systems zur Forschungsevaluation

Eine abschließende Empfehlung betrifft das in diesem Bericht vorgeschlagene System zur Messung der Forschungsleistung als Ganzes. Wie insbesondere die ökonomische Forschung in vielen Fällen zeigen konnte, weicht die Reaktion von Menschen auf die Einführung von Anreizsystemen zum Teil von den durch das Anreizsystem intendierten Wirkungen ab. Dies ist auch im Fall der Einführung eines Systems zur Messung der Forschungsleistung zumindest denkbar. Das Ziel, die Forschungsleistung zu verbessern, indem gezielt Anreize zur Verbesserung gesetzt werden, muss nicht zwangsläufig durch das System erreicht werden. Es ist zumindest denkbar, dass die Forscher/innen zukünftig nur noch die Leistungen erbringen, die auch im System erfasst werden, da nur diese als Leistungen „zählen“ (vgl. Frey 2007, S. 209).

---

<sup>33</sup> Dieses Vorgehen entspricht auch der Logik, nach welcher die DFG ihre Bewilligungsdurchschnitte erfasst.

Darüber hinaus kann ebenfalls nicht ausgeschlossen werden, dass die Einführung eines Systems zur Messung der Forschungsleistung an der HHU nicht den gewünschten Effekt hat, die Qualität der Forschungsleistung zu erhöhen. So argumentiert Frey, dass durch eine regelmäßige Evaluation von Forschungsleistungen die Qualität und Originalität der Forschung sinke, da innovative wissenschaftliche Arbeit intrinsischer Motivation bedürfe, welche durch die Evaluation verdrängt werde (vgl. Frey 2007, S. 211).

Zuletzt muss berücksichtigt werden, dass es auch zu einer Anreizverschiebung kommen könnte, wenn aufgrund von Evaluationen Forschungsleistungen als wichtiger erachtet werden als Leistungen in der Lehre und Selbstverwaltung (vgl. Dilger 2001, S. 138; Pull 2009, S. 311). Dies stellt noch einmal die Bedeutung eines Systems zur Messung der Lehrleistung sowie die explizite Berücksichtigung der akademischen Selbstverwaltung als eigenständige Leistung heraus.

Vor dem Hintergrund dieser möglichen nicht-intendierten Effekte der Einführung eines Systems zur Messung der Forschungsleistung wird eine kritische Evaluation der Effekte des Systems nach dessen Einführung dringend empfohlen. Diese Evaluation könnte beispielsweise auf Grundlage von Befragungen im Kreis der Professoren/innen durchgeführt werden. Darüber hinaus empfiehlt sich eine Befragung der Dekane/innen und Prodekanen/innen, da diese einen Überblick über ihre Fakultäten haben und mögliche Verhaltensänderungen unter den Kollegen und Kolleginnen beobachten könnten.



---

## LITERATUR

- BiHui J, LiMing L, Rousseau R, Egghe L (2007) The R- and AR-indicies: Complementing the h-index. *Chinese Science Bulletin* 52 (6): 855-863.
- Bommer R, Ursprung HW (1998) Spieglein, Spieglein an der Wand: Eine publikationsanalytische Erfassung von Forschungsleistungen volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz. *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 118 (1): 1-28.
- Bogumil J, Heinze RG (2009) Einleitung. In: Bogumil J, Heinze RG (Hrsg.) *Neue Steuerung von Hochschulen – Eine Zwischenbilanz*. Edition Sigma, Berlin: 7-12.
- Borgwardt A (2011) Kriterien, Aussagekraft und Wirkungen von Rankings. In: Borgwardt A (Hrsg.) *Rankings im Wissenschaftssystem – Zwischen Wunsch und Wirklichkeit*. Publikationen zur Konferenz der Friedrich-Ebert-Stiftung am 17. März 2011. Schriftenreihe des Netzwerk Exzellenz an Deutschen Hochschulen, Friedrich-Ebert-Stiftung, Berlin: 23-42.
- Bräuninger M, Haucap J, Muck J (2011) Was lesen und schätzen Ökonomen im Jahr 2011? *DICE Ordnungspolitische Perspektiven* Nr. 18: 1-56.
- Broad WJ (1981) The publishing game: Getting more for less. *Science* 211 (4487): 1137-1139.
- Bull HP, Mehde V (2000) Reform der Hochschulorganisation – die populären Modelle und ihre Probleme. *JZ* 2000: 650-659.
- Daniel HD, Fisch R (1990) Research performance evaluation in the German university sector. *Scientometrics* 19 (5-6): 349-361.
- Denninger E, (2001) in: Stein E, Denninger E, Hoffmann-Riem W, Schneider H-P (2001) *Kommentar zum Grundgesetz für die Bundesrepublik Deutschland (AK-GG)*. 3. Aufl. Neuwied 2001, Loseblatt, Art. 5 Abs. 3 I (2001).
- Deutsche Forschungsgemeinschaft (2004) *Empfehlungen zu einer „Leistungsorientierten Mittelvergabe“ (LOM) an den Medizinischen Fakultäten*. Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft. Deutsche Forschungsgemeinschaft, Bonn.
- Deutsche Forschungsgemeinschaft (2009) *Förder-Ranking 2009*. Institutionen – Regionen – Netzwerke. Fachliche Profile von Hochschulen und außeruniversitären Forschungseinrichtungen im Licht öffentlich geförderter Forschung. Wiley-VCH, Weinheim.
- Deutscher Juristen-Fakultätentag (2010) *Mitglieder des Deutschen Juristen-Fakultätentages*. Zugriff am 11.05.2011 unter <http://www.djft.de/>
- DIE ZEIT (2011) *CHE-Hochschulranking 2011/12*. Zugriff am 25.05.2011 unter <http://ranking.zeit.de/che2011/de/>
- Dilger A (2000) Plädoyer für einen Sozialwissenschaftlichen Zitationsindex. *Die Betriebswirtschaft* 60 (4): 473-484.
- Dilger A (2001) Was lehrt die Prinzipal-Agenten-Theorie für die Anreizgestaltung in Hochschulen? *Zeitschrift für Personalforschung* 15 (2): 132-148.
- Dilger A (2010) Rankings von Zeitschriften und Personen in der BWL. *Zeitschrift für Management* 5 (1): 91-102.
- Dilger A (2011) Befragung der Kommission Hochschulmanagement zu VHB-JOURQUAL. *Diskussionspapier des Instituts für Organisationsökonomik* 7/2011: 1-22.
- Dyckhoff H, Rassenhövel S, Sandfort K (2009) Empirische Produktionsfunktion betriebswirtschaftlicher Forschung: Eine Analyse der Daten des Centrums für Hochschulentwicklung. *Zeitschrift für betriebswirtschaftliche Forschung* 61 (2): 22-56.
- Dyckhoff H, Schmitz C (2007) Forschungsleistungsmessung mittels SSCI oder SCI-X? Internationale Sichtbarkeit und Wahrnehmung der Betriebswirtschaftslehre von 1990-2004. *Die Betriebswirtschaft* 67 (6): 638-662.

- Dyckhoff H, Thieme A, Schmitz C (2005) Die Wahrnehmung deutschsprachiger Hochschullehrer für Betriebswirtschaft in der internationalen Forschung. Eine Pilotstudie zu Zitationsverhalten und möglichen Einflussfaktoren. *Die Betriebswirtschaft* 65 (4): 350-372.
- Egghe L (2006) Theory and practice of the g-index. *Scientometrics* 69 (1): 131-152.
- Ellison G (2010) Is peer review in decline? *Economic Inquiry* 49 (3): 635-657.
- Enserink M (2007) Who ranks the university rankers? *Science* 317 (5841): 1026-1028.
- Erichsen H-U, Scherzberg A (1990) Verfassungsrechtliche Determinanten staatlicher Hochschulpolitik. *NVwZ* 1990: 8-17.
- Fiedler M, Welpel IM, Picot A (2006) Terra Incognita. Forschungsleistungen und Qualifizierungswege des deutschsprachigen Hochschullehrernachwuchses für Betriebswirtschaftslehre. *Die Betriebswirtschaft* 66 (4): 464-486.
- Fiedler M, Welpel IM, Lindlbauer K, Sattler K (2008) Denn wer da hat, dem wird gegeben: Publikationsproduktivität des BWL-Hochschullehrernachwuchses und deren wissenschaftlicher Betreuer. *Zeitschrift für Betriebswirtschaft* 78 (5): 477-508.
- Franck G (1998) *Ökonomie der Aufmerksamkeit: Ein Entwurf*. München, Carl Hanser.
- Frey B (2007) Evaluierungen, Evaluierungen ... *Evaluativ*. Perspektiven der Wirtschaftspolitik 8 (3): 207-220.
- Frey B, Rost K (2010) Do rankings reflect research quality? *Journal of Applied Economics* 13 (1): 1-38.
- Gärditz KF (2009) Evaluationsbasierte Forschungsfinanzierung im Wissenschaftsrecht. *WissR* 42 (2009): 353-392.
- Gärditz KF (2005) Hochschulmanagement und Wissenschaftsadäquanz. *NVwZ* 2005: 407-410.
- Geis M-E (2004) Das Selbstbestimmungsrecht der Universitäten. *WissR* 37 (2004): 2-25.
- Gläser J. (2006) Die Fallstricke der Bibliometrie. *Soziologie* 35(1): 42-51.
- Gläser J, Lange S, Laudel G (2009) Auswirkungen der evaluationsbasierten Forschungsfinanzierung an Universitäten auf die Inhalte der Forschung. Australien und Deutschland im Vergleich. *WissR* 42 (2009): 329-352.
- Görlich C (2003) Wissenschaftsfreiheit und Hochschulmanagement. *DÖV* 2003: 583-588.
- Groß T (2002) Wissenschaftsadäquates Wissenschaftsrecht. *WissR* 35 (2002): 313-332.
- Haucap J, Hartwich T, Uhde A (2005) Besonderheiten und Wettbewerbsprobleme des Marktes für wissenschaftliche Fachzeitschriften. *Vierteljahreshefte zur Wirtschaftsforschung* 74 (3): 84-107.
- Heiber H (1986) Messung universitärer Forschungsleistungen mit Hilfe der Zitatennalyse. In: Fisch R, Daniel HD (Hrsg) *Messung und Förderung von Forschungsleistungen*. Universitätsverlag Konstanz, Konstanz: 135-149.
- Heinze T (2002) Evaluation von Forschungsleistungen. Konzeptionelle Überlegungen und Situationsbeschreibung für Deutschland. *Wissenschaftsmanagement* 8 (6): 14-22.
- Hennig-Thurau T, Walsh G, Schrader U (2004) VHB-JOURQUAL: Ein Ranking von betriebswirtschaftlich-relevanten Zeitschriften auf der Grundlage von Expertenurteilen. *Zeitschrift für betriebswirtschaftliche Forschung* 56 (9): 520-545.
- Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102 (46): 16569-16572.
- Hoffacker W (2003) Zur Rolle von Vertrauen und Recht, Markt und Kontrakt als Steuerungselemente des Hochschulsystems. *WissR* 36 (2003): 92-104.
- Hornbostel, S. (1997) *Wissenschaftsindikatoren: Bewertungen in der Wissenschaft*. Opladen, Westdeutscher Verlag.
- Hornbostel S (2006) Leistungsmessung in der Forschung. Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung. In: Hochschulrektorenkonferenz (Hrsg.) *Beiträge zur Hochschulpolitik 1/2006, Hochschulrektorenkonferenz, Bonn*: 219-228.
- Hornbostel S, Heise S (2006) Die Rolle von Drittmitteln in der Steuerung von Hochschulen. In: Berthold C (Hrsg.) *Handbuch Wissenschaftsfinanzierung*. Raabe-Verlag, Berlin: 1-34.



- Hudson J (2007) Be known by the company you keep: Citations – quality or chance? *Scientometrics* 71 (2): 231-238.
- Hufeld U (2002) Staatlicher Schutz der Universitas litterarum. *DÖV* 2002: 309-318.
- Huff AS (1999) *Writing for scholarly publication*, Sage, Thousand Oaks.
- Jaeger M (2006) Leistungsorientierte Budgetierung: Analyse der Umsetzung an ausgewählten Universitäten und Fakultäten/Fachbereichen. *HIS Kurz-Information A 1/2006*, Hannover.
- Jaeger M (2008) Wie wirksam sind leistungsorientierte Budgetierungsverfahren an deutschen Hochschulen? *Zeitschrift für Hochschulentwicklung* 3 (1): 89-104.
- Jaeger M (2009) Steuerung durch Anreizsysteme an Hochschulen. In: Bogumil J, Heinze R (Hrsg.) *Neue Steuerung von Hochschulen*. Edition Sigma, Berlin: 45-65.
- Jaeger M, In der Smitten S (2009) Evaluation der leistungsbezogenen Mittelvergabe an die Berliner Hochschulen. *HIS: Forum Hochschule 1/2009*, Hannover.
- Jaeger M, In der Smitten S (2010) Evaluation der leistungsbezogenen Mittelvergabe an die Hochschulen in Mecklenburg-Vorpommern. *HIS: Forum Hochschule 4/2010*, Hannover.
- Jaeger M, Leszczensky M, Orr D, Schwarzenberger A (2005) Formelgebundene Mittelvergabe und Zielvereinbarungen als Instrumente der Budgetierung an deutschen Universitäten: Ergebnisse einer bundesweiten Befragung. *HIS Kurz-Information A 13/2005*, Hannover.
- Jansen D, Wald A, Franke K, Schmoch U, Schubert T (2007) Drittmittel als Performanzindikator der wissenschaftlichen Forschung – Zum Einfluss von Rahmenbedingungen auf Forschungsleistungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59 (1): 125-149.
- Kieser A (1998) Going Dutch – Was lehren niederländische Erfahrungen mit der Evaluation universitärer Forschung? *Die Betriebswirtschaft* 58 (2): 208-224.
- Kleiner M (2010) „Qualität statt Quantität“ – Neue Regeln für Publikationsangaben in Förderanträgen und Abschlussberichten. Zugriff am 02.05.2012 unter [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/2010/statement\\_qualitaetstatt\\_quantitaet\\_mk\\_100223.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2010/statement_qualitaetstatt_quantitaet_mk_100223.pdf)
- Kousha K, Thelwall M (2009) Google Book Search: Citation Analysis for Social Science and the Humanities. *Journal of the American Society for Information Science and Technology* 60(8): 1537-1549.
- Kreysing M (2008) Forschungsförderung mittels leistungsorientierter Mittelvergabe. *Zeitschrift für Hochschulentwicklung* 3 (1): 19-28.
- Kuhlmann S, Heinze T (2004) Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf. Teil I: Konzeptionelle Grundlagen. *WissR* 37 (2004): 53-69.
- Kuhlmann S, Heinze T (2004) Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf. Teil II: Produktion und Verwendung evaluativer Information sowie Möglichkeiten ihrer künftigen Organisation. *WissR* 37 (2004): 125-149.
- Kulkarni AV, Aziz B, Shams I, Busse JW (2009) Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *The Journal of the American Medical Association* 302 (10): 1092-1096.
- Ladeur K-H (2005) Die Wissenschaftsfreiheit der „entfesselten Hochschule“. *DÖV* 2005: 753-764.
- Lindholm-Romantschuk Y, Warner J (1996) The Role of Monographs in Scholarly Communication: An Empirical Study of Philosophy, Sociology and Economics. *Journal of Documentation* 52(4): 389-404.
- Linmans AJM (2010) Why with bibliometrics the Humanities does not need to be the weakest link Indicators for research evaluation based on citations, library holdings, and productivity measures. *Scientometrics* 83(2): 337-354.
- Litzenberger T, Sternberg R (2005) Die Forschungsleistung der Soziologie an zehn deutschen Universitäten. Ein bibliometrischer Vergleich auf Basis des Social Science Citation Index (SSCI). *Soziologie* 34 (2): 174-190.
- Löwer W (2000) Zwei aktuelle Fragen der Hochschulreformdebatte. *WissR* 33 (2000): 302-331.
- Mehde V (2011) Die Evaluation von Verwaltungsleistungen. *Verw* 44 (2011): 179-211.

- Meho LI, Yang K (2007) Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology* 58 (13): 2105-2125.
- Meier F (2009) Die Universität als Akteur. Zum institutionellen Wandel der Hochschulorganisation, Wiesbaden.
- Meier F, Schimank U (2009) Matthäus schlägt Humboldt? New Public Management und die Einheit von Forschung und Lehre. *Beiträge zur Hochschulforschung* 31(1): 42-61.
- Merton RK (1968) The Matthew Effect in science. *Science* 159 (3810): 56-63.
- Minssen H, Wilkesmann U (2003) Folgen der indikatorisierten Mittelzuweisung in nordrhein-westfälischen Hochschulen. *Beiträge zur Hochschulforschung* 25 (3): 106-128.
- Moed, H. (2009) Measuring contextual citation impact of scientific journals.
- Müller H (2010): Wie valide ist das Handelsblatt-BWL-Ranking? Zeitschriften- und zitationsbasierte Personenrankings im Vergleich. *Betriebswirtschaftliche Forschung und Praxis* 62 (2): 150-164.
- Müller-Camen M, Salzgeber S (2005) Changes in academic work and the chair regime: The case of German business administration academics. *Organization Studies* 26 (2): 271-290.
- Müller-Terpitz R (2011) Neue Leitungsstrukturen als Gefährdungen der Wissenschaftsfreiheit. *WissR* 44 (2011): 236-263.
- Nederhof AJ, van Leeuwen TN, van Raan AF (2010) Highly cited non-journal publications in political science, economics and psychology: a first exploration. *Scientometrics* 83(2): 363-374.
- Oppenländer KH (2002) Einige Bemerkungen zur Evaluierung von Wirtschaftsforschungsinstituten. *Jahrbücher für Nationalökonomie und Statistik* 222 (3): 382-392.
- Oswald A (2007) An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers, *Economica* 74 (293): 21-31.
- Pernice I (2004), in: H. Dreier (Hrsg.), *Grundgesetz Kommentar*, 2. Aufl. Tübingen 2004, Art. 5 III (Wissenschaft).
- Pommerehne WW, Renggli MFP (1986) Die Messung universitärer Forschungsleistungen am Beispiel der Wirtschaftswissenschaften. In: Fisch R, Daniel HD (Hrsg.) *Messung und Förderung von Forschungsleistungen*. Universitätsverlag Konstanz, Konstanz: 89-134.
- Pull K (2009) Editorial: Risiken und Nebenwirkungen der Leistungsmessung von Professoren: Warum wir sie brauchen und warum wir dennoch aufpassen müssen. *Die Betriebswirtschaft* 69 (3): 311-313.
- Ridder H (2006) Universitäten zwischen Eigensinn und strategischer Orientierung. In: Welte H, Auer M, Meister-Scheytt C (Hrsg.) *Management von Universitäten*. Hampp, München: 101-114.
- Rassenhövel S (2010) Performancemessung im Hochschulbereich. *Theoretische Grundlagen und empirische Befunde*. Gabler, Wiesbaden.
- Rassenhövel S, Dyckhoff H (2006) Die Relevanz von Drittmittelindikatoren bei der Beurteilung der Forschungsleistung im Hochschulbereich. In: Zelewski S, Akca N (Hrsg.) *Fortschritt in den Wirtschaftswissenschaften – Wissenschaftstheoretische Grundlagen und exemplarische Anwendungen*. Gabler, Wiesbaden: 85-112.
- Riordan P, Ganser C, Wolbring T (2011) **Measuring the quality of research-A critical analysis of the Forschungsgrating of the German Wissenschaftsrat**. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63 (1): 147-172.
- Rost K, Frey BS (2011) Quantitative and qualitative rankings of scholars. *Schmalenbach Business Review* 63 (1): 63-91.
- Sandberger G (2011) Die Neuordnung der Leitungsorganisation der Hochschulen durch die Hochschulrechtsnovellen der Länder. *WissR* 44 (2011): 118-155.
- Sandberger G (2002) Organisationsreform und -autonomie – Bewertung der Reformen in den Ländern. *WissR* 35 (2002): 125-150.
- Schenke W-R (2005) Neue Fragen an die Wissenschaftsfreiheit. *Neue Hochschulgesetze im Lichte des Art. 5 III GG*. *NVwZ* 2005: 1000-1009.



- Schläpfer F (2010) How much does journal reputation tell us about the academic interest and relevance of economic research? Empirical analysis and implications for environmental economic research. *GAIA – Ecological Perspectives for Science and Society* 19 (2): 140-145.
- Schlüter A (2011) Neue Leitungsstrukturen an den Hochschulen – Eine Zwischenbilanz. *WissR* 44 (2011): 264-279.
- Schmoch U (2009) Geeignete Ansätze zur Messung wissenschaftlicher Leistung. *Beiträge zur Hochschulforschung* 31(1): 26-41.
- Schrader U, Hennig-Thurau T (2009) VHB-JOURQUAL 2: Method, results, and implications of the German Academic Association for Business Research's journal ranking. *Business Research* 2 (2): 180-204.
- Schulze-Fielitz H, in: Benda E, Maihofer W, Vogel H-J, *Handbuch des Verfassungsrechts* (2), 2. Aufl. Berlin 1995, § 27 Freiheit der Wissenschaft.
- Seidler H H (1999) Die Ersetzung des Rechts durch die Ökonomie. *WissR* 32 (1999): 261-270.
- Sieweke S (2010) Managementstrukturen und outputorientierte Finanzierung im Hochschulbereich. Zum Instrumentarium des Neuen Steuerungsmodells im Hinblick auf Wissenschaftsfreiheit und Hochschulautonomie. Nomos, Baden-Baden.
- Starbuck WH (2005) How much better are the most-prestigious journals? The statistics of academic publishing. *Organization Science* 16 (2): 180-200.
- Stock WG (2000) Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In: Fuchs-Kittowski H, Parthey H, Umstätter W (Hrsg.) *Wissenschaftsforschung Jahrbuch 1998*. Gesellschaft für Wissenschaftsforschung, Berlin: 239-282.
- Süß S (2006) Wandel der Forschung und dysfunktionale Effekte des zunehmenden Wettbewerbs um wissenschaftliche Reputation. *Die Hochschule – Journal für Wissenschaft und Bildung* 15 (1): 84-97.
- Torres-Salinas D, Moed H (2009) Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of book titles in Economics. *Journal of Informetrics* 3(2009): 9-26.
- University of Texas (2011) The UTD Top 100 Business School Research Rankings. Zugriff am 15.11.2011 unter <http://som.utdallas.edu/top100Ranking/>
- Ursprung HW (2003) Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators. *Perspektiven der Wirtschaftspolitik* 4 (3): 177-189.
- Valadkhani A, Worthington A (2006) Ranking and clustering Australian university research performance, 1998-2002. *Journal of Higher Education Policy and Management* 28 (2): 189-210.
- Voeth M, Herbst U, Loos J (2011) Bibliometrische Analyse der Zeitschriftenrankings VHB-JOURQUAL 2.1 und Handelsblatt-Zeitschriftenranking BWL. Untersucht am Beispiel der Forschungsdisziplin Marketing. *Die Betriebswirtschaft* 71 (5): 439-458.
- Wall HJ (2009) Don't get skewed over by journal rankings. *B.E. Journal of Economic Analysis and Policy* 9 (1): Artikel 34
- Waltman L, Yan E, van Eck NJ (2011) A recursive field-normalized bibliometric performance indicator: an application to the field of library and information science. *Scientometrics* 89(2): 301–314.
- Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AFJ (2011) Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics* 5(1): 37-47.
- Wirtschafts- und Sozialwissenschaftlicher Fakultätentag (2010) Mitgliedsfakultäten. Zugriff am 10.05.2011 unter [http://www.fakultaetentag.de/test\\_WISOFT/mitglieder.html](http://www.fakultaetentag.de/test_WISOFT/mitglieder.html)
- Wissenschaftsrat (2005a) Empfehlungen zu Rankings im Wissenschaftssystem. Wissenschaftsrat, Köln.
- Wissenschaftsrat (2005b) Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland. Wissenschaftsrat, Köln.
- Wissenschaftsrat (2007) Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Chemie. Wissenschaftsrat, Köln
- Wissenschaftsrat (2008) Empfehlungen zum Forschungsrating. Rostock, Wissenschaftsrat.

Wissenschaftsrat (2008) Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Soziologie. Wissenschaftsrat, Köln.

Wissenschaftsrat (2010) Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften. Wissenschaftsrat, Köln.

Wolf J, Rohn A, Macharzina K (2005) Institution und Forschungsproduktivität. Befunde und Interpretationen aus der deutschsprachigen Betriebswirtschaftslehre. *Die Betriebswirtschaft* 65 (1): 62-77.

Zuccala A, van Leeuwen TN (2011) Book Reviews in Humanities Research Evaluations. *Journal of the American Society for Information Science and Technology* 62(10): 1979-1991.



ANHANG 1:  
DRITTMITTEL JE PROFESSOR/IN NACH FÄCHERGRUPPEN UND FACHBEREICHEN

| Fächergruppe<br>Lehr- und Forschungsbereich <sup>1)</sup>      | Drittmittel je Professor/in<br>(in 1.000 Euro) |        |
|--|--|--------|
|  | 2008   | 2009   |
| <b>Sprach- und Kulturwissenschaften</b>                        | 57,81  | 66,28  |
| darunter:  |  |        |
| Sprach- und Kulturwissenschaften allgemein                     | 135,91   | 190,34 |
| Evangelische Theologie   | 32,79  | 37,24  |
| Katholische Theologie  | 17,02  | 21,75  |
| Philosophie  | 50,22  | 47,92  |
| Geschichte   | 85,14  | 96,70  |
| Allgemeine und vergleichende Literatur- und Sprachwissenschaft | 112,65   | 82,57  |
| Altphilologie (Klassische Philologie)                          | 45,19  | 46,12  |
| Germanistik (Deutsch, germanische Sprachen ohne Anglistik)     | 40,12  | 44,99  |
| Anglistik, Amerikanistik                                       | 23,66  | 22,58  |
| Romanistik   | 21,51  | 29,72  |
| Sonstige/Außereuropäische Sprach- und Kulturwissenschaften     | 83,85  | 96,45  |
| Psychologie  | 99,55  | 109,81 |
| Erziehungswissenschaften                                       | 52,55  | 69,90  |
| Sonderpädagogik  | 25,78  | 26,73  |
| <b>Sport</b>   | 125,02   | 124,83 |
| <b>Rechts-, Wirtschafts- und Sozialwissenschaften</b>          | 62,71  | 70,20  |
| darunter:  |  |        |
| Rechts-, Wirtschafts- und Sozialwissenschaften allgemein       | 66,73  | 26,56  |
| Politikwissenschaften  | 68,98  | 67,09  |
| Sozialwissenschaften   | 67,79  | 72,54  |
| Rechtswissenschaften   | 34,05  | 35,97  |
| Wirtschaftswissenschaften                                      | 70,67  | 84,92  |

Anhang 1: Fortsetzung

| Fächergruppe<br>Lehr- und Forschungsbereich <sup>1)</sup>                       | Drittmittel je Professor/in<br>(in 1.000 Euro) |               |
|---|--|---------------|
|   | 2008   | 2009          |
| <b>Mathematik, Naturwissenschaften</b>  | 219,25   | 239,66        |
| darunter:   |  |               |
| Mathematik  | 66,50  | 80,67         |
| Informatik  | 233,22   | 253,26        |
| Physik, Astronomie  | 309,15   | 311,88        |
| Chemie  | 250,03   | 283,53        |
| Pharmazie   | 148,02   | 140,81        |
| Biologie  | 269,80   | 294,42        |
| Geowissenschaften (ohne Geografie)  | 267,19   | 364,88        |
| Geografie   | 86,99  | 94,59         |
| <b>Humanmedizin<br/>(einschl. zentrale Einrichtungen der Hochschulkliniken)</b> | 442,16   | 474,55        |
| <b>Veterinärmedizin</b>   | 102,91   | 112,25        |
| <b>Agrar-, Forst- und Ernährungswissenschaften</b>                              | 220,20   | 250,82        |
| <b>Ingenieurwissenschaften</b>  | 402,66   | 460,13        |
| darunter:   |  |               |
| Maschinenbau/Verfahrenstechnik  | 663,27   | 761,53        |
| Elektrotechnik  | 347,79   | 388,97        |
| Architektur   | 52,54  | 62,32         |
| Bauingenieurwesen   | 290,14   | 301,44        |
| <b>Kunst, Kunstwissenschaft</b>   | 34,16  | 39,74         |
| darunter:   |  |               |
| Musik, Musikwissenschaft  | 19,16  | 23,35         |
| <b>Fächergruppen insgesamt</b>  | <b>197,79</b>                                  | <b>217,59</b> |
| <b>Zentrale Einrichtungen</b>   | <b>25,84</b>                                   | <b>26,34</b>  |
| <b>Insgesamt</b>  | <b>220,72</b>                                  | <b>240,94</b> |

<sup>1)</sup> Lehr- und Forschungsbereiche mit mindestens 100 Professoren/innenstellen. Quelle: Statistisches Bundesamt: Finanzstatistische Kennzahlen für den Hochschulbereich auf der Basis der Hochschulfinanzstatistik. Laufende Grundmittel und Drittmittel je Professor/in der Universitäten nach Fächergruppen und Lehr- und Forschungsbereichen (ohne drittmittelfinanzierte Professuren)



## ANHANG 2: ERLÄUTERUNG DES „SOURCE NORMALIZED IMPACT PER PAPER“ (SNIP)

Die Konstruktionslogik des SNIP geht davon aus, dass sich das Zitationspotential von Publikationen nach Fächern unterscheidet. Mit Zitationspotential ist die Wahrscheinlichkeit gemeint, dass ein Beitrag von einem anderen Artikel zitiert wird. Indem das Fachgebiet empirisch aus den Referenzen der Artikel einer Publikationsquelle bestimmt wird, können auch Unterschiede zwischen den Teilgebieten eines Faches abgebildet werden.

Die Berechnung des SNIP erfolgt in drei Schritten. Erstens wird der Brutto-Impact-Faktor („Raw Impact per Paper“ – RIP) bestimmt. Dieser entspricht dem Impact-Faktor des SSCI, d.h. der durchschnittlichen Anzahl von Zitaten im Referenzjahr pro publiziertem Artikel in einer Zeitschrift in den zwei bzw. fünf vorangegangenen Jahren. Der Referenzzeitraum, der für die Berechnung des SNIP zugrunde gelegt wird, beträgt jedoch drei Jahre. In den meisten Disziplinen entfallen die meisten Zitate auf diese Zeitspanne (Lancho-Barrantes, Guerrero-Bote, & Moya-Anegon, 2010, p. 379).

Zweitens wird das Zitationspotential einer Zeitschrift bestimmt, d.h. die Wahrscheinlichkeit, dass ein Artikel aus dieser Zeitschrift zitiert wird. Dies geschieht, indem aus dem Fachgebiet einer Zeitschrift die durchschnittliche Anzahl der Referenzen bestimmt wird. Die Abgrenzung des Fachgebiets einer Zeitschrift erfolgt dabei nicht, indem eine a priori Klassifikation von Fächern vorgenommen wird, sondern es werden die Referenzen aus den Zeitschriften gezählt, welche die untersuchte Zeitschrift zitieren. Die Definition des Fachgebiets ergibt sich also aus den Zitationsbeziehungen zwischen den Zeitschriften.<sup>34</sup> Da die Bestimmung des Zitationspotentials umso reliabler gelingt, je mehr Referenzen auf die Zeitschrift in der Datenbank erfasst werden (das Zitationspotential hängt folglich von der Abdeckung eines Fachgebiets in der Datenbank ab), wird das empirisch ermittelte Zitationspotential anhand des Abdeckungsgrades der Datenbank gewichtet. Auch der Abdeckungsgrad der Datenbank wird empirisch ermittelt, indem der Anteil der Referenzen aus den letzten drei Jahren ermittelt wird, der in der Datenbank erfasst ist, auch „Database Citation Potential“ oder DCP genannt.<sup>35</sup> Da auch das DCP der Zeitschriften von ihrem Fachgebiet abhängt, werden die Werte normalisiert, indem der DCP einer Zeitschrift durch den Median DCP geteilt wird. Dies ergibt das „Relative Database Citation Potential“ (RDCP). Die Zeitschrift mit einem Median DCP hat folglich einen RDCP von 1.

Im dritten Schritt wird nun der „Raw Impact per Paper“ (RIP) durch das „Relative Database Citation Potential“ (RDCP) geteilt:

$$SNIP = \frac{RIP}{RDCP}$$

Auf diese Weise werden Unterschiede in der Publikationskultur der Fächer im Impactfaktor korrigiert. Im Median jedes Fachgebiets (wenn RDCP = 1) entspricht daher der SNIP dem RIP.

Die Korrektur des Brutto-Impact-Faktors um das „Relative Database Citation Potential“ führt indes nicht mechanisch dazu, z.B. deutschsprachige Titel gegenüber englischsprachigen Zeitschriften ‚aufzuwerten‘, wie in Tabelle 11 (Blöcke 1 und 3) am Beispiel von soziologischen und politikwissenschaftlichen Zeitschriften gezeigt wird. Allerdings zeigt sich die Wirksamkeit des SNIPs im Vergleich über Fächergrenzen hinweg (zusätzlich Block 3 in Tabelle 11). Der große Unterschied zwischen „Behavioral and Brain Sciences“ – einer Zeitschrift aus dem Bereich Life-Sciences – und den geistes- und sozialwissenschaftlichen Zeitschriften wird weitgehend nivelliert, wenn das Zitationspotential korrigiert wird.

<sup>34</sup> Wobei die Zitate, die auf die Zielzeitschrift entfallen, in der Regel nur einen sehr kleinen Anteil an der Gesamtzahl von Referenzen haben.

<sup>35</sup> Der DCP wird also nur auf Grund von Referenzen ermittelt, die in den von Scopus erfassten Quellen aus den letzten drei Jahren verfügbar sind. Dadurch findet je nach Fachgebiet eine erhebliche Einschränkung des DCP statt.

| Zeitschrift   | Scopus            | Web of Science           |                          |
|---|-------------------|--------------------------|--------------------------|
|   | SNIP <sup>a</sup> | JIF-2 Jahre <sup>b</sup> | JIF-5 Jahre <sup>b</sup> |
| American Journal of Sociology                           | 3,884             | 2,808                    | 5,046                    |
| European Sociological Review                            | 2,190             | 0,816                    | 1,345                    |
| Kölner Zeitschrift für Soziologie und Sozialpsychologie | 0,676             | 1,188                    | 0,867                    |
| American Journal of Political Science                   | 3,629             | 2,397                    | 3,363                    |
| European Journal of Political Research                  | 1,205             | 2,514                    | 2,734                    |
| Politische Vierteljahresschrift                         | 0,332             | 0,625                    | 0,429                    |
| Bioethics   | 0,774             | 1,013                    | 1,731                    |
| Behavioral and Brain Sciences                           | 0,700             | 12,818                   | 19,355                   |

Anmerkungen: <sup>a</sup> SNIP=Source Normalized Impact per Paper. Referenzzeitraum für den SNIP sind drei Jahre; JIF=Journal-Impact-Faktor.

Tabelle 11: Vergleich der Impact-Faktoren von Scopus und dem Web of Science (Referenzjahr 2008)

Ein praktischer Vorteil liegt in der Verfügbarkeit des Indikators in der Scopus Datenbank, weshalb der SNIP ohne größeren Aufwand zur Gewichtung von Publikationen in einem Publikationsscore verwendet werden kann.<sup>36</sup> Die Bestimmung von Zeitschriftenlisten für Fächer und Teilgebiete der Fächer sowie die Berechnung eines mittleren Impact-Faktors für diese Zeitschriftenlisten entfällt.

<sup>36</sup> Der Vollständigkeit halber sei erwähnt, dass Scopus einen zweiten Indikator SJP bereitstellt, der das Prestige eines Journals über die Impact-Faktoren der Journals bestimmt, aus denen Zitate auf die Zeitschrift erfolgen (dem Ranking der Zeitschriften liegt dann also ein rekursiver Mechanismus zugrunde, wie der im PageRank-Algorithmus von Google umgesetzt ist). Dieser Indikator ist für den Fächervergleich ungeeignet, weil die Spreizung des Indikators aufgrund der fachspezifischen Publikationskulturen tendenziell verstärkt und nicht verringert wird. Die Daten, die der Berechnung von SNIP und SJR zu Grunde gelegt werden, können von der Homepage [www.journalmetrics.com](http://www.journalmetrics.com) als MS-EXCEL-Tabelle heruntergeladen werden.



## ANHANG 3: BEFRAGUNG DER PROFESSOREN/INNEN DER PHILOSOPHISCHEN FAKULTÄT

### Leistungsindikatoren für die Philosophische Fakultät Ergebnisse einer Befragung der Professoren/innen an der Philosophischen Fakultät

Pascal Siegers  
27. Oktober 2011

#### 1. Einleitung

Die Entwicklung von Leistungsindikatoren für die Philosophische Fakultät der Heinrich-Heine-Universität Düsseldorf entspringt aus der Diskussion um die Vergabe besonderer Leistungszulagen im Rahmen der 2005 eingeführten W-Besoldung. In vom Rektorat vorgelegten Verfahrensvorschlag für die Vergabe von Leistungszuschlägen werden die Fakultäten vor die Wahl aus zwei Verfahrensvarianten gestellt. Zum einen können besondere Leistungszulagen vergeben werden, wenn Professoren/innen Großprojekte aus Förderprogrammen der DFG (Sonderforschungsbereiche, Forschergruppen, Graduiertenkollegs) oder äquivalente Projekte bei anderen Drittmittelgebern einwerben. Zum anderen können die Fakultäten selbst Leistungsindikatoren entwickeln, anhand derer die Forschungsleistung der Professoren/innen vergleichend beurteilt wird. Im Fall der Philosophischen Fakultät können in dieser Variante drei Professoren/innen im Jahr besondere Leistungszulagen gewährt werden. Deren Höhe und die Bezugsdauer werden im Verfahrensvorschlag des Rektorates nicht konkretisiert.

Wenn die Philosophische Fakultät ein tragfähiges Systems zur Leistungsevaluation entwickeln will, müssen die Leistungskriterien mindestens den folgenden drei Ansprüchen genügen.

- 1) Leistungsindikatoren müssen gerecht sein, d.h. sie dürfen nicht zu einer systematischen Benachteiligung einzelner Professoren/innen, Institute oder Fakultäten bei der Vergabe von Leistungszulagen führen. Alle Professoren/innen müssen realistische Chancen haben, mit besonderen Leistungen eine Leistungszulage zu erreichen. In der Praxis bedeutet dies, die Leistungsindikatoren in ihrer fachspezifischen Einbettung zu interpretieren, d.h. indem Leistungen mit Referenz auf fächerspezifische Leistungsdurchschnitte bewertet werden und nicht auf das Wissenschaftssystem als Ganzes bezogen werden. Dies gilt umso mehr, wenn finanzielle Indikatoren für die Leistungsbewertung herangezogen werden. Nur durch die Implementation gerechter Leistungsindikatoren kann die Akzeptanz der Leistungsbewertung sichergestellt werden, was für den langfristigen Erfolg der Leistungsbewertung eine *conditio sine qua non* darstellt.
- 2) Die Indikatoren müssen einfach gestaltet sein. Der Aufwand für die Datenerhebung muss übersichtlich bleiben. Die Leistungsbewertung darf nicht mit hohem und kostspieligem Verwaltungsaufwand verbunden sein. Im Idealfall lässt sich das System mit den vorhandenen personellen Ressourcen umsetzen.
- 3) Die Indikatoren müssen transparent sein, d.h. alle Beteiligten müssen auch ohne, z.B. bibliometrisches Spezialwissen in der Lage sein, die Konstruktionsprinzipien der Leistungsindikatoren nachzuvollziehen.

Es ist offenkundig, dass in der Entwicklung von Leistungsindikatoren ein trade-off zwischen der fachspezifischen Einbettung (Punkt 1) und dem Verwaltungsaufwand in der Erhebung und Auswertung der Daten (Punkt 2) erfolgen muss. Die Entwicklung fachspezifischer Bewertungsstandards als Benchmarks ist mit verhältnismäßig hohem Aufwand verbunden, da diese Informationen zurzeit noch nicht bereitgestellt werden, sondern zum Teil noch entwickelt werden müssen.

Darüber hinaus müssen Leistungsindikatoren auf Ergebnisse wissenschaftlicher Arbeit bezogen sein, die aus Sicht der Professoren/innen eine aussagekräftige Bewertung der Leistungen zulassen. Ansonsten werden Leistungsindikatoren nicht als legitime Entscheidungsgrundlage in Fragen der Ressourcenverteilung akzeptiert.

Um in einem frühen Stadium der Entwicklung von Leistungsindikatoren sicherzustellen, dass die Indikatoren auf die größtmögliche Akzeptanz in der Fakultät treffen, wurde eine Befragung aller Professoren/innen der Philosophischen Fakultät durchgeführt, die unter anderem Fragen nach Art und Gewichtung möglicher Indikatoren enthielt. Im Folgenden werden die wichtigsten Ergebnisse dieser Befragung dargestellt (Abschnitt 2) und daraus Leitlinien für die Entwicklung von Leistungsindikatoren abgeleitet (Abschnitt 3). Zum Schluss wird ein kurzer Überblick über die Planung der weiteren Arbeitsschritte gegeben (Abschnitt 4).

## 2. Ergebnisse der Befragung der Professoren/innen der Philosophischen Fakultät

### 2.1 Anlage der Untersuchung

Die Befragung der Professoren/innen der Philosophischen Fakultät (ausgenommen waren lediglich die außerplanmäßigen Professoren/innen) wurde aus zwei Gründen durchgeführt. Erstens sollte ausgehend von der Literatur zur Bewertung wissenschaftlicher Leistungen untersucht werden, welche gängigen Indikatoren von den Professoren/innen der Fakultät zur Bewertung ihrer Leistungen akzeptiert werden. Insbesondere sollte auch die Bedeutung von Monografien innerhalb der Publikationskultur der verschiedenen Fächer erhoben werden (Abschnitt 2.2), weil Monografien mit Standardverfahren aus der Szientometrie nur schwer erfasst werden können. Zweitens dienen die Angaben zu Drittmittelgebern und Publikationsverhalten als empirische Grundlage einer Machbarkeitsstudie zur Umsetzbarkeit unterschiedlicher Leistungsindikatoren für die Fakultät.<sup>37</sup> Um ein differenziertes Bild über die Einstellungen der Professoren/innen zur Leistungsbewertung zu erhalten, enthielt der Fragebogen unter anderem Fragen zu Komponenten, Zwecken, Wirkungen und Verfahren der Leistungsbewertung.

Die Erhebung wurde vom 4. bis zum 20. September 2011 online durchgeführt. Dazu wurden zunächst alle 62 Professoren/innen der W- und C-Besoldung vom Dekan der Fakultät angeschrieben und um die Teilnahme an der Befragung gebeten. Am 4. September wurde allen Professoren/innen der Link zum Fragebogen geschickt. Nach 7 Tagen erfolgte eine erste Erinnerung für die Teilnehmer/innen, die noch nicht geantwortet hatten. Eine zweite Erinnerung erfolgte zwei Tage vor Ablauf der Teilnahmefrist. Insgesamt haben 55 der 62 Professoren/innen an der Befragung teilgenommen, was einer Ausschöpfungsquote von knapp 90 Prozent entspricht. Auch für Mitarbeiter/innenbefragungen ist dies ein guter Wert. Die durchschnittliche Beantwortungsdauer des Fragebogens betrug 21 Minuten.

### 2.2 Akzeptanz von Leistungsindikatoren an der Philosophischen Fakultät

In der Praxis werden unterschiedliche Indikatoren für die Bewertung wissenschaftlicher Leistungen verwendet, die jeweils mit spezifischen Vor- und Nachteilen einhergehen. Angesichts der Vielfalt der Fächerkulturen an der Philosophischen Fakultät, stellt sich daher die Frage, welche der einzelnen Indikatoren aus Sicht der Professoren/innen angewendet werden können, um wissenschaftliche Leistungen zu erfassen.

---

<sup>37</sup> Die Erörterung, welche Leistungsindikatoren unter welchen Bedingungen umgesetzt werden können, ist noch nicht Bestandteil dieses Berichts.



Dazu wurde in der Befragung ein Katalog von 16 Komponenten erstellt, aus dem die Befragten solche auswählen sollten, die ihrer Ansicht nach als Grundlage für die Bewertung von wissenschaftlichen Leistungen geeignet sind.

| Komponenten  | Prozent Zustimmung |
|--|--------------------|
| Monografien  | 91                 |
| Publikationen in Fachzeitschriften mit anonymen Begutachtungsverfahren             | 84                 |
| Publikationen in Herausgeber/innenbänden   | 76                 |
| Drittmittel der DFG  | 69                 |
| Drittmittel von gemeinnützigen Stiftungen  | 67                 |
| Publikationen in sonstigen Fachzeitschriften                                       | 58                 |
| Anzahl erfolgreicher Promotionen (Betreuung)                                       | 51                 |
| Drittmittel von Bundes- und Landesministerien                                      | 49                 |
| Drittmittel der Europäischen Union   | 49                 |
| Gutachten (Peer-Reviews) zur wissenschaftlichen Qualität von Publikationen         | 38                 |
| Anzahl erfolgreicher Habilitationen (Betreuung)                                    | 38                 |
| Sonstige Faktoren <sup>a</sup>   | 36                 |
| Sonstige Drittmittel   | 31                 |
| Drittmittel privater Unternehmen   | 22                 |
| Zitationen von Publikationen durch andere Wissenschaftler/innen (z.B. der h-Index) | 22                 |

Anmerkungen: <sup>a</sup> Insgesamt 19 Befragte haben auf die offene Frage nach weiteren Bewertungskriterien geantwortet. Dabei wird in erster Linie die Berücksichtigung von Leistungen in der Lehre eingefordert (5 Nennungen inkl. Betreuung von Abschlussarbeiten), die Gutachter/innentätigkeit für DFG und andere Organisation (2 Nennungen), Teilnahme an Fachkonferenzen (2 Nennungen), Organisation wissenschaftlicher Tagungen, Herausgabe von Fachzeitschriften, Rezensionen von Monografien, Job-Placement von Mitarbeitern, Zahl externer Rufe, interdisziplinäre Zusammenarbeit in Forschungsprojekten, Einladungen zu Vorträgen/Kongressen (mit jeweils einer Nennung). Anteile über 100 Prozent ergeben sich aus Mehrfachnennungen.

Tabelle 1: Zustimmung zu Komponenten der Bewertung wissenschaftlicher Leistungen (geordnet nach Höhe der Zustimmung in Prozent)

Die Ergebnisse in Tabelle 1 zeigen, dass insbesondere Publikationen als Grundlage zur Leistungsbewertung gewählt werden. Monografien werden mit 91 Prozent am häufigsten genannt, Publikationen in Fachzeitschriften mit Peer-Review folgen mit 84 Prozent und Publikationen in Herausgeber/innenbänden mit 76 Prozent. Die Einwerbung von Drittmitteln wird mit 69 Prozent Nennungen für Drittmittel der DFG und 67 Prozent für Drittmittel gemeinnütziger Stiftungen weniger einhellig als Indikator für wissenschaftliche Leistung anerkannt als die genannten Publikationsformen. Die genannten fünf Komponenten sind folglich jeweils für mehr als zwei Drittel der Professoren/innen als Indikatoren brauchbar. Deutlich wird in Tabelle 1 auch, welche Komponenten nicht zur Leistungsbewertung herangezogen werden sollten. Dazu gehören insbesondere personenbezogene Zitationsanalysen, Drittmittel von privaten Unternehmen oder sonstigen Organisationen, die Anzahl Habilitationen oder Gutachten zu Publikationen. Diese Komponenten werden von der Mehrheit

der Professoren/innen nicht genannt. Jeweils ungefähr die Hälfte der Professoren/innen nennt als Indikatoren auch Publikationen in sonstigen Fachzeitschriften, Anzahl der betreuten Promotionen und Drittmittel von Bundesministerien bzw. der EU.

Neben der einfachen Auswahl aus der Liste wurden die Befragten gebeten, die ausgewählten Komponenten nach ihrer relativen Wertigkeit zu gewichten. So wird nicht nur erfasst, welche Komponenten als Indikatoren brauchbar sind, sondern auch, wie stark sie in die Leistungsbewertung eingehen sollen, d.h. wie aussagekräftig sie für die Leistungsbewertung sind. Das Muster aus Tabelle 1 wird durch die Gewichtung nur leicht modifiziert, wenn die Ergebnisse für die Fakultät als Ganzes betrachtet werden, wie aus der letzten Zeile von Tabelle 2 hervorgeht. Zwar erhalten Monografien das höchste Gewicht (0,79), gefolgt von Aufsätzen in Fachzeitschriften mit Peer-Review (0,65), an dritter Stelle folgen jedoch Drittmittel der DFG (0,48). Beiträge in Herausgeber/innenbänden erreichen die vierthöchste Gewichtung (0,43) und Drittmittel von Stiftungen die fünfthöchste (0,42).

Stärker als in den einfachen Nennungen zeigen sich in den Gewichtungen Unterschiede zwischen den Instituten der Fakultät. Zwar werden Monografien für (fast) alle Fächer der Fakultät relativ hoch gewichtet, allerdings werden Fachzeitschriften mit Peer-Review in der Germanistik, Philosophie und den Sozialwissenschaften höher gewichtet. In den allgemeinen Sprachwissenschaften werden beide Publikationsformen annähernd gleich gewichtet. Insbesondere in den kleinen Fächern werden Artikel in Fachzeitschriften deutlich geringer gewichtet als Monografien (wobei die Jüdischen Studien hier die Ausnahme von der Regel darstellen). Drittmittel (insbesondere der DFG) als Indikator werden vor allem in der Romanistik, Philosophie, Medien- und Kulturwissenschaften und der Kunstgeschichte hoch gewichtet. Diese spielen wiederum in den weiteren kleinen Fächern, aber auch der Anglistik und der Germanistik, nur eine untergeordnete Rolle. Nachwuchsförderung wird nur in der Kunstgeschichte und in geringerem Ausmaß in den Jüdischen Studien als Indikator relativ stark gewichtet. In diesem Zusammenhang ist bemerkenswert, dass die Drittmittel von Stiftungen vielfach höher gewichtet wurden als Mittel der Ministerien oder der Europäischen Union (Anglistik, Sozialwissenschaften, Geschichtswissenschaften, Modernes Japan, Romanistik und z.T. auch Philosophie und Kunstgeschichte).

Tabelle 2 liefert folglich ein differenziertes Bild der Publikations- und Drittmittelkulturen der einzelnen Institute der Fakultät. Erstens finden sich Fächer, in denen die Akzeptanz der Leistungsevaluation maximiert wird, indem diese anhand von Publikationen stattfindet. In den Geschichtswissenschaften, der Klassischen Philologie und dem Modernen Japan sollten in erster Linie Monografien berücksichtigt werden. In der Germanistik, der Philosophie, den Sozialwissenschaften und den Allgemeinen Sprachwissenschaften darüber hinaus auch Artikel in Fachzeitschriften mit Peer-Review. Im Verhältnis von Monografien zu Aufsätzen in Fachzeitschriften – in den Naturwissenschaften das dominante Publikationsmedium – zeigen sich die Unterschiede zwischen den Instituten der Fakultät am deutlichsten.

Nur in den Instituten für Sozialwissenschaften (6 von 12) und für Sprache und Information (3 von 5) geben die Hälfte oder mehr Professoren/innen an, dass Fachzeitschriften das wichtigste Forum für die Publikation wissenschaftlicher Arbeiten ist. In den meisten anderen Instituten werden verschiedene Publikationsformen gleichberechtigt bewertet (Ergebnisse nicht tabellarisch ausgewiesen). Je wichtiger Monografien für die Publikationskultur des eigenen Faches eingeschätzt werden, umso höher fällt auch deren Gewichtung gegenüber anderen Komponenten der Leistungsevaluation aus.<sup>38</sup>

<sup>38</sup> Die Korrelation beträgt 0,315 und ist trotz der geringen Fallzahl am Fünfprozent-Niveau signifikant.



| Institut                            | Drittmittel |                  |            |                      |                       |               | Publikation                    |                             |                  |                                  |                |                 | Nachwuchsförderung    |                     |               |           |
|-------------------------------------|-------------|------------------|------------|----------------------|-----------------------|---------------|--------------------------------|-----------------------------|------------------|----------------------------------|----------------|-----------------|-----------------------|---------------------|---------------|-----------|
|                                     | DFG         | Minis-<br>terien | EU         | Stif-<br>tun-<br>gen | Unter-<br>neh-<br>men | Son-<br>stige | Zeitschriften<br>(Peer-Review) | Zeitschriften<br>(sonstige) | Mono-<br>graphen | Heraus-<br>geber/in-<br>nenbände | Cut-<br>achten | Zita-<br>tionen | Pro-<br>mo-<br>tionen | Habilita-<br>tionen | Son-<br>stige | N         |
| Anglistik und Amerikanistik         | .17         | .17              | .17        | .33                  | .07                   | .07           | .57                            | .27                         | .73              | .37                              | .23            | .27             | .30                   | .20                 | .60           | 5         |
| Germanistik                         | .21         | .21              | .11        | .11                  | .00                   | .00           | .80                            | .56                         | .70              | .66                              | .07            | .00             | .07                   | .07                 | .00           | 5         |
| Geschichtswissenschaften            | .56         | .21              | .33        | .64                  | .08                   | .33           | .59                            | .42                         | .98              | .42                              | .14            | .00             | .17                   | .17                 | .29           | 6         |
| Jüdische Studien                    | .17         | .13              | .21        | .17                  | .00                   | .00           | .56                            | .42                         | .63              | .42                              | .25            | .21             | .44                   | .44                 | .83           | 3         |
| Klassische Philologie               | .00         | .00              | .00        | .00                  | .00                   | .00           | .52                            | .35                         | 1.00             | .35                              | .33            | .00             | .17                   | .00                 | .17           | 2         |
| Kultur- und<br>Medienwissenschaften | .75         | .75              | .75        | .75                  | .19                   | .25           | .25                            | .56                         | .63              | .56                              | .00            | .00             | .50                   | .25                 | .50           | 2         |
| Kunstgeschichte                     | .72         | .69              | .17        | .69                  | .53                   | .69           | .56                            | .69                         | 1.00             | .58                              | .00            | .00             | .89                   | .33                 | .33           | 3         |
| Modernes Japan                      | .30         | .00              | .00        | .30                  | .00                   | .00           | .53                            | .30                         | .93              | .67                              | .00            | .17             | .33                   | .33                 | .00           | 2         |
| Philosophie                         | .75         | .55              | .64        | .64                  | .00                   | .08           | .89                            | .13                         | .69              | .30                              | .23            | .13             | .30                   | .17                 | .13           | 4         |
| Romanistik                          | .80         | .27              | .30        | .50                  | .20                   | .23           | .33                            | .37                         | .80              | .30                              | .35            | .20             | .23                   | .30                 | .33           | 5         |
| Sozialwissenschaften                | .56         | .24              | .40        | .46                  | .05                   | .17           | .79                            | .22                         | .68              | .30                              | .42            | .15             | .13                   | .11                 | .24           | 12        |
| Sprache und Information             | .45         | .15              | .33        | .25                  | .05                   | .05           | .83                            | .05                         | .88              | .58                              | .26            | .27             | .33                   | .33                 | .17           | 5         |
| <b>Insgesamt</b>                    | <b>.48</b>  | <b>.26</b>       | <b>.30</b> | <b>.42</b>           | <b>.08</b>            | <b>.16</b>    | <b>.65</b>                     | <b>.33</b>                  | <b>.79</b>       | <b>.43</b>                       | <b>.23</b>     | <b>.13</b>      | <b>.27</b>            | <b>.21</b>          | <b>.28</b>    | <b>55</b> |

Anmerkungen: Ausgewiesen werden die durchschnittlichen Gewichtungen der einzelnen Komponenten zur Bewertung wissenschaftlicher Leistungen. Um sicherzustellen, dass die durchschnittlichen Gewichtungen auf einer gemeinsamen Metrik beruhen (d.h. die relativen Gewichtungen durch die Befragten jeweils proportional in die Berechnung des Durchschnitts eingehen) wurde für jeden Befragten die höchste Gewichtung ermittelt und alle weiteren Gewichte durch den Höchstwert geteilt. Folglich erhält für jeden Befragten die Komponente mit der höchsten Gewichtung den Wert 1. Die Differenzen von 1 spiegeln proportional die Differenzen zwischen den einzelnen Komponenten wieder, wie sie auf den Skalenregeln des Fragebogens vergeben wurden. Die neue Variable hat deshalb einen Wertebereich von 0 (soll nicht in die Bewertung eingehen) bis 1 (soll in der Bewertung stark maximal werden).

Tabelle 2: Durchschnittliche Gewichtung der Komponenten zur Bewertung wissenschaftlicher Leistungen nach Instituten der Philosophischen Fakultät.

Zweitens gibt es Fächer, in denen auch Drittmittel, insbesondere der DFG, als Indikator anerkannt werden: der Romanistik, der Philosophie, der Kunstgeschichte und den Medien- und Kulturwissenschaften und in geringerem Maße auch die Sozialwissenschaften und die Geschichtswissenschaften. *Die institutsspezifischen Mittelwerte müssen jedoch sehr vorsichtig interpretiert werden, da die Fallzahlen zum Teil sehr klein ausfallen* (siehe die letzte Spalte in Tabelle 2).

Die Ergebnisse aus den Tabellen 1 und 2 zeigen, welche Komponenten für die Ausarbeitung von Leistungskriterien berücksichtigt werden müssen und welche fachspezifischen Anpassungen ggf. vorgesehen werden müssen, um einzelne Fächer nicht systematisch zu benachteiligen. In erster Linie müssen Indikatoren die Publikationsaktivität erfassen, ohne dabei Monografien und Herausgeber/innenbände zu vernachlässigen, wie dies in der Regel in Bewertungsverfahren in den Naturwissenschaften der Fall ist. Besonders hervorzuheben ist dabei die Rolle von Monografien in den Geisteswissenschaften. 80 Prozent der Professoren/innen geben an, dass Monografien in ihrem Fach eine sehr wichtige (60 Prozent) oder wichtige (20 Prozent) Rolle für die Vermittlung von Wissen spielen (Ergebnisse nicht tabellarisch ausgewiesen). Kein einziger hingegen berichtet, dass Monografien für das Fach unwichtig sind.

Wenn Drittmittel als Kriterien verwendet werden, sind insbesondere die DFG und gemeinnützige Stiftungen als Drittmittelgeber relevant. Die Gewichtung sowohl der DFG-Drittmittel als auch der anderen Drittmittel hängt nicht vom (selbstberichteten) Drittmittelauflkommen der Professoren/innen ab (die Korrelation ist nahe null), d.h. verhältnismäßig großer Erfolg in der Beantragung von Drittmitteln erhöht nicht die Akzeptanz dieser Komponente als Leistungsindikator.

Da neben der Forschung auch die Lehre zu den Kernaufgaben von Hochschullehrern/innen gehört, wurde gefragt, ob auch die Lehre in die Leistungsevaluation einfließen soll. Wie aus Tabelle 3 hervorgeht sind knapp 90 Prozent der Befragten der Meinung, dass eine Leistungsevaluation nicht ohne Berücksichtigung von Leistungen in der Lehre stattfinden soll. Als Leistungsdimensionen stehen dabei Ergebnisse aus der Lehrevaluation und die (quantitative) Lehrbelastung gemessen an der Zahl abgenommener Prüfungen und betreuter Abschlussarbeiten im Vordergrund. Ein Nachteil der quantitativen Indikatoren ist, dass die kleinen Fächer der Fakultät aufgrund geringerer Studierendenzahlen systematisch benachteiligt werden.

|   | Zustimmung in Prozent |
|---|-----------------------|
| Leistungen in der Lehre berücksichtigen?        | 86                    |
| Kriterien:                                      |                       |
| Ergebnisse der Lehrevaluation <sup>a</sup>      | 77                    |
| Größe der Veranstaltungen <sup>a</sup>          | 30                    |
| Anzahl abgenommener Prüfungen <sup>a</sup>      | 77                    |
| Anzahl betreuter Abschlussarbeiten <sup>a</sup> | 87                    |
| Andere Kriterien <sup>a,b</sup>                 | 26                    |

Anmerkungen: <sup>a</sup> Prozentuierung erfolgt auf der Basis der 86 Prozent der Befragten, die Leistungen in der Lehre für die Bewertung wissenschaftlicher Leistungen berücksichtigen wollen (Zeile 1 der Tabelle). <sup>b</sup> Insgesamt nennen 12 Befragte andere Kriterien: Anzahl der betreuten Praktika/Hausarbeiten, Auszeichnungen von Abschlussarbeiten, Exkursionen, Innovationen in der Lehre, Betreuung von E-Learning Kursen, Weiterbildungen und Auszeichnungen.

Tabelle 3: Zustimmung zur Berücksichtigung von Leistungen in der Lehre und Kriterien (Angaben in Prozent)



In der Frage zu welchen Zwecken Leistungsindikatoren eingesetzt werden, besteht relativ großer Konsens in der Fakultät (Tabelle 4). Die meisten Nennungen aus der Mehrfachauswahl entfallen auf die Vergabe von besonderen Leistungszuschlägen im Rahmen der W-Besoldung. An zweiter Stelle steht die leistungsorientierte Mittelvergabe innerhalb der Fakultät und die Evaluation von Professoren/innen im Rahmen von Bleibeverhandlungen (jeweils 65 Prozent), gefolgt von Berufungsverhandlungen (62 Prozent) und der Evaluation von Bewerbern/innen in Berufungsverfahren bzw. der Evaluation von Juniorprofessoren/innen (jeweils 60 Prozent).

| Verwendung für ...  | Zustimmung in Prozent |
|---|-----------------------|
| Vergabe von Leistungszuschlägen im Rahmen der W-Besoldung                         | 76                    |
| Leistungsorientierte Mittelvergabe innerhalb der Fakultät                         | 65                    |
| Bleibeverhandlungen   | 65                    |
| Berufungsverhandlungen  | 62                    |
| Evaluation von Bewerbern/innen im Rahmen von Berufungsverfahren                   | 60                    |
| Evaluation von Juniorprofessorinnen und Juniorprofessoren                         | 60                    |
| Vergabe interner Auszeichnungen   | 31                    |
| Leistungsorientierte Mittelvergabe unter den Fakultäten innerhalb der Universität | 29                    |
| Andere Zwecke <sup>a</sup>  | 15                    |
| Benchmarking verschiedener Lehrstühle innerhalb der Fakultät                      | 11                    |
| Vergleich der Fakultäten untereinander  | 5                     |

Anmerkungen: <sup>a</sup> Die unter „andere Zwecke“ genannten Komponenten können den Anmerkungen zu Tabelle 1 entnommen werden.

Tabelle 4: Zwecke der Verwendung von Leistungsindikatoren; geordnet nach Höhe der Zustimmung (Angaben in Prozent)

Eine Reihe von Zwecken wird von der Mehrheit der Befragten nicht unterstützt, z.B. die Vergabe interner Auszeichnungen, die Mittelvergabe unter den Fakultäten der Universität, der Vergleich der Fakultäten bzw. das Benchmarking der Institute. Tabelle 4 zeigt folglich, dass die Nützlichkeit von Leistungsindikatoren in der Fakultät grundsätzlich anerkannt wird, sowohl was die leistungsabhängigen Gehaltsanteile als auch was die leistungsorientierte Mittelvergabe innerhalb der Fakultät angeht.

Die grundsätzliche Akzeptanz von Leistungsindikatoren zeigt sich auch darin, dass etwa ein Drittel der Professoren/innen angibt, ihre wissenschaftliche Tätigkeit an Leistungsindikatoren auszurichten (vgl. Tabelle 5), wenn diese eingeführt werden. Ein weiteres Viertel nimmt eine neutrale Position zu dieser Frage ein. Nur ein Drittel gibt an, dass Leistungsindikatoren keinen Einfluss auf die wissenschaftliche Tätigkeit haben werden.

|                           | Prozent |
|---------------------------|---------|
| 1 = Ja, auf jeden Fall    | 13      |
| 2                         | 20      |
| 3                         | 27      |
| 4                         | 15      |
| 5 = Nein, auf keinen Fall | 16      |
| 6 = Weiß ich nicht        | 9       |

Anmerkungen: N=55

Tabelle 5: Ausrichtung der wissenschaftlichen Tätigkeit an Leistungsindikatoren

Um diesen Befund besser interpretieren zu können, wurden die Professoren/innen gefragt, ob sie nach der Einführung von Leistungsindikatoren beabsichtigen, Drittmittelprojekte zu beantragen und Artikel in Fachzeitschriften mit Peer-Review einzureichen. Davon hängt ab, ob die Einführung der Leistungsindikatoren eine Steuerungswirkung entfalten kann. Die Ergebnisse in Tabelle 6 (oberer Teil) zeigen, dass 36 bis 40 Prozent der Professoren/innen, die ihre Arbeit an Leistungsindikatoren ausrichten wollen, mit hoher Wahrscheinlichkeit vermehrt Drittmittel beantragen werden (Kategorien 1 und 2 der Antwortskala).

Die Bereitschaft vermehrt in Fachzeitschriften mit Peer-Review zu veröffentlichen, fällt geringer aus. Allerdings spiegelt dies insbesondere Unterschiede in der Publikationskultur wieder, da vor allem in den Instituten diese Bereitschaft gering ausfällt, in denen Publikationen in Fachzeitschriften für die Leistungsevaluation nur gering gewichtet werden (Ergebnisse nicht ausgewiesen).

| Wie wahrscheinlich ist es, dass Sie...  | 1 = sehr wahrscheinlich | 2    | 3    | 4    | 5 = sehr unwahrscheinlich |            | N  |
|---|-------------------------|------|------|------|---------------------------|------------|----|
| ... verstärkt Drittmittel bei der DFG beantragen?   | 12 %                    | 24 % | 27 % | 18 % | 18 %                      | -          | 33 |
| ... verstärkt Drittmittel bei anderen Organisationen als der DFG beantragen?  | 12 %                    | 28 % | 27 % | 15 % | 18 %                      | -          | 33 |
| ... vermehrt Aufsätze bei internationalen Fachzeitschriften mit anonymen Begutachtungsprozessen (Peer-Review) einreichen? | 9 %                     | 16 % | 41 % | 19 % | 15 %                      | -          | 33 |
| ... vermehrt Aufsätze bei nationalen Fachzeitschriften mit anonymen Begutachtungsprozessen (Peer-Review) einreichen?      | 6 %                     | 16 % | 41 % | 16 % | 22 %                      | -          | 32 |
| Wie hoch ist die Wahrscheinlichkeit, erfolgreich...   | 1 = sehr hoch           | 2    | 3    | 4    | 5 = sehr gering           | weiß nicht | N  |
| ... Drittmittel der DFG zu beantragen?  | 4 %                     | 22 % | 37 % | 11 % | 4 %                       | 22.2       | 27 |
| ... Drittmittel anderer Organisationen als der DFG zu beantragen?   | 4 %                     | 26 % | 44 % | 4 %  | 4 %                       | 18.5       | 27 |
| ... vermehrt in internationalen Fachzeitschriften mit Peer-Review zu publizieren?   | 15 %                    | 26 % | 33 % | 7 %  | 4 %                       | 14.8       | 27 |
| ... vermehrt in deutschsprachigen Fachzeitschriften mit Peer-Review zu publizieren?                                       | 12 %                    | 24 % | 40 % | 4 %  | 0 %                       | 20.0       | 25 |

Tabelle 6: Steuerungswirkung von Leistungsindikatoren (bezogen auf Drittmittel und Publikationen in Fachzeitschriften mit Peer-Review)



Den potentiellen Erfolg der Bemühungen um weitere Drittmittel bzw. die verstärkte Publikation von Artikeln in Fachzeitschriften mit Peer-Review (vgl. Tabelle 6 unterer Teil) abzuschätzen, fällt den Befragten relativ schwer, was an der starken Besetzung der Mittelkategorie und der „Weiß nicht“-Kategorie erkennbar ist. Auffällig ist, dass die Erfolgswahrscheinlichkeit von Publikationen in Zeitschriften allgemein höher eingeschätzt wird, als für die Beantragung von Drittmitteln, was auf die zahlreichen Kontingenzen eines wettbewerblich organisierten Vergabeverfahrens für Drittmittel verweist.

Was das Verfahren der Leistungsbewertung angeht, bevorzugt die Mehrheit der Professoren/innen die Erstellung eines Forschungsprofils aus den verschiedenen Dimensionen der Leistungsbewertung (Publikationen, Drittmittel, Lehre). Nur ein Siebtel spricht sich dafür aus, einen Summenindex aus mehreren Leistungsindikatoren zu bilden (vgl. Tabelle 7).

| Bewertungsverfahren                                 | Zustimmung in Prozent |
|---|-----------------------|
| Erstellung eines Tätigkeits- bzw. Forschungsprofils | 60                    |
| Bildung eines Leistungsindex                        | 16                    |
| Erstellung einer Liste mit K.O.-Kriterien           | 0                     |
| Andere <sup>a</sup>                                 | 9                     |
| Weiß nicht  | 15                    |

Anmerkungen: N=55, <sup>a</sup> 5 Befragte haben die Möglichkeit einer offenen Eingabe anderer Bewertungsverfahren verwendet, von denen einer der Erstellung eines Tätigkeits- bzw. Forschungsprofils gleicht. Ein Befragter möchte keine Leistungsbewertung. Andere Vorschläge sind ein Scoring unter Berücksichtigung des akademischen Alters und ein argumentierendes und begründendes Gutachten in Textform.

Tabelle 7: Zustimmung zu verschiedenen Bewertungsverfahren

Ein Vorteil in der Erstellung eines Tätigkeitsprofils liegt darin, dass keine (arbiträre) Gewichtung für die Summierung einzelner Komponenten gewählt werden muss (z.B. Drittmittel gegenüber Publikationen). Das Profil zeigt für jeden der einzelnen Indikatoren an, wie ein/e Professor/in im Vergleich zur Fakultät steht. Auf diese Weise können in der abschließenden Gesamtbewertung Besonderheiten der Fächer berücksichtigt werden.

### 3. Schlussfolgerungen: Leitlinien für die Entwicklung von Leistungsindikatoren

Aus der Befragung der Professoren/innen können zusammenfassend folgende Leitlinien für die Ausarbeitung von Leistungsindikatoren abgeleitet werden:

*Erstens* sollte die Leistungsbewertung in erster Linie auf Grundlage der Publikationsaktivität erfolgen. Eine Einschränkung auf Fachzeitschriften mit Peer-Review-Verfahren, wie sie für die Naturwissenschaften und zum Teil für die Wirtschaftswissenschaften üblich ist, kann für die Fächer der Philosophischen Fakultät keine valide Messung der Leistung ergeben. Neben Fachzeitschriften müssen deshalb mindestens Monografien in der Evaluation berücksichtigt werden. Soweit sinnvoll umsetzbar sollten auch weitere Publikationsformen in der Leistungsbewertung Berücksichtigung finden, wobei diskutiert werden muss, in welchem Verhältnis verschiedene Publikationsformen gewichtet werden.

*Zweitens*, werden Drittmittel in der Fakultät nur partiell als Maß wissenschaftlicher Leistungen anerkannt, was sich mit der kritischen Einschätzung von Drittmitteln als Indikator z.B. des Wissenschaftsrates deckt.

Werden Drittmittel dennoch zur Leistungsmessung verwendet, sind es vor allem Drittmittel der DFG und gemeinnütziger Stiftungen, die in die Bewertung einfließen sollten, da diese in einem kompetitiven, offenen Antragsverfahren vergeben werden und folglich keine Auftragsforschung darstellen.

*Drittens* sollte ein System der Leistungsbewertung auch Leistungen in der Lehre berücksichtigen, insbesondere Ergebnisse aus der Lehrevaluation und Indikatoren für die Lehrbelastung.

*Viertens* wird kein zusammenfassender Summen-Score über die verschiedenen Leistungsdimensionen (Publikationen, Drittmittel, Lehre) gewünscht. Vielmehr sollten die Dimensionen in Form eines wissenschaftlichen Tätigkeitsprofils getrennt ausgewiesen werden, wodurch u.a. den Besonderheiten einzelner Fächer (insbesondere der kleinen Fächer) in der Leistungsbewertung Rechnung getragen werden kann.

Insgesamt muss ein System der Leistungsbewertung folglich die verschiedenen Dimensionen wissenschaftlicher Arbeit abbilden, ohne dabei arbiträre Gewichtungen der einzelnen Dimensionen vorzunehmen. Die Herausforderung in der Entwicklung solcher Leistungsindikatoren liegt insbesondere darin, Verfahren zum Umgang mit verschiedenen Publikationsformen zu finden, da Monografien und Herausgeber/innenbände in der bibliometrischen Literatur bislang nicht ausreichend berücksichtigt wurden.



## ANHANG 4: ZUR BEFRAGUNG DER REKTORATE UND FAKULTÄTEN VERWENDETER FRAGEBOGEN

[Anmerkung: Diese Version des Fragebogens wurde an die Rektorate der Universitäten versandt. Der Fragebogen für die Fakultäten unterscheidet sich nur in den Größenskalen (Anzahl Studierende, Professoren/innen etc.) von diesem Fragebogen. Für die Befragung der philosophischen Fakultäten wurde ein leicht modifizierter Fragebogen eingesetzt, der auch Items zur Vergabe von besonderen Leistungszulagen in der W-Besoldung enthielt. Unterschiede im Fragebogendesign werden in der Beschreibung der empirischen Ergebnisse berichtet, wenn diese relevant sind.]

### Allgemeine Fragen zu Ihrer Universität

1. Handelt es sich bei Ihrer Universität um eine Technische Universität?

- Ja  Nein

2. In welchem Bundesland liegt Ihre Universität?

\_\_\_\_\_

3. Befindet sich Ihre Universität in staatlicher oder privater Trägerschaft?

- staatlich  privat  sonstige: \_\_\_\_\_

4. In welchem Jahr wurde Ihre Universität gegründet?

\_\_\_\_\_

5. Wie viele Professuren umfasst Ihre Universität (ohne Juniorprofessuren und Professuren der Medizinischen Fakultät)?

- bis 50  101 – 200  301 – 400  
 51 – 100  201 – 300  > 400

6. Wie viele Studierende waren im Sommersemester 2011 in den Studiengängen Ihrer Universität eingeschrieben?

- bis 5.000  10.000 – 20.000  30.000 – 40.000  
 5.000 – 10.000  20.000 – 30.000  > 40.000

### Fragen zur Beurteilung der Forschungsleistung an Ihrer Universität

7. Wird an Ihrer Universität die Forschungsleistung von Professoren/innen / Lehrstühlen regelmäßig beurteilt?

- Ja (weiter mit Frage 8)                       Nein (weiter mit Frage 21)

8. Wie lange gibt es an Ihrer Universität bereits ein System zur regelmäßigen Beurteilung der Forschungsleistung?

- weniger als 1 Jahr                       3 – 4 Jahre                       > 6 Jahre  
 1 – 2 Jahre                       5 – 6 Jahre

9. Welche Bedeutung haben folgende Kriterien im Rahmen der regelmäßigen Beurteilung der Forschungsleistung an Ihrer Universität?

|  | keine<br>Bedeutung       | geringe<br>Bedeutung     | mittlere<br>Bedeutung    | hohe<br>Bedeutung        |
|--|--------------------------|--------------------------|--------------------------|--------------------------|
| Höhe der eingeworbenen Drittmittel, die zur Beurteilung der Forschungsleistung herangezogen werden | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Anzahl der Publikationen in nationalen referierten Zeitschriften                                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Anzahl der Publikationen in internationalen referierten Zeitschriften                              | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Anzahl anderer Publikationen   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Gesamtzahl der Zitationen pro Professor/in / Lehrstuhl   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Andere bibliometrische Kennzahlen, wie z.B. h-Index, g-Index etc.                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Anzahl der erfolgreichen Promotionen pro Jahr  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Anzahl der erfolgreichen Habilitationen pro Jahr   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Erhaltene Auszeichnungen, Preise, Stipendien etc.  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| andere, und zwar:  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| _____  |                          |                          |                          |                          |
| _____  |                          |                          |                          |                          |



10. Falls mehrere Kriterien zur regelmäßigen Beurteilung der Forschungsleistung an Ihrer Universität herangezogen werden: Wie werden diese Kriterien miteinander verknüpft?

- Aus den genannten Kriterien wird eine Liste mit K.O.-Kriterien erstellt (Ausschluss-Methode)
- Aus den genannten Kriterien wird ein (Forschungs-)Profil erstellt (Profil-Methode)
- Die genannten Kriterien werden gewichtet und zu einer Kennzahl zusammengefasst (Scoring-Methode)
- Andere, und zwar: \_\_\_\_\_

11. Verwenden Sie im Rahmen der Beurteilung der Forschungsleistung an Ihrer Universität Zeitschriften-Rankings?

- Ja, und zwar folgende: \_\_\_\_\_
- Nein

12. Welche der nachfolgenden Drittmittel werden an Ihrer Universität im Rahmen der regelmäßigen Beurteilung der Forschungsleistung berücksichtigt? (Mehrfachnennungen möglich)

- Drittmittel werden zur Beurteilung der Forschungsleistung nicht herangezogen
- Drittmittel der Deutschen Forschungsgemeinschaft
- Drittmittel der Europäischen Union
- Drittmittel des Bundesministeriums für Bildung und Forschung
- Drittmittel anderer Bundesministerien
- Drittmittel von Landesministerien
- Drittmittel der Thyssen- und/oder VW-Stiftung
- Drittmittel anderer Stiftungen, und zwar v. a.: \_\_\_\_\_
- Drittmittel privater Unternehmen
- Drittmittel des Deutschen Akademischen Austausch Dienstes
- andere, und zwar: \_\_\_\_\_

13. Werden die Drittmittel aus den verschiedenen Quellen in irgendeiner Weise gewichtet?

- Ja, und zwar folgendermaßen: \_\_\_\_\_
- Nein

14. Für welche Zwecke wird das System zur regelmäßigen Beurteilung der Forschungsleistung an Ihrer Universität eingesetzt? (Mehrfachnennungen möglich)

- Leistungsorientierte Mittelvergabe innerhalb einer Fakultät
- Leistungsorientierte Mittelvergabe unter den Fakultäten innerhalb der Universität
- Vergabe interner Boni
- Vergabe interner Auszeichnungen
- Bleibeverhandlungen
- Berufungsverhandlungen
- Evaluation von Bewerbern/innen im Rahmen von Berufungsverfahren
- Evaluation von Juniorprofessoren/innen
- Benchmarking verschiedener Lehrstühle innerhalb einer Fakultät
- Vergleich der Fakultäten untereinander
- andere, und zwar: \_\_\_\_\_

15. Wird das System zur regelmäßigen Beurteilung der Forschungsleistung auf alle Fakultäten Ihrer Universität gleich angewendet?

- Ja, es gibt keine fakultätsspezifischen Unterschiede
- Nein, Unterschiede ergeben sich in folgenden Bereichen:  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

16. Wie hoch ist die Summe der Mittel, die innerhalb Ihrer Universität insgesamt leistungsorientiert vergeben wird?

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> es werden keine Mittel leistungsorientiert vergeben | <input type="checkbox"/> 500.000 – 1.000.000 EUR   | <input type="checkbox"/> 2.000.000 – 2.500.000 EUR |
| <input type="checkbox"/> bis 500.000 EUR                                     | <input type="checkbox"/> 1.000.000 – 1.500.000 EUR | <input type="checkbox"/> 2.500.000 – 3.000.000 EUR |
|  | <input type="checkbox"/> 1.500.000 – 2.000.000 EUR | <input type="checkbox"/> > 3.000.000 EUR           |

17. Gemessen an der Mittelzuweisung, die Ihre Universität von ihrem Träger erhält: Wie hoch ist der Anteil der Mittel, der an Ihrer Universität leistungsorientiert vergeben wird?

- |                                  |                                   |                                |
|----------------------------------|-----------------------------------|--------------------------------|
| <input type="checkbox"/> 1 – 5%  | <input type="checkbox"/> 11 – 20% | <input type="checkbox"/> > 30% |
| <input type="checkbox"/> 6 – 10% | <input type="checkbox"/> 21 – 30% |                                |



18. Wie bewerten Sie die Wirkung des Systems zur regelmäßigen Beurteilung der Forschungsleistung an Ihrer Universität?

|                                  | deutlich<br>verringert   |                          | unverändert              |                          | deutlich<br>erhöht       |                          |
|----------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Forschungsorientierung           | <input type="checkbox"/> |
| Drittmittelorientierung          | <input type="checkbox"/> |
| Publikationsorientierung         | <input type="checkbox"/> |
| Lehrorientierung                 | <input type="checkbox"/> |
| Andere, und zwar: _____<br>_____ | <input type="checkbox"/> |

19. Wie zufrieden sind Sie mit dem derzeitigen System zur regelmäßigen Beurteilung der Forschungsleistung an Ihrer Fakultät?

|                     |                          |                          |                     |                          |                          |                |                          |
|---------------------|--------------------------|--------------------------|---------------------|--------------------------|--------------------------|----------------|--------------------------|
| gar nicht zufrieden | <input type="checkbox"/> | <input type="checkbox"/> | teilweise zufrieden | <input type="checkbox"/> | <input type="checkbox"/> | sehr zufrieden | <input type="checkbox"/> |
|---------------------|--------------------------|--------------------------|---------------------|--------------------------|--------------------------|----------------|--------------------------|

20. Welche Verbesserungen könnten Sie sich vorstellen?

---



---



---

Füllen Sie bitte die folgenden drei Fragen nur aus, wenn an Ihrer Universität kein System zur regelmäßigen Beurteilung der Forschungsleistung existiert!

21. Ist es geplant, an Ihrer Universität ein System zur regelmäßigen Beurteilung der Forschungsleistung einzuführen?

- Ja, und zwar innerhalb der nächsten \_\_\_\_\_ Jahre.
  Ja, jedoch steht noch kein Zeitpunkt fest.
  Nein

22. Was sind aus Ihrer Sicht die Gründe, die gegen die Einführung eines Systems zur regelmäßigen Beurteilung der Forschungsleistung an Ihrer Universität sprechen?

|   | trifft<br>überhaupt<br>nicht zu |                          |                          | trifft voll<br>und ganz zu |                          |                          |
|---|---------------------------------|--------------------------|--------------------------|----------------------------|--------------------------|--------------------------|
| Die Notwendigkeit hat sich für unsere Universität bislang noch nicht gestellt.  | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| Die Notwendigkeit wurde bereits angesprochen. Die Umsetzung ist jedoch an internem Widerstand gescheitert.  | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| Es existieren bislang keine geeigneten Kriterien zu einer Beurteilung der Forschungsleistung.   | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| Die Beurteilung der Forschungsleistung ist mit einem zu hohen Aufwand verbunden.  | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| Die Beurteilung der Forschungsleistung führt zu einer einseitigen Konzentration auf die verwendeten Kriterien; andere wichtige Aspekte werden dabei vernachlässigt. | <input type="checkbox"/>        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |

23. Wie sollte Ihrer Meinung nach ein System zur regelmäßigen Beurteilung der Forschungsleistung ausgestaltet sein?

---



---



---



---

24. Haben Sie noch Anmerkungen / Kommentare zu unserer Studie?

---



---

Vielen Dank für Ihre Teilnahme!