

# Das Fächerrating des Landes Hessen – Ausgestaltung, Einsatz und Leistungsfähigkeit eines Instruments im Rahmen der Landes- hochschulsteuerung

Axel Oberschelp

---

Im Zuge des „New Public Management“ sind verschiedene Verfahren der Leistungsmessung und des Leistungsvergleichs für Zwecke der Hochschulsteuerung eingeführt worden. Deren Einsatz wird häufig kritisch diskutiert. In jüngster Zeit kommen auch Ratings im Kontext der Landeshochschulsteuerung zum Einsatz, wobei noch unklar ist, wie leistungsfähig dieses Instrument im Steuerungskontext tatsächlich ist. In diesem Beitrag wird ausführlich das in den Jahren 2013 bis 2016 in mehreren Phasen durchgeführte hessische Fächerrating besprochen, bei dem Peer Review Bestandteil einer fachbezogenen Bewertung war. Ausgehend von einer Typologie der Steuerungsinstrumente werden zunächst die Besonderheiten von Rating-Verfahren herausgearbeitet. Die detaillierte Darstellung des Vorgehens in Hessen hat dann zum Ziel, zu einer Einschätzung der Leistungsfähigkeit von Ratings zu kommen, wobei auch zu thematisieren ist, inwieweit die Anwendung im Steuerungskontext zu Verfahrensmodifikationen führt und wie weitreichend diese sind. Als weiteren Aspekt nimmt die Untersuchung die Rolle von Peer Review-Verfahren im Ratingkontext in den Blick und fragt nach hierbei festzustellenden besonderen Anforderungen an die Mitglieder von Bewertungskommissionen.

---

## 1 Einleitung

Im deutschen Wissenschaftssystem sind für Zwecke des Leistungsvergleichs konzipierte Instrumente mittlerweile auf den unterschiedlichen Ebenen etabliert. Ihre Einführung ging einher mit Veränderungen, die das Verhältnis zwischen Staat und Hochschulen einerseits und die Governance-Strukturen innerhalb der Hochschulen andererseits betreffen und die unter dem Begriff des „New Public Management“ (NPM) subsummiert werden (*Grande et al. 2014; Schedler/Proeller 2011; Burtscheidt 2010*). Im Kontext der Landeshochschulsteuerung haben insbesondere die auf Leistungsvergleichen basierenden Verfahren der Leistungsorientierten Mittelvergabe (LOM) sowie Zielvereinbarungen weite Verbreitung gefunden (*Jaeger 2009*). Im Rahmen von Hochschulvergleichen sind Ratings bzw. Rankings<sup>1</sup> mit unterschiedlichen

---

<sup>1</sup>Auf die Gemeinsamkeiten und Unterschiede beider Verfahren wird im zweiten Abschnitt näher eingegangen.

Akteuren als Adressaten gebräuchlich. Das Forschungsrating des Wissenschaftsrats als ein ambitioniertes und im Wissenschaftssystem exponiertes Verfahren ist dabei besonders hervorzuheben. Innerhalb einzelner Organisationen ermöglichen Benchmarking-Verfahren spezifische Formen des Leistungsvergleichs. Mit den mittlerweile punktuell eingeführten Leistungsbezügen bei der Besoldung von Professoren und Professorinnen ist ein entsprechendes Instrumentarium auch auf der Ebene einzelner Individuen zu finden.

Derzeit wird die Wirksamkeit der Instrumente des NPM im Kontext der Steuerung von Hochschulen intensiv diskutiert (*Grande et al. 2014*). Die vorliegende Studie ist innerhalb dieser Diskussion zu verorten und gibt mit Blick auf das in Hessen eingeführte Ratingverfahren, dessen fächer- und standortbezogene Ergebnisse und Bewertungen die Landeshochschulentwicklungsplanung unterstützen sollten, Erfahrungen und Einschätzungen wieder, die unmittelbar aus der Projektarbeit resultieren. Über dieses aus hochschulpolitischer Sicht möglicherweise wegweisende Vorhaben des Landes Hessen möchte der vorliegende Beitrag somit als Erfahrungs- und Werkstattbericht informieren.<sup>2</sup>

Er nimmt die Landeshochschulsteuerung in den Blick und untersucht den seit einigen Jahren festzustellenden, auf dieser Ebene neuen Einsatz des Instruments Rating. Mit der Durchführung eines sogenannten Hochschulratings in den Jahren 2008/09 hat der Freistaat Sachsen erstmals ein Rating zu Steuerungszwecken eingesetzt. Von diesem Verfahren gingen erhebliche Impulse auf das hier im Mittelpunkt der Darstellung stehende hessische Fächerrating aus, das in den Jahren 2013/14 konzipiert und implementiert wurde. Das für die Konzeption und den Einsatz verantwortliche Hessische Ministerium für Wissenschaft und Kunst (HMWK) war außer an einer für den Hochschulvergleich geeigneten Datengrundlage auch an einer umfassenden und fachbezogenen Standort*bewertung* interessiert. Die Ergebnisse sollten für die staatliche Aufgabe der Landeshochschulentwicklungsplanung Verwendung finden.

Die Einführung von Ratings im Kontext der Landeshochschulpolitik ging von der Annahme aus, dass es sich hierbei um ein leistungsfähiges Instrument zur Unterstützung hochschulpolitischer Aufgaben und Zielsetzungen handelt.<sup>3</sup> Die Überprüfung dieser Hypothese steht im Mittelpunkt des vorliegenden Beitrags. In diesem Zusam-

<sup>2</sup>Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) war an der Konzeptionierung des hessischen Fächerratings von Beginn an beteiligt, hat das Hessische Ministerium für Wissenschaft und Kunst (HMWK) während aller Phasen des Ratings beratend unterstützt und die Durchführung des Ratings federführend übernommen. Der Schwerpunkt der Tätigkeit des Verfassers lag insbesondere auf der Durchführung und Unterstützung des Peer Review-Verfahrens. Eine wissenschaftliche Begleitforschung zu diesem Projekt, die notwendig wäre, um die hier getroffenen Einschätzungen abzusichern bzw. weitere Erkenntnisse zu generieren, war dabei weder konzeptionell angelegt, noch ist sie zum jetzigen Zeitpunkt geplant.

<sup>3</sup>Um welche Aufgaben und Zielsetzungen es sich dabei handelt, wird in Abschnitt 2.2 erläutert.

menhang wird außerdem der Frage nachgegangen, ob Peer Review als ein Element der hier untersuchten Ratingverfahren für eine Bewertung der Leistungsfähigkeit von Organisationen geeignet ist oder ob sich möglicherweise die Funktion von Peer Reviews im Kontext von Ratings ändert. Damit kommt ein Aspekt zur Sprache, der für das Wissenschaftssystem insgesamt von erheblicher Bedeutung ist.

Es erfolgt zunächst eine Verortung von Ratings innerhalb des für Leistungsmessungen und -vergleiche zur Verfügung stehenden Instrumentariums (Abschnitt 2), wobei detailliert auf die Bereiche Leistungsmessung, Leistungsbewertung, Ergebnisverwendung und Partizipation eingegangen wird. Auf die Besonderheiten von Ratings und die Unterschiede zu anderen Verfahren, insbesondere zu Rankings, wird dabei ebenso eingegangen wie auf die Erwartungen, die aufseiten hochschulpolitischer Akteure, insbesondere der zuständigen Ministerien, mit dem Einsatz dieses Instruments verbunden waren. Mit dem sächsischen Hochschulrating wird dabei ein für Vergleichszwecke heranzuziehendes und für die Ausgestaltung des hessischen Verfahrens beispielgebendes Rating in den Blick genommen.

Anschließend erfolgt die Darstellung des in Hessen eingeführten Rating-Verfahrens (3), ebenfalls den oben bereits genannten Bereichen folgend. Im Abschnitt 3.2 dieses Kapitels kommen dabei die Besonderheiten des Einsatzes von Peer Review im Ratingkontext zur Sprache.

In der Zusammenfassung im letzten Abschnitt (4) wird zunächst die Frage diskutiert, inwieweit das hessische Fächerrating überhaupt dem zuvor definierten Typus eines Ratings entspricht. Anschließend werden die wichtigsten Ergebnisse und deren Bewertung für eine Überprüfung der oben genannten Hypothese zur Leistungsfähigkeit von Ratings im Steuerungskontext herangezogen. Hieraus sind einerseits Empfehlungen für den künftigen Instrumenteneinsatz abzuleiten, andererseits sind weitere Forschungsfragen bzw. „Leerstellen“ der vorliegenden Untersuchung zu identifizieren, deren Bearbeitung zu einer noch verlässlicheren Überprüfung der hier behandelten Hypothese führen könnte.

## **2 Instrumente der Leistungsmessung und des Leistungsvergleichs im Steuerungskontext**

### **2.1 Überblick**

Der seit den 1980er-Jahren festzustellende Wandel im Wissenschaftssystem hat die Rahmenbedingungen und die Praxis der staatlichen Steuerung von Hochschulen deutlich verändert. Einhergehend mit einer größeren Autonomie der Organisationen wird zunehmend Transparenz über die erbrachten Leistungen eingefordert. Staatliche Steuerungsmechanismen orientieren sich nicht mehr an den Details der Rechen-

schaftslegung, sondern an den Outputs (Leistungen) der Hochschulen, für deren Messung und Bewertung – besonders unter Aspekten der Vergleichbarkeit – spezifische Instrumente entwickelt wurden.

In diesem Abschnitt werden die Charakteristika der wichtigsten Instrumente der Leistungsmessung und des Leistungsvergleichs im Steuerungskontext (Evaluationen, Leistungsorientierte Mittelvergabe, Rankings, Zielvereinbarungen) skizziert, um anschließend die Besonderheiten des Instruments Rating besser herausstellen zu können. Die Darstellung der Instrumente orientiert sich in der Reihenfolge am Zeitpunkt ihrer Einführung in die Landeshochschulsysteme und folgt den Gliederungsaspekten Leistungsmessung, Leistungsbewertung, Ergebnisverwendung und Partizipation. Benchmarking-Verfahren, ein weiteres Instrument im Baukasten des NPM, werden an dieser Stelle hingegen nicht behandelt. Sie spielen für den Steuerungskontext Land-Hochschulen keine maßgebliche Rolle.

Der Einsatz der hier behandelten Instrumente war von Beginn an von Diskussion und kritischen Einschätzungen zu grundsätzlichen methodischen Problemen bezüglich der Messung und Bewertung hochschulischer Leistungen begleitet (*Münch/Pechmann 2009*). Nach wie vor Gültigkeit hat eine Einschätzung aus dem Jahr 2009, wonach eine für die Bewältigung dieser Schwierigkeiten notwendige theoretische Fundierung zu den „Produktions-, Austausch und Bewertungsprozessen in der Wissenschaft“ (*Hornbostel et al. 2009*) mit Blick auf Forschung und Lehre noch nicht vorliegt. Entsprechend ist auch der Einsatz von Indikatoren, die diese Leistungen abbilden sollen, häufig anlassbezogen gewählt und mit Blick auf das Wissenschaftssystem insgesamt unsystematisch. Andererseits speist sich die Kritik aus der konkreten Ausgestaltung einzelner Instrumente sowie aus den nicht-intendierten Folgen des Instrumenteneinsatzes (*Krempkow et al. 2013*). Insbesondere hinsichtlich des Einsatzes des Instruments Ranking, jedoch auch im Zusammenhang mit anderen leistungsmessenden Verfahren, ist diese Kritik teils vehement formuliert worden. Demnach führt die Messung hochschulischer Leistungen zu ausschnitthaften und stark vereinfachenden Darstellungen, schafft damit eine eigene Realität und bildet die Wirklichkeit nur verzerrt oder verkürzt ab (*Borgwardt 2011 S. 7*). Nach *Hornbostel (2007) S. 7* führt bei Rankings die Orientierung an den verwendeten Indikatoren dazu, dass es zu einer Optimierung der Ausprägungen dieser Indikatoren anstatt zu der intendierten Qualitätsverbesserung komme. Mit Blick auf einzelne Rankings zielt die Kritik auf deren ungenügende Datengrundlage und die teilweise nicht gegebene Gewährleistung grundsätzlicher Standards (Objektivität, Reliabilität und Validität) wissenschaftlicher Untersuchungen (*Lenzen 2012*).

## 2.1.1 Evaluationen<sup>4</sup>

Viele Landesbehörden haben schon vor der Einführung des NPM evaluative Verfahren eingesetzt, um externe Expertise für die Weiterentwicklung ihrer Hochschulsysteme zu nutzen. Die externe Expertise wird in Form von Begutachtungskommissionen organisiert, die elementarer Bestandteil dieser nach wie vor gebräuchlichen Verfahren sind. Neben der Analyse von Detailfragen (z. B. Lehrerbildung, Dienstrecht) wird dieses Instrument auch bei der Beurteilung von Fachdisziplinen, einzelnen Organisationen oder Landeshochschulsystemen eingesetzt, wobei der formative Charakter dieses Typs von Evaluation gegenüber der retrospektiven Bewertung der Leistungsfähigkeit deutlich im Vordergrund steht. Beispiele für anlassbezogene Evaluationen sind die Tätigkeit der Expertenkommission „Wissenschaftsland Bayern 2020“, die im Jahr 2005 eine Analyse der Perspektiven der Hochschullandschaft des Freistaats vorlegte (*Expertenkommission Bayern 2005*), sowie die Evaluation des Hochschulsystems in Schleswig-Holstein durch eine „Expertenkommission Hochschulentwicklung“ im Jahr 2003 (*Expertenkommission Schleswig-Holstein 2003*).

Ein in diesem Kontext noch eher neues Verfahren liegt mit den Länderstrukturbegutachtungen des Wissenschaftsrats vor. Da diese Form der Begutachtung in den letzten Jahren intensiv zum Einsatz kam – 2016 für die MINT-Bereiche der Hochschulen Hamburgs (*Wissenschaftsrat 2016*), 2014 für das Saarland (*Wissenschaftsrat 2014*), 2013 für die Hansestadt Bremen (*Wissenschaftsrat 2013a*) und für das Land Sachsen-Anhalt (*Wissenschaftsrat 2013b*) – wird auf dieses Verfahren näher eingegangen. Neben einer allgemeinen Zielsetzung, die auf die Bewertung des Hochschulsystems eines Bundeslandes und auf das Aufzeigen von Entwicklungsperspektiven gerichtet ist, wird bei den Begutachtungen durch den Wissenschaftsrat mit dem jeweiligen Auftraggeber eine spezifische Konkretisierung der Zielsetzung vereinbart.<sup>5</sup> Die Bewertung der einzelnen Standorte erfolgt durch eine Sachverständigenkommission auf Grundlage schriftlicher Selbstberichte und darauf folgender Begehungen der begutachteten Einrichtungen. Die Selbstberichte beinhalten unter anderem einen Datenteil, bestehend aus Daten der amtlichen Statistik sowie eigenen Daten der Hochschule, und nehmen auch die Rahmenbedingungen der Standorte in den Blick. Die Einschätzungen und Empfehlungen der Kommissionen werden der beauftragenden Behörde in Form eines Gutachtens zur Verfügung gestellt (*Wissenschaftsrat 2016*). Bestandteile der Bewertung sind Stärken-Schwächen-Analysen sowie die Identifizierung von Entwicklungsperspektiven. Im Gutachten erfolgt keine vergleichende Gegenüberstellung der zu den Standorten vorliegenden quantitativen Daten. Die Analysen und Empfeh-

<sup>4</sup>In diesem Abschnitt werden ausschließlich extern initiierte Verfahren behandelt, nicht jedoch innerhalb der Hochschulen angesiedelte Verfahren institutioneller Evaluationen.

<sup>5</sup>Für die Evaluation universitätsmedizinischer Einrichtungen hat der Wissenschaftsrat z. B. ein eigenes, in höherem Maße standardisiertes Verfahren entwickelt (*Wissenschaftsrat 2012*).

lungen führen zu keiner Ergebnisverdichtung in Form von Ranglisten oder -gruppen, sondern kommen ausschließlich in textlicher Form zum Ausdruck. Auch hierbei wird auf eine vergleichende Perspektive weitgehend verzichtet, sodass Bewertungen und Empfehlungen mit einem deutlichen Fokus auf die jeweiligen institutionenspezifischen Befunde vorgenommen werden. Die Mitwirkungsmöglichkeiten der Hochschulen liegen einerseits bei der Redaktion des Selbstberichts, andererseits können im Rahmen der Begehungen die Einschätzungen und Standpunkte der Hochschule in das Verfahren eingebracht werden.

Die von 2000 bis 2012 durchgeführten Forschungsevaluationen der wissenschaftlichen Kommission des Landes Niedersachsen sind ein Beispiel für den zyklischen Einsatz dieses Instruments, das bei einer mit den Begutachtungen des Wissenschaftsrats vergleichbaren Verfahrensstruktur ausschließlich den Bereich Forschung in den Blick nimmt. Eine Besonderheit dieses Verfahrens ist es, dass die durchführende Einrichtung in die Umsetzungsprozesse (Monitoring) der im Zuge des Bewertungsverfahrens getroffenen Empfehlungen involviert ist (*Wissenschaftliche Kommission Niedersachsen 2006 S. 45 und 48*). Evaluierungsagenturen führen ebenfalls in dieser Art strukturierte, fachbezogene Bewertungen durch, bei denen jedoch kein Vergleich von Standorten stattfindet.

### **2.1.2 Leistungsorientierte Mittelzuweisung (LOM)**

Es handelt sich um ein in der Steuerungspraxis aller Bundesländer seit den ausgehenden 1990er-Jahren eingesetztes Instrument (vgl. die aktuelle Übersicht bei *Dohmen (2015) S. 4 f.*). Dabei lässt sich in der Praxis zwar eine Vielzahl von Varianten und Verfahrensausprägungen feststellen, diese folgen jedoch dem grundlegenden Konstruktionsprinzip von LOM-Verfahren, der Zuweisung von Haushaltsmitteln anhand von definierten, quantitativen Leistungsparametern. Es handelt sich um ein anreizorientiertes Verfahren, bei dem überdurchschnittliche Leistungen honoriert werden. Charakteristisch für LOM-Verfahren ist eine ausgeprägt wettbewerbliche Situation der eingebundenen Hochschulen. Die Herstellung von Vergleichbarkeit ist eine zentrale Anforderung an die Leistungsmessung im Rahmen von LOM, was unter anderem zur Folge hat, dass vorwiegend leicht zu messende, quantitative Sachverhalte erfasst werden. Dadurch kommen nur Ausschnitte des gesamten Leistungsvermögens einer Einrichtung zur Abbildung. Ein Kritikpunkt zielt auf den Automatismus von Leistungsmessung und Mittelzuweisung, der die LOM anfällig für Fehlsteuerungen macht. Wird z. B. die Anzahl der Studienabschlüsse honoriert, kann das aufseiten der Hochschulen zu Anstrengungen führen, den quantitativen Output zulasten des Ausbildungslevels zu steigern. Die real festzustellenden Umverteilungen durch LOM-Verfahren sind jedoch, nicht zuletzt, weil in vielen Finanzierungsmodellen Kappungsgrenzen implementiert sind (*Bogumil et al. 2013 S. 33*), zumeist von geringem Umfang. Die Steue-

rungswirkung von LOM-Verfahren sollte deshalb nicht überschätzt werden. Wirksam ist dieses Instrument nach Einschätzung von *Bogumil et al. (2014) S. 54–55* noch am ehesten im Bereich Forschung sowie als Anreiz für Leistungssteigerung bei quantitativ gut messbaren Sachverhalten. Hingegen ist für den Bereich Lehre und für qualitative Aspekte der Leistungserstellung von einer derzeit eher schwachen Steuerungs- und Anreizfunktion auszugehen. Der partizipative Charakter ist bei diesem Instrument eher schwach ausgeprägt. Mitwirkungsmöglichkeiten der zu bewertenden Einrichtungen bestehen – je nach Verfahrenszuschnitt – allenfalls in der Phase der Instrumentenentwicklung, nicht jedoch beim laufenden Einsatz, der durch einen Automatismus von Leistungsmessung und Ergebnisverwendung und von dem Fehlen einer Interpretation der Ergebnisse der Leistungsmessung gekennzeichnet ist.

### 2.1.3 Rankings

Beginnend mit der Einführung des Hochschulrankings des Centrums für Hochschulentwicklung (CHE) im Jahre 1998 (*Berghoff et al. 2001; Wissenschaftsrat 2004 S. 19*) hat dieses Instrument auch im deutschsprachigen Raum kontinuierlich an Bedeutung gewonnen.<sup>6</sup> Damit wurde eine zuvor schon in den angelsächsischen Ländern festzustellende Entwicklung übernommen. Vereinfachend lassen sich hinsichtlich des Adressatenkreises die folgenden Unterschiede der verschiedenen Verfahren erkennen. Lehrbezogene Rankings wie z. B. das CHE-Hochschulranking oder das Universitäten-Ranking der Zeitschrift *Wirtschaftswoche* richten sich an eine breite Öffentlichkeit. Hingegen sind Vergleiche, die Leistungsunterschiede im Bereich Forschung darstellen, wie z. B. das Forschungsrating des Wissenschaftsrats, das CHE-Ranking „Vielfältige Exzellenz“ (bis 2009: Forschungsranking) sowie die Rankings des Handelsblatts für die Fächer Volkswirtschaftslehre und Betriebswirtschaftslehre an einen eher kleineren Adressatenkreis gerichtet. Dies gilt auch für Verfahren, die spezifische Teilaspekte in den Blick nehmen. Dazu zählen der Förderatlas der Deutschen Forschungsgemeinschaft, das Ranking der Alexander-von-Humboldt-Stiftung und das Gleichstellungsaspekte behandelnde Ranking (CEWS-Ranking) des Leibniz-Instituts für Sozialwissenschaften (gesis).

Es ist ein Charakteristikum von Rankings, dass eine Operationalisierung der Leistungen von Hochschulen durch quantitative Sachverhalte abbildende Indikatoren erfolgt, und dass die Ergebnisse der Leistungsmessung in einer stark verdichtenden und vergleichenden Form ausgewiesen werden. Während für die Messung selbst je nach Instrumentenzuschnitt eine Vielzahl von Verfahren (Befragungen, Auswertungen öffentlich zugänglicher Datenbanken, Datenabfrage bei den zu bewerteten Einrichtungen) zum Einsatz kommen kann, ist das Ziel in allen Fällen die Verortung einer Organisation in einer Rangliste. Die Ergebnisse von Rankings führen zu einer vertikalen Differenzierung

---

<sup>6</sup>Einen Überblick über die Entwicklung der globalen Rankings, die hier nicht im Mittelpunkt der Darstellung stehen, gibt *Hazelkorn (2013)*.

der Hochschullandschaft und stellen damit ein stark wettbewerbles Element dar. Bezüglich des Einsatzes des Instruments Ranking im Kontext der Landeshochschulsteuerung ist zunächst zu konstatieren, dass eine direkte Verwendung, bei der Ranking-Ergebnisse unmittelbar Eingang in Struktur- oder Ressourcenentscheidungen fänden – etwa über die Kopplung mit anderen Instrumenten –, nicht stattfindet. Nachweislich haben Rankings über ihre große Öffentlichkeitswirksamkeit jedoch einen erheblichen Einfluss auf die Entscheidungsträger auf dem Feld der Hochschulpolitik. Das von globalen Ranking-Ergebnissen beeinflusste Engagement des Bundes im Rahmen der Exzellenzinitiative ist ein Beispiel hierfür (*Hazelkorn/Ryan 2013*). Zwar sind die zu bewertenden Einrichtungen insofern in das Verfahren involviert, als sie – zumindest bei einigen Rankings – bei der Datenerhebung mitwirken. Sie haben jedoch weder Einfluss auf die Auswahl der zu erhebenden Daten noch auf die Methodik ihrer Verrechnung. Auch die Ergebnisdarstellung kann von den Hochschulen nicht mitgestaltet werden, sodass insgesamt ein geringes Maß an Partizipation vorliegt.

#### **2.1.4 Zielvereinbarungen**

Wie bei der leistungsorientierten Mittelvergabe handelt es sich bei Zielvereinbarungen (ZV; auch die Bezeichnung Hochschulverträge ist gebräuchlich) um anreizbasierte Verfahren, die seit den ausgehenden 1990er-Jahren im Hochschulbereich eingeführt wurden und mittlerweile in allen deutschen Ländern im Steuerungsfeld Land-Hochschule im Einsatz sind (*König 2009; In der Smitten/Jaeger 2012*). Die mit dem Einsatz dieses Instruments verbundenen Ziele sind vor allem das Setzen von Leistungsanreizen und die Förderung der dezentralen Strategiebildung. Kennzeichnend für ZV ist, dass die zu erreichenden Ziele auf Verhandlungsbasis zwischen den Partnern vereinbart werden. Da die Hochschulen von der landesseitigen Finanzierung abhängig sind, ist Gleichrangigkeit in Verhandlungen nicht möglich (*In der Smitten/Jaeger 2012 S. 5*). Gleichwohl lassen sich sowohl Länder identifizieren, in denen die Ziele eher hierarchisch („top-down“) vorgegeben werden, als auch solche mit einer stärker partnerschaftlichen Zielformulierung, bei der auch „bottom-up“ eingebrachte Ziele Berücksichtigung finden (*König 2009 S. 34*). Grundsätzlich werden in diesen Vereinbarungen Ziele formuliert, nicht jedoch Maßnahmen zur Zielerreichung. Die Ausgestaltung dieser Maßnahmen liegt vielmehr im Ermessen desjenigen Vertragspartners, der die Leistungen zu erbringen hat. Die Messung hochschulischer Leistungen ist bei diesem Verfahren an der Stelle von Bedeutung, wenn der Grad der Zielerreichung zu bestimmen ist. Welche Messverfahren dabei zum Einsatz kommen und wie diese instrumentenspezifisch zuzuschneiden sind, hängt wesentlich davon ab, inwieweit die Ziele konkretisiert respektive zählbar formuliert wurden. Je allgemeiner die Ziele formuliert sind, umso schwerer ist die Überprüfung der Zielerreichung. Allerdings spielt das Kriterium der guten Zählbarkeit eine eher untergeordnete Rolle im Vergleich zu LOM-Verfahren. Der Fokus ist bei ZV stärker auf innovative und zukunftsorientierte



Vorhaben gerichtet, häufig auch unter Einbeziehung qualitativer Aspekte, die einer Messbarkeit ohnehin nur in eingeschränktem Maße zugänglich sind. Die Bewertung der erzielten Leistungen erfolgt gemeinsam durch die Hochschule und das zuständige Ministerium. Ein Hinzuziehen externer Expertise ist hierbei als eine zusätzliche Option vereinzelt feststellbar, aber nicht der Regelfall (*In der Smitten/Jaeger 2012 S. 73*). Sofern die Ziele nur in geringem Maße operationalisierbar formuliert sind, kann sich ein zum Teil erheblicher Interpretationsspielraum bei der Bestimmung des Grades der Zielerreichung ergeben. Würden für einzelne Ziele zudem erst noch zu schaffende Voraussetzungen vereinbart, z. B. baulicher Art, für deren Schaffung nicht die Hochschule verantwortlich ist, ergeben sich weitere Schwierigkeiten für eine Leistungsbeurteilung. Eine Kopplung von Zielerreichung und Mittelzuweisung ist möglich, wird aber in den Ländern durchaus unterschiedlich gehandhabt und ist keinesfalls das Standardverfahren. Daneben ist der Aspekt der Strategieentwicklung als eine weitere Form der Ergebnisverwendung zu nennen. Es handelt sich bei ZV um ein Instrument der partizipativen Organisationsführung, bei dem die relevanten Akteure in den Hochschulen in vielfältiger Weise einbezogen werden: Zuvorderst ist hier der Prozess der Aushandlung von Zielen zu nennen, dann die Messung und Berichterstattung über den Grad der Zielerreichung, in der Regel beides im Verantwortungsbereich der Hochschulen liegend, und schließlich die gemeinsame Bewertung der erzielten Leistungen durch die Hochschulen und die zuständigen Behörde.

## 2.2 Das Instrument Rating im Kontext der Steuerung von Hochschulen

Während Rankings die Verortung einer Organisation in einer Rangliste zum Ziel haben, erfolgt bei Ratings die Bewertung einzelner Leistungsdimensionen nach einer vorgegebenen Notenskala, wobei Noten auch mehrfach vergeben werden können (*Hornbostel 2007 S. 6*). Anders als bei Rankings ist die Verrechnung mehrerer Leistungsdimensionen zu einem Gesamtwert nicht Bestandteil von Rating-Verfahren. Insbesondere diese hohe Aggregation von Ergebnissen war Gegenstand der zum Teil massiven Kritik an einzelnen Ranking-Verfahren in der Vergangenheit. Mit multidimensionalen Ergebnisdarstellungen und dem Verzicht auf Ranglisten zugunsten der Bildung von Ranggruppen kam es daraufhin zu einer Modifizierung etlicher Ranking-Verfahren – vor allem des in Deutschland sehr einflussreichen CHE-Hochschulrankings – und damit zu einer Annäherung der Instrumente Ranking und Rating.<sup>7</sup> Ein zentrales Merkmal von Ratings, das diese jedoch deutlich von Ranking-Verfahren unterscheidet, liegt in der Integration von Peer Reviews in den Verfahrensablauf, mit denen die Qualität der Bewertungen und Urteile sichergestellt werden soll. Damit stehen Ratings innerhalb der standardisierten Formen des Leistungsvergleichs in einer besonderen Nähe zu Evaluationen (*Lange 2010 S. 322*), bei denen grundsätzlich Peer Reviews implementiert sind.

---

<sup>7</sup>Diese inhaltliche Angleichung beider Instrumente zum Ausdruck bringend werden bei *Bogumil et al. 2013 S. 37 f.*, beide Verfahren zusammenfassend abgehandelt.

Der Einsatz von Ratings zur Unterstützung der Steuerung von Hochschulsystemen stellt eine Neuerung dar. Das Forschungsrating des Wissenschaftsrats wurde in den Jahren 2005 bis 2012 zur Bewertung der Fächer Chemie, Soziologie, Anglistik/Amerikanistik und Elektrotechnik/Informationstechnik eingesetzt. Adressaten dieses Ratings waren neben den „Entscheidungsträgern in den Universitäten [...] und in den außeruniversitären Forschungseinrichtungen“ auch die Ministerien (*Wissenschaftsrat 2004* S. 44). Unmittelbar aus dem Steuerungskontext resultierende Fragestellungen standen bei der Konzeption dieses Verfahrens jedoch nicht im Mittelpunkt. Da der Fokus ausschließlich auf der Forschung liegt, kann dieses Rating zudem keine umfassende Informationsbasis für Steuerungszwecke bereitstellen. Unzweifelhaft hat dieses als Pilotstudie angelegte Vorhaben das Instrument Rating in den Fokus der Wahrnehmung des Wissenschaftssystems und der Diskussionen über Leistungsmessung und Leistungsvergleiche gerückt. Ohne die Schwierigkeiten auszublenden, die auch mit diesem Rating-Verfahren einhergehen, kann mit Blick auf die methodische Qualität und Aussagekraft seiner Ergebnisse von einer vergleichsweise hohen Akzeptanz im deutschen Wissenschaftssystem gesprochen werden (*Hochschulrektorenkonferenz 2014*).

Im deutschen Hochschulsystem sind nur zwei Beispiele anzuführen, bei denen Ratings explizit im Kontext der Landeshochschulsteuerung konzipiert und durchgeführt wurden, nämlich das in den Jahren 2008/09 durchgeführte Sächsische Hochschulrating und das 2013/14 implementierte Fächerrating des Landes Hessen. Während das hessische Verfahren im folgenden Abschnitt ausführlich behandelt wird, soll an dieser Stelle das sächsische Rating überblicksartig dargestellt werden. Dies ermöglicht unter anderem – zunächst in Form einer Arbeitshypothese, die dann unter Einbezug der hessischen Befunde auszuarbeiten ist – eine Präzisierung derjenigen hochschulpolitischen Aufgaben und Zielsetzungen, die mit dem Einsatz von Ratings unterstützt werden sollen.

In den Jahren 2008/09 wurde im Freistaat Sachsen ein als Sächsisches Hochschulrating bezeichnetes Verfahren durchgeführt (die folgenden Ausführungen basieren auf der Darstellung von *Banscherus/Reiß (2012)*), bei dem es sich, anders als bei der Namensgebung zunächst zu vermuten, um eine Bewertung mit explizitem Fachbezug gehandelt hat. Die Ergebnisse und Einschätzungen des Ratings sollten Eingang in die Landeshochschulentwicklungsplanung des Landes finden. Dabei waren einzelne Strukturentscheidungen bereits vor der Durchführung des Ratings in der Diskussion. Konkret handelte es sich um eine stärkere Kooperation und Abstimmung benachbarter Hochschulen sowie um die Konzentration von Studienangeboten, womit unter anderem die Schließung von Außenstandorten und einzelnen Studiengängen verbunden war. Die Ergebnisse des Ratings sollten der Absicherung und Legitimationen dieser Vorschläge dienen. Diese mit dem Rating verbundene hochschulpolitische Zielsetzung in Sachsen – insbesondere die Frage, welche Strukturentscheidungen infolge des demografischen Umbruchs möglich und sinnvoll sind – hat den Fokus des

Verfahrens in besonderer Weise auf lehrbezogene Aspekte gelenkt. In einem Datenkatalog, basierend auf Daten der amtlichen Statistik und ergänzt um Daten, die von den Hochschulen eigens für das Rating erhoben wurden, wurden die Leistungen der Hochschulen dargestellt. Dieser Datenkatalog stellte die Bewertungsgrundlage für die sich anschließende Tätigkeit von Kommissionen mit einem fächergruppenbezogenen Zuschnitt dar. Insgesamt waren 17 Expertinnen und Experten für die Bewertung aller 21 in das Rating einbezogenen Fächer zuständig, wobei eine Anhörung Bestandteil des Bewertungsprozesses war. Die Bewertungen der Peers mündeten in einen Bericht, der um Korrekturen und Stellungnahmen der Hochschulen ergänzt werden konnte. War zunächst eine Verdichtung der Urteile angedacht (mit einer sechsstufigen Skala: „Exzellent“, „Sehr gut“, „Gut“, „Befriedigend“, „Nicht befriedigend“, „Nicht bewertbar“), wurde diese Ergebnisdarstellung nach Einwänden der Kommissionsmitglieder, die eine plakative Einschätzung der Leistungsfähigkeit nicht nach außen geben wollten, zugunsten einer rein textlichen Bewertung aufgegeben.

Im Anschluss an die Berichterstattung durch die Kommissionen diskutierte das Ministerium den Bewertungsbericht mit den Rektorinnen und Rektoren der sächsischen Hochschulen. Als Ergebnis des Ratings fanden die von den Expertinnen und Experten formulierten Empfehlungen zum Teil Eingang in den von der Staatsregierung vorgeschlagenen Hochschulentwicklungsplan (*Banscherus/Reiß 2012 S. 84*).

Das sächsische Verfahren hat den Hochschulen somit weitgehende Beteiligungsmöglichkeiten eingeräumt. In der Phase der Datenerhebung wie auch im anschließenden Bewertungsverfahren hatten die Hochschulen Gelegenheit, erklärende Stellungnahmen zu formulieren, Ergänzungen einzureichen und im Rahmen einer Anhörung an einer qualitativen Validierung der indikatorengestützten Bewertung mitzuwirken. Bei der Formulierung des Bewertungsberichts gab es für die Hochschulen ebenso Mitwirkungsmöglichkeiten wie bei der Diskussion der Ergebnisse und bei deren Überführung in die Landeshochschulentwicklungsplanung.

Am sächsischen Beispiel wird deutlich, welche Anforderungen und Zielsetzungen mit der Implementierung des Instruments Rating in den Kontext der Landeshochschulsteuerung zusammenhängen:

- Die mit dem NPM eingeführten Steuerungsgrundsätze – Output- statt Detailsteuerung – führen zu spezifischen, teilweise neuen Informationsbedarfen der Landesbehörden, welche nun eine belastbare Datenbasis für künftige Strukturentscheidungen und Entscheidungsprozesse, insbesondere eine fundierte Bewertung zur Leistungsfähigkeit der Standorte benötigen. Das Instrument Rating bietet sich für diese Zielstellung an, weil es einen Leistungsvergleich verschiedener Einheiten untereinander ermöglicht und mithilfe eines integrierten Peer Review-Verfahrens eine Beurteilung der Leistungen vor dem Hintergrund politischer Ziele vornehmen kann.

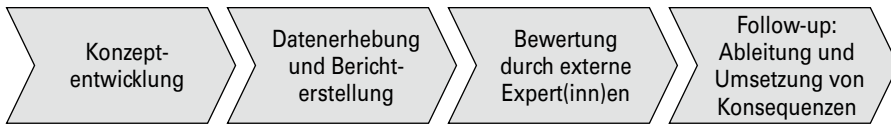
- Da eine Diskussion und Verwendung der Ergebnisse in größeren hochschulpolitischen Zusammenhängen vorgesehen ist, wird ein Verfahren benötigt, das überblicksartig informiert und bewertet. Die hohe Verdichtung der Ergebnisse bei Rating-Verfahren kommt dieser Anforderung prinzipiell entgegen.
- Bewertungen von Landeshochschulsystemen, wenn es sich zudem um Flächenländer mit einer vergleichsweise stark ausdifferenzierten Hochschullandschaft handelt, müssen außerdem unter Aufwands Gesichtspunkten erfolgen. Deshalb sind Ratings als weniger detailtiefe Verfahren gegenüber aufwändigen Evaluationen oder auch den kommunikationsintensiven Benchmarking-Verfahren im Vorteil. Allerdings sind Ratings weniger geeignet, wenn es darum geht, Erklärungen für die festzustellenden Unterschiede bereitzustellen, gerade weil sie eine geringere Detailtiefe aufweisen.

### 3 Die Ausgestaltung des Fächerratings in Hessen

Die stärksten Impulse für die Implementierung eines Fächerratings im Land Hessen gingen von veränderten Informationsbedarfen des zuständigen Ministeriums (HMWK) aus. Weil die im Zuge des NPM ausgeweitete Autonomie der Hochschulen zunehmend weniger Einblicke des Ministeriums in das Geschehen in den Hochschulen zuließ, gewannen aus Sicht der Behörde Informationen über die erbrachten Leistungen an Bedeutung. Dies stand in unmittelbarem Zusammenhang mit einer sich in zunehmendem Maße als Output-orientiert verstehenden Landeshochschulsteuerung, auch charakterisiert als Steuerung „aus der Distanz“ (Hüther/Krücken 2016 S. 60; Blümel et al. 2011 S. 105), und führte zu geänderten Informationsbedarfen, die mithilfe des bestehenden Landesberichtswesens nicht abgedeckt werden konnten. Es galt eine belastbare Datenbasis zu schaffen, die eine fundierte Bewertung der Leistungsfähigkeit der Standorte ermöglicht und damit künftige Strukturentscheidungen unterstützen könnte. Dabei war es eine schon zu Beginn und in Abstimmung mit den Hochschulen des Landes getroffene Entscheidung, ein „schlankes Verfahren“<sup>8</sup> zu implementieren, um den Aufwand sowohl für die Hochschulen als auch für das durchführende Ministerium zu begrenzen.

Mit der Konzeptionierung und Implementierung eines Ratings wurde mit dem DZHW ein externer Partner beauftragt, der über fachliche Expertise und Erfahrung auf diesem Gebiet verfügte. Im Rahmen einer 2013 beginnenden Konzeptionierungsphase wurde gemeinsam mit den Hochschulen des Landes ein für geeignet befundenes Verfahren vereinbart und in den Jahren 2013/14 implementiert. Der vom Ministerium beschlossene Rating-Ablauf lässt sich wie im Folgenden dargestellt als Vier-Phasen-Prozess beschreiben:

<sup>8</sup>Mündliche Aussage von Vertretern des HMWK.

**Abbildung 1:** Phasengliederung des hessischen Fächerratings

Der erste Rating-Durchlauf in den Jahren 2013/14 war als Pilotphase angelegt, in dessen Verlauf bzw. Nachgang an mehreren Stellen eine Feinjustierung des Verfahrens möglich sein sollte. Diese Pilotphase hatte die Fächer Mathematik, Physik, Chemie und Elektrotechnik zum Gegenstand und ist deshalb auch als „MINT-Fächerrating“ bezeichnet worden. Im zweiten Durchlauf in den Jahren 2015/16 fand eine Bewertung der Fächer Rechtswissenschaft, Wirtschaftswissenschaften und Geschichtswissenschaft statt, wobei zusätzlich zu allgemeinen Verfahrensmodifikationen einige (wenige) fachspezifische Anpassungen des Instruments vorgenommen wurden. In den fachbezogenen Vergleich wurden alle staatlichen Universitäten und Hochschulen für angewandte Wissenschaften des Landes einbezogen.

### 3.1 Leistungsmessung

Am Beginn eines Ratings, wie bei allen anderen Bewertungsverfahren, sollte die Formulierung möglichst konkreter, mit dem Rating verbundener Ziele stehen, deren Operationalisierung, z. B. durch Kennzahlen, dann im weiteren Verlauf der Konzeptionierung erfolgen kann (Lange 2010 S. 325). Dieser erste in der Phase der Konzeptentwicklung zu leistende Arbeitsschritt ist von erheblicher Relevanz für alle nachfolgenden Phasen des Verfahrens. Insbesondere kann durch die Unterlegung mit konkreten steuerungspolitischen Zielen Vorbehalten bzw. Bedenken hinsichtlich der späteren Verwendung der Ergebnisse entgegengetreten werden. Vorbehalte aufseiten der Peers können z. B. mit der Verdichtung und Plakativität der Ergebnisdarstellung in Rating-Verfahren zusammenhängen, sich aber auch auf die aufseiten der scientific community stets gefürchteten Debatten um Kürzungen und Standortschließungen beziehen. Die Definition der Informationsbedarfe und die Festlegung der Verfahren der Leistungsmessung sollten dann der Formulierung steuerungspolitischer Ziele folgen.

In der Praxis des hessischen Ratings waren konkrete Zielsetzungen zunächst nicht formuliert. Zwar hatte das HMWK allgemeine Ziele schon vor Beginn der Pilotphase in Form von Leitfragen beschrieben, die an die für die Bewertung zuständigen Expertinnen und Experten gerichtet waren. Diese Leitfragen waren jedoch keinen klar konturierten Steuerungszielen zugeordnet und nicht mit spezifischen Datenanforderungen gekoppelt. Aus Sicht mehrerer Mitglieder der Bewertungskommissionen wurden deshalb in der Pilotphase die Ziele des HMWK als nicht hinreichend deutlich

bzw. zu wenig konkretisiert bezeichnet. Dies hat das HMWK veranlasst, vor Durchführung des zweiten Turnus eine weitere Klärung und Konkretisierung der Ziele vorzunehmen.

Das von den Hochschulen zu beliefernde Datenset wurde in einem kommunikativen und auf Konsens ausgerichteten Prozess zwischen dem HMWK und den hessischen Hochschulen vereinbart. Die Vorgehensweise war von pragmatischen Überlegungen zur Datenverfügbarkeit und vom Verbreitungsgrad einzelner Indikatoren geprägt, nicht jedoch von theoretischen Überlegungen. Grundsätzlich ist hierzu anzumerken, dass das Wissenschaftssystem derzeit keine akzeptierten und belastbaren Verfahren der Auswahl von Indikatoren und Kennzahlen für Zwecke der Leistungsmessung anbieten kann.<sup>9</sup> Es fehlen insbesondere theoretische Begründungen für die Auswahl und Gewichtung von Indikatoren (*Lange 2010 S. 326*). In der Folge haben sich im hessischen Rating einzelne Indikatoren bzw. Messverfahren im Verlauf des Verfahrens nicht bewährt bzw. als weniger aussagekräftig als erwartet herausgestellt. Dies betraf z. B. den Indikator „gemeldete Patente“, der im Anschluss an die Pilotphase aus dem Datenset herausgenommen wurde.

Vereinbart wurde ein Set von Indikatoren, das quantitative und qualitative Information zu den Bereichen Studium und Lehre, Forschung und Transfer, Nachwuchsförderung, Chancengleichheit und Internationalisierung sowie zu den Rahmenbedingungen des jeweiligen Faches bereitstellte und das in Form von Datenberichten (jeweils separat für einzelne Hochschulen und Fächer) ausgewiesen wurde. Für die Erstellung dieser Datenberichte wurde primär auf Quellen der amtlichen Statistik zurückgegriffen. Die Abgrenzung der Fächer für die Zwecke des Ratings erfolgte über die Zuordnung der Studiengänge zur Systematik der amtlichen Statistik. Die Festlegung zu Besonderheiten und Problemlagen einzelner Fächer (z. B. Zuordnung interdisziplinärer Studiengänge, Ausstattungsmerkmale von Fächern) erfolgte unter Einbezug der betroffenen Hochschulen. Ergänzend zu diesen quantitativen Daten enthielten die Berichte textliche Erläuterungen zu einzelnen Sachverhalten (z. B. zu den Schwerpunkten des Fachs in Forschung und Lehre, zu Kooperationsbezügen, zu Maßnahmen zur Betreuung der Studierenden, zur Gewinnung und Betreuung des Nachwuchses usw.). Diese Informationen wurden von den Hochschulen einheitlich in Texten von vorgegebener Länge zusammengestellt. Bevor es zu einer Auswertung der Berichte durch die Kommissionen kam, wurden sie an die Hochschulen zurückgekoppelt. Diese erhielten damit die Gelegenheit, Fehler oder Leerstellen in den Berichten zu identifizieren und zu korrigieren.

<sup>9</sup>Mittlerweile sind Fragen der Leistungsmessung stärker in den Fokus der Wissenschaftsforschung gerückt, sodass für künftige Ratings möglicherweise auf derartige Verfahren oder Modelle zurückgegriffen werden könnte. Siehe hierzu insbesondere die im Rahmen der Förderlinie des BMBF „Leistungsbewertung in der Wissenschaft“ durchgeführten Forschungsvorhaben (<http://www.hochschulforschung-bmbf.de/de/1333.php>).

Die Auswertung der Datenberichte war Aufgabe der fachspezifischen Kommissionen und verlief nicht ohne Schwierigkeiten: In den Berichten waren die Indikatoren fach- und hochschulbezogen zusammengestellt, sie enthielten jedoch – in der Pilotphase – keine vergleichenden Datenzusammenstellungen. Die fehlende Vergleichsperspektive, die im Rahmen der Kommissionssitzungen erst herzustellen war, wurde von den Peers kritisiert. Daneben lag für die Expertinnen und Experten eine weitere Schwierigkeit darin, die in textlicher Form vorliegenden und stark unterschiedlich akzentuierten Informationen zu qualitativen Sachverhalten (z. B. Studienqualität, Zugang zum Arbeitsmarkt) vergleichend zu bewerten. Außerdem wurde eine unzureichende Passung zwischen den Berichten einerseits und den Leitfragen andererseits bemängelt. Nachdem in der Pilotphase diese Schwierigkeiten offenkundig geworden waren, wurde das Datenkonzept für den zweiten Rating-Durchlauf in den genannten Punkten angepasst. Aber auch im Rahmen des zweiten Rating-Durchlaufs (2015/16) zogen die Bewertungskommissionen ein Fazit, welches kritische Aspekte hinsichtlich der Abbildung der von den Hochschulen erbrachten Leistungen in den Datenberichten beinhaltete. Eine besonders kritische Bewertung erfuhr in beiden Projekt-Durchläufen die Abbildung der Publikationstätigkeit der Fächer, der infolge des Verzichts auf eine bibliometrische Aufarbeitung wenig Aussagekraft zugebilligt wurde. Welche Form einer bibliometrischen Aufbereitung im Ratingkontext jedoch zielführend und angemessen gewesen wäre, konnte nicht herausgearbeitet werden. Insgesamt benannten die Kommissionen beider Durchläufe Schwierigkeiten, die vorliegende Datenbasis zu einem Gesamtbild zusammenzufügen, und regten eine noch stärker fachspezifische Ausrichtung der Informationsgrundlage an.

### **3.2 Leistungsbewertung**

Eine Bewertung der Leistungsfähigkeit der Hochschulstandorte, ausgehend von den oben beschriebenen Datenberichten, erfolgte durch ein Peer Review-Verfahren. Derartige Verfahren der Leistungsbewertung werden als „Informed Peer Review“ bezeichnet und sind mittlerweile im Wissenschaftssystem etabliert. Dieses Verfahren hat den Vorteil, dass mit der Heranziehung von quantitativen Informationen bzw. Kontextinformationen die Risiken einer subjektiven Urteilsfindung durch die Peers abgemindert werden (*Wissenschaftsrat 2011 S. 17*). *Matthies, Simon und Knie (2008 S. 334)* vertreten allerdings die Einschätzung, dass es sich hierbei nur um eine „vermeintliche Objektivierung“ handle, da auch quantitative Informationen auf vorangegangenen Peer Review-Verfahren basieren können. Indem mit der Bewertung durch Expertinnen und Experten ein Kernelement der Selbststeuerung von Wissenschaft im Rating-Kontext zum Einsatz gebracht wird, findet in jedem Fall eine Aufwertung des gesamten Verfahrens mit Blick auf Wissenschaftsadäquatheit und Akzeptanz statt.

Die für die Auswertung der Datenberichte und die Bewertung der Hochschulstandorte zuständigen Kommissionen bestanden aus fünf bis zehn Personen, waren fachbezogen zusammengesetzt und ausschließlich für die Bewertung des jeweiligen Fachs zuständig. Die Entscheidung über die Zusammensetzung der Kommissionen erfolgte in Abstimmung mit den Hochschulen, lag aber letztlich beim HMWK. Die Sitzungen der Kommissionen fanden in Wiesbaden statt und hatten eine Dauer von einem (Pilotphase) bzw. eineinhalb Tagen (zweiter Durchlauf). Die längere Dauer der Kommissionssitzung im zweiten Rating-Durchlauf resultierte aus der Einführung einer Anhörung von Vertreterinnen und Vertretern der Hochschulen. Ein Ergebnis der Auswertung der Pilotphase war, dass den Kommissionen eine Möglichkeit gegeben werden sollte, einzelne Sachverhalte noch im Verlauf der Sitzung durch Rückfragen an die Hochschulen einer Klärung zuzuführen.

Im Verlauf der Kommissionssitzungen wurden – nach Möglichkeit einstimmig – Urteile zur Leistungsfähigkeit der Hochschulstandorte getroffen, die im Anschluss an die Sitzungen in einem Kommissionsbericht ausformuliert wurden. Die Hochschulen bekamen im Anschluss die Möglichkeit, zu dem konsentierten Bericht sachliche Richtigstellungen und inhaltliche Stellungnahmen zu formulieren. In der Pilotphase erfolgte dies im Nachgang zur Berichterstellungsphase. Im zweiten Rating-Durchgang erhielten die Hochschulen die Bewertungsberichte dann bereits in einer Entwurfsfassung, was es den Kommissionsmitgliedern ermöglichte, die sachlichen Hinweise in die finale Fassung der Berichte aufzunehmen. Auf diese Weise waren die Stellungnahmen der Hochschulen als Moment der Qualitätssicherung in die Berichtserstellungsphase integriert. Die Kommissionsberichte enthielten ausschließlich textliche Bewertungen, auf die Bildung von Ranglisten oder -gruppen wurde hingegen ausdrücklich verzichtet.

Durchgängig lässt sich eine eher zurückhaltende, vorsichtige Bewertungspraxis durch die Kommissionen feststellen. Hierfür können einerseits verfahrensspezifische Gründe angeführt werden: Dazu zählen die oben bereits angeführten Schwierigkeiten der Kommissionsmitglieder, die vorliegenden Informationen zu einem Gesamtbild zusammenzufügen. Auch die Unschärfe der Ziele des Ratings könnte eine derartige Bewertungspraxis befördert haben. Andererseits sind strukturelle Gründe zu vermuten, die mit dem Instrument Peer Review zusammenhängen. Die Verwendung im Rating-Kontext geht mit einer Verschiebung der Funktionsbestimmung von Peer Review-Verfahren einher. In traditionellen Einsatzfeldern von Peer Review – bei der Bewertung von Publikationen, im Kontext von Berufungsverfahren, aber auch bei Verfahren wie Evaluationen – sind Selektions- und Konstruktionsfunktionen (*Neidhardt 2010 S. 282*) von zentraler Bedeutung: Artikel werden ausgewählt, professorales Personal wird gekürt, die Empfehlungen der Peers gehen in die Restrukturierung wissenschaftlicher Organisationen ein. Im Rating-Kontext haben die Peers hingegen vornehmlich die Aufgabe, zu informieren und vergleichend zu bewerten. Anders als bei der Bewertung von Publika-



tionen oder bei der Programm- und Forschungsförderung stehen die Peers vor der Aufgabe, komplexe Organisationen bzw. das jeweilige Landeshochschulsystem zu bewerten. Dies sind Anforderungen, die über eine fachliche Expertise der Kommissionsmitglieder deutlich hinausgehen. Zwar wurden in Hessen bei der Auswahl der Kommissionsmitglieder Erfahrungen in der akademischen Selbstverwaltung, in der Hochschulleitung und -entwicklung sowie die Mitarbeit in Gremien, die von Relevanz für das Wissenschaftssystem insgesamt sind, (Wissenschaftsrat, DFG, Akkreditierungsrat) berücksichtigt, im Ergebnis des Auswahlprozesses waren die Kommissionen jedoch keine ausschließlich aus Organisationsanalysten bestehenden Gremien. Es ist zu konstatieren, dass die Anforderungen an die Zusammenstellung von Bewertungskommissionen im Rating-Kontext vielfältig und vergleichsweise hoch sind. Vor dem Hintergrund der im Wissenschaftssystem wahrgenommenen Gefahr einer „Überlastung des Gutachterwesens“ (Neidhardt 2010 S. 286), könnten diese besonderen Anforderungen eine Restriktion hinsichtlich der Ausweitung des Einsatzes von Ratings darstellen.

Der Frage, inwieweit mit einem Wechsel des Begutachtungsgegenstands – von Individuen, Publikationen oder Forschungsvorhaben hin zur Gesamtleistung einer wissenschaftlichen Einrichtung – ein Wandel der Bewertungspraktiken einhergeht, hat sich *Marc Torka (2011)* zugewandt. Er kommt zu der Feststellung, dass trotz eines organisationsspezifischen Fokus einzelne, wissenschaftsspezifische Aspekte (z. B. Forschungsprogramme und -ergebnisse) in den Mittelpunkt der Bewertung gestellt werden (*Torka 2011 S. 80*). *Reichert (2013 S. 37)* stellt mit Rückgriff auf die Untersuchung Torkas die Vermutung auf, dass gerade „Informed Peer Review“-Verfahren dahin tendieren, überwiegend Überblicks-Outputdaten zu einer – möglicherweise verzerrenden – Interpretation zu verwenden. Eine derartige Bewertungspraxis lässt sich zwar in Hessen nicht feststellen. Allerdings bestätigen die bereits erwähnte, eher zurückhaltende Bewertungspraxis im Fächerrating sowie die aufseiten der Peers festzustellenden Schwierigkeiten, die Datenbasis zu einem geschlossenen Gesamtbild zusammenzufügen, die Vermutungen Torkas.

Peer Review-Verfahren ist die Gefahr von subjektiven Urteilen z. B. infolge von Befangenheit (Bias) inhärent. Verantwortlich hierfür ist die – allerdings notwendige – große fachliche Nähe der Peers zu dem zu bewertenden Gegenstand, die sich auf die Bewertung entweder in überkritischen (bspw. aufgrund von Konkurrenzsituationen) oder in verharmlosenden Urteilen (kollegialer Bias) auswirken kann (*Neidhardt 2006 S. 420*). Bei Verfahren wie in Hessen, die vonseiten der zuständigen Landesbehörden initiiert werden, ist insbesondere mit einem kollegialen Bias der Gutachterinnen und Gutachter zu rechnen. Um hinsichtlich der Bewertungspraxis in Hessen zu wissenschaftlich belastbaren Aussagen zu kommen, wären detaillierte Untersuchungen erforderlich (Beispiele hierfür bei *Fischer 2011*), die in dem hier besprochenen Fall nicht vorliegen. Es kann deshalb nur eine – ebenfalls subjektive – Einschätzung des Verfas-

sers wiedergegeben werden. Demnach haben Kollegialitätserwägungen die Kommissionsberatungen durchaus unterschwellig begleitet und zu einer grundsätzlichen Zurückhaltung bei Bewertungen und zu schwachen Formulierungen negativ konnotierter Urteile geführt. Als Missachtung der Intentionen des Auftraggebers ist dieses keinesfalls zu werten. Vielmehr kommt darin das strukturelle Problem des Instruments Peer Review zum Ausdruck, die erforderliche „kompetente Unbefangenheit“ (Neithardt 2006 S. 420) der Evaluatoren herzustellen.

### 3.3 Ergebnisverwendung

Die Zielsetzung des hessischen Ratings bestand im Wesentlichen darin, eine verlässliche und für Vergleichszwecke geeignete Informationsgrundlage zu generieren. Die Ergebnisse sollten jedoch erst im Nachgang des eigentlichen Ratings in die Entwicklungsplanungen der Hochschulen bzw. des Ministeriums überführt werden. Dem Rating kam somit die Funktion zu, die Verfahren, in deren Verlauf Entscheidungen zur Struktur des Landeshochschulsystems getroffen werden, zu rationalisieren und transparent zu gestalten, wodurch die Akzeptanz für Strukturentscheidungen erhöht werden sollte.

Die von diesem Verfahren konkret ausgehende Steuerungswirkung lässt sich an dieser Stelle nicht verlässlich abschätzen. Dies hat mehrere Gründe. Zum einen war bei Abfassung dieses Beitrags der zweite Durchlauf des Ratings noch nicht abgeschlossen. Zum anderen wären für eine derartige Bewertung vertiefte Kenntnisse innerministerieller Prozesse und Einschätzungen erforderlich. Eine diese Fragestellungen explizit aufgreifende Begleitforschung, die die von den verantwortlichen ministeriellen Stellen auf Grundlage der Rating-Ergebnisse initiierten Maßnahmen auf ihre Wirksamkeit hin untersucht, könnte hierzu Erkenntnisse beisteuern. Vorerst kann nur auf die folgenden Befunde hingewiesen werden: Die Ergebnisse des ersten Rating-Durchlaufs (Pilotphase) wurden in mehreren Arbeitsgruppen zur Landeshochschulentwicklungsplanung von Vertreterinnen und Vertretern des Ministeriums und der Hochschulen diskutiert und zum Teil in konkrete Maßnahmen überführt, die im Rahmen von Zielvereinbarungen festgeschrieben wurden. Daneben ist auf eine weitere Möglichkeit der Ergebnisverwendung hinzuweisen, die sich zurzeit noch in der Konzeptionierungsphase befindet: So ist aufseiten des HMWK beabsichtigt, die im Zuge der Datenerhebung für das Rating gemachten Erfahrungen in die Weiterentwicklung des Landesberichtssystems einfließen zu lassen.

### 3.4 Partizipation

Es ist ein Charakteristikum der sich im Wandel befindlichen Governance-Strukturen im Hochschul- und Wissenschaftsbereich, dass hierarchische Steuerungsmechanismen durch nicht-hierarchische, auf Kooperation beruhenden Formen der Produktion

öffentlicher Güter abgelöst werden (*Grande et al. 2014 S. 20*). Die Einführung von Steuerungsinstrumenten wird deshalb nicht „top down“ von den zuständigen staatlichen Stellen beschlossen, sondern „bottom up“ unter Einbezug der beteiligten Hochschulen konzipiert und durchgeführt. In Hessen ist dabei die Betonung des partizipativen Gedankens auch in den entsprechenden Gesetzestexten verankert (§ 7 Abs. 1 HHG HE 2009). Derartige Akteurskonstellationen sind für Zielvereinbarungen und für Verfahren der leistungsbezogenen Mittelvergabe bereits ausführlich dargestellt worden (*Jaeger 2009 S. 57*). Das Beispiel Hessen zeigt, dass die Einführung des stärker wettbewerblichen Instruments Rating sich ebenfalls an diesem Muster orientiert. Bei der Implementierung und Durchführung des Ratings war aus Sicht der Akteure im HMWK die mittels Partizipation herzustellende Akzeptanz des Verfahrens eine zentrale Voraussetzung für den erfolgreichen Instrumenteneinsatz.

Demzufolge waren im Ablauf des hessischen Ratings zahlreiche Möglichkeiten zur Mitwirkung und zur Einflussnahme durch die Hochschulen des Landes in allen Projektphasen angelegt: Mit ihrer Teilnahme an den vorbereitenden Sitzungen, ihrer Mitwirkung bei dem zu vereinbarenden Datenset sowie bei den Festlegungen zur Ergebnisdarstellung waren die Hochschulen intensiv an der Konzeptentwicklung beteiligt. In der Phase der Berichterstattung wurden die Datenberichte von den beteiligten Hochschulen geprüft und gegebenenfalls geändert, bevor sie den Kommissionen für das Bewertungsverfahren zur Verfügung gestellt wurden. In der Bewertungsphase hatten die Hochschulen des Landes im ersten Durchlauf (Pilotphase) nur wenige Möglichkeiten zur Mitwirkung. Es wurde in der Rückschau auf die Pilotphase deutlich, dass partizipative Elemente, die auch der Qualitätssicherung dienen sollten, noch gestärkt werden sollten. Deshalb wurden auf Wunsch der Peers im zweiten Durchlauf des Ratings die Kommissionssitzungen um eine Anhörung der Hochschulen ergänzt. Schließlich waren die Hochschulen auch in der vierten Phase des Ratings, der Ableitung und Umsetzung von Konsequenzen, beteiligt, indem die Ratingergebnisse in den Arbeitsgruppen zur Landeshochschulentwicklungsplanung diskutiert und als Grundlage für die mit dem Land getroffenen Zielvereinbarungen genutzt wurden.

#### **4 Fazit**

Die Untersuchung des hessischen Fächerratings erfolgte mit der Zielstellung, die Eignung und Leistungsfähigkeit von Ratings für Zwecke der Landeshochschulsteuerung zu bewerten. In diesem Zusammenhang sollte auch untersucht werden, ob sich im Kontext von Ratings möglicherweise eine Änderung der Funktionsbestimmung von Peer Reviews feststellen lässt.

Bevor die Befunde zum hessischen Fächerrating vor dem Hintergrund dieser Fragestellungen diskutiert werden, ist zu klären, ob es sich bei dem hessischen Verfahren

überhaupt um ein Rating im eigentlichen Sinne handelt. Ein Charakteristikum von Ratings liegt in der Verdichtung der Ergebnisse zu Ranggruppen auf der Grundlage einer Benotungsskala. Das hessische Rating hat eine solche Form der Ergebnisverdichtung nicht vorgenommen, wobei es bemerkenswert ist, dass dies auch zu keinem Zeitpunkt der Konzeption und der Durchführung des Verfahrens ernsthaft erwogen wurde. Eine Verdichtung der Bewertungen zu Ranggruppen oder die Abbildung auf einer Notenskala hätte die Akzeptanz des Verfahrens aufseiten der Hochschulen vermutlich deutlich gemindert. Ein vergleichbares Vorgehen kennzeichnet im Übrigen auch das Verfahren in Sachsen. Möglicherweise ist dies ein Hinweis auf strukturelle Schwierigkeiten, bei Ratings, die im Steuerungskontext zu verorten sind, Leistungsunterschiede plakativ zur Darstellung zu bringen. Dahingegen ist es bei einem Verfahren mit einer anders gelagerten Zielrichtung wie dem Forschungsrating des Wissenschaftsrats durchaus möglich, zu verdichtenden und plakativen Urteilen zu gelangen. Zum einen fehlt also ein zentrales Merkmal von Ratings, zum anderen ist insbesondere mit Blick auf den Prozessablauf eine große strukturelle Ähnlichkeit des hessischen Ratings mit evaluativen Verfahren nicht zu übersehen. Allerdings ist in Hessen ein Verfahren entwickelt worden, das sich gegenüber Evaluationen durch einen deutlich reduzierten Ressourceneinsatz auszeichnet. Trotz strukturell großer Nähe zu evaluativen Verfahren handelt es sich somit um ein spezifisch formiertes Verfahren, das sowohl Charakteristika von Ratings als auch von Evaluationen aufweist.

Es ist somit festzustellen, dass in Hessen ein Rating-Modell spezifischen Zuschnitts zum Einsatz kam, für dessen Verwendung im Steuerungskontext besondere Problemlagen und Anforderungen zu konstatieren sind. Die Durchführung eines Ratings, das als ressourcenschonende Alternative zu Evaluationen konzipiert wurde, jedoch trotzdem aussagekräftige Bewertungen eines komplexen Landeshochschulsystems zulässt, hat in der Praxis mehrere Schwachstellen erkennen lassen: So blieben einzelne Felder der Leistungsmessung unterkomplex und damit einer angemessenen Bewertung durch die Kommissionen nicht zugänglich; dies trifft beispielsweise für Publikationsdaten zu, aber auch für andere, in ihrer Qualitätsdimension zu bewertende Sachverhalte (Studienqualität, Zugang zum Arbeitsmarkt). Die Tätigkeit der Bewertungskommission ist hier als weiterer Punkt anzuführen. Es ist offenkundig, dass mit einer zweistündigen gemeinsamen Anhörung von Vertretern und Vertreterinnen mehrerer Hochschulen nicht die gleiche Detailtiefe der Bewertung zu erzielen ist wie bei den im Rahmen von Evaluationen üblichen ganz- oder mehrtägigen Begehungen einzelner Fachbereiche. Deutlich geworden ist am hessischen Beispiel auch, dass sich Bewertungskommissionen im Rating-Kontext hohen Anforderungen gegenübersehen. Ihre Aufgaben sind durchaus vergleichbar mit denen, die an Evaluierungskommissionen gestellt werden. Anders als beim Peer Review wissenschaftlicher Publikation oder bei der Begutachtung von Forschungsanträgen, bei denen ausschließlich fachliche Aspekte zu bewerten sind, geht es im Rating-Kontext um die Bewertung eines komplexen

Landeshochschulsystems. Im Vergleich zu Evaluationen hatten die Kommissionen im hessischen Fächerrating diese Aufgaben jedoch mit einem deutlich reduzierten Aufwand zu bewältigen. Ob die in Hessen in ausschließlich textlicher Form getroffenen Urteile der Kommissionen das Informationsbedürfnis des beauftragenden HMWK letztlich bedienen konnten, ist aus Sicht des Verfassers nicht abschließend zu beurteilen. Die durchgängig vorsichtige, eher zurückhaltende Bewertungspraxis aller Kommissionen im hessischen Fächerrating gibt jedoch Anlass zu einer eher skeptischen Einschätzung. Das sächsische Beispiel als einziger zur Verfügung stehender Vergleichsmaßstab ist in diesem Punkt deutlich anders gelagert, da die Ziele des Ratings – Legitimierung und empirische Absicherung von bereits im Vorfeld diskutierten Strukturentscheidungen – hier klarer abgegrenzt und anders konturiert waren.

Ein abschließendes Urteil zur Leistungsfähigkeit von Ratings im Steuerungskontext kann auf Grundlage dieser Befunde nicht getroffen werden – zum einen, weil aus lediglich zwei Fallbeispielen keine allgemeine Gültigkeit beanspruchenden Schlussfolgerungen zu ziehen sind, zum anderen, weil sich einige Details des Verfahrensablaufs der Kenntnis des Verfassers entziehen. Dies betrifft insbesondere die Art der Ergebnisverwendung sowie die Einschätzungen der Kommissionsmitglieder zu verschiedenen Aspekten ihrer Tätigkeit im hessischen Fächerrating. Um bezüglich dieser Punkte zu belastbaren Ergebnissen zu kommen, bedürfte es einer Begleitforschung, die sich explizit den damit verbundenen Fragestellungen zuwendet. Da eine solche Begleitforschung derzeit weder vonseiten des HMWK noch vom DZHW angedacht ist, soll und kann an dieser Stelle deshalb nur eine vorläufige Einschätzung wiedergeben werden, die auf den hessischen Erfahrungen basiert. Demnach hat sich die Leistungsfähigkeit des Instruments Rating in dem dargestellten Verfahrenszuschnitt noch nicht erwiesen. Als problematisch haben sich nach Meinung des Verfassers insbesondere die Bereiche Leistungsmessung und -bewertung herausgestellt, bei letzterem speziell die Kopplung der Ergebnisdarstellung mit den Informationsbedarfen des Auftraggebers.

Aus der Darstellung und Diskussion des Fächerratings in Hessen lassen sich Aspekte ableiten, die nach Einschätzung des Verfassers Erfolgsfaktoren für den Einsatz eines derartigen Instruments sein könnten und die hier in Form von Empfehlungen zusammengefasst werden. Sollten zukünftig weitere Länder Ratings in ihrer Steuerungspraxis einsetzen, wodurch sich neue Möglichkeiten ergäben, die Leistungsfähigkeit des Instruments abzuschätzen, könnten diese Empfehlungen die Konzeptionierung einzelner Verfahren begleiten und unterstützen. Die Empfehlungen lauten wie folgt:

- Eine Klärung der Ziele und der Aufgabenstellung des Ratings sollte zu Beginn des Verfahrens erfolgen. Die Ziele sollten operationalisierbar und messbar gewählt sein.
- Gekoppelt mit den Zielen sollte sich die Leistungsmessung auf ausgewählte Sachverhalte konzentrieren. Dies ermöglicht auch die Implementierung anspruchsvoller

und qualitätsgesicherter Verfahren der Datenaufbereitung (z. B. im Bereich Bibliometrie), soweit dies für die Zielstellung des Ratings erforderlich sein sollte. Abzura-ten ist hingegen von dem Versuch, die gesamte Komplexität des Hochschulsystems mit eng begrenzten Ressourcen abbilden zu wollen.

- Fachspezifische Besonderheiten sind bei den Verfahren der Datenerhebung und Leistungsmessung unbedingt zu berücksichtigen. Unter Umständen sollte bereits hier die Tätigkeit der Bewertungskommissionen ansetzen.
- Sofern Ratings nicht nur Leistungen messen, sondern diese auch unter Einbezug qualitativer Aspekte bewerten sollen, ist eine Alternative zu Peer Review-Verfahren derzeit nicht in Sicht. Eine zentrale Aufgabe ist darin zu sehen, den Kommissionen gute und zielführende Arbeitsbedingungen zu schaffen. Hierzu gehören auch ausreichende zeitliche Ressourcen, um eine vorwiegend quantitative Datengrundlage mit den Ergebnissen eines Austauschs mit Vertretern und Vertreterinnen der Hochschulen für die Bewertung zusammenzuführen. Von großer Bedeutung ist auch die adäquate Zusammensetzung der Bewertungskommissionen mit Personen, die neben ihrer fachlichen Kompetenz über Kenntnisse organisationalen Handelns und zu Länderspezifika verfügen und die mit den Möglichkeiten und Restriktionen staatlichen Gestaltens im Hochschulbereich vertraut sind.
- Für die Gewinnung von kompetenten Peers sollten in ausreichendem Maße zeitliche Ressourcen vorgesehen werden. Dies gilt auch mit Blick auf die in den letzten Jahren feststellbare Ausweitung des Gutachterwesens.
- Für die Ergebnisdarstellung sollten die Möglichkeiten geprüft werden, zu stark verdichtenden Bewertungen (z. B. Ranggruppen) zu kommen und diese ggf. in Abstimmung mit den zu bewertenden Einrichtungen zu vereinbaren.
- Die partizipative Ausgestaltung von Ratings im Steuerungskontext ist (allerdings ausgehend von den durch die Landesbehörde festzulegenden Zielen) in allen Verfahrensschritten mitzudenken.

## Literatur

*Banscherus, Ulf; Reiß, Miriam (2012):* Das sächsische Hochschulrating als Beispiel für einen evidenzbasierten Ansatz in der Hochschulentwicklungsplanung. Ein Werkstattbericht. Beiträge zur Hochschulforschung 34 (3), S. 72–91

*Berghoff, Sonja; Federkeil, Gero; Giebisch, Petra; Hachmeister, Cort-Denis; Müller-Böling, Detlef (2001):* Der Studienführer 2001: Das Hochschulranking. Vorgehensweise und Indikatoren. CHE-Arbeitspapier 24. Gütersloh

*Blümel, Albrecht; Kloke, Katharina; Krücken, Georg (2011):* Professionalisierungsprozesse im Hochschulmanagement in Deutschland. In: Langer, Andreas; Schröder,

Andreas (Hrsg.): Professionalisierung im Nonprofit Management (Soziale Investitionen). Wiesbaden, S. 105–127

*Bogumil, Jörg; Burgi, Martin; Heinze, Rolf G.; Gerber, Sascha; Gräf, Ilse-Dore; Jochheim, Linda; Schickentanz, Maren (2014): Zwischen Selbstverwaltungs- und Managementmodell. Umsetzungsstand und Bewertungen der neuen Steuerungsinstrumente in deutschen Universitäten. In: Grande, Edgar et al. (Hrsg.): Neue Governance der Wissenschaft. Reorganisation – externe Anforderungen – Medialisierung. Bielefeld, S. 49–72*

*Bogumil, Jörg; Burgi, Martin; Heinze, Rolf. G.; Gerber, Sascha; Gräf, Ilse-Dore; Jochheim, Linda; Schickentanz, Maren; Wannöffel, Manfred (2013): Modernisierung der Universitäten. Umsetzungsstand und Wirkungen neuer Steuerungsinstrumente (Modernisierung des öffentlichen Sektors: Sonderband, Bd. 41). Berlin*

*Borgwardt, Angela (2011): Kriterien, Aussagekraft und Wirkungen von Rankings. In: Dies. (Hrsg.): Rankings im Wissenschaftssystem – Zwischen Wunsch und Wirklichkeit. Publikation zur Konferenz der Friedrich-Ebert-Stiftung am 17. März 2011 (Schriftenreihe des Netzwerk Exzellenz an deutschen Hochschulen, Bd. 4). Bonn, S. 23–42.*

*Burtscheidt, Christine (2010): Humboldts falsche Erben. Eine Bilanz der deutschen Hochschulreform. Frankfurt am Main*

*Dohmen, Dieter (2015): Anreize und Steuerung in Hochschulen – Welche Rolle spielt die leistungsbezogene Mittelzuweisung? (FiBS-Forum 54). Berlin*

*Expertenkommission Hochschulentwicklung in Schleswig-Holstein unter dem Vorsitz von Prof. Dr. Hans-Uwe Erichsen (Hrsg.) (2003): Zur Entwicklung der Hochschulen in Schleswig-Holstein. Empfehlungen der von der Landesrektorenkonferenz und der Ministerin für Bildung, Wissenschaft, Forschung und Kultur eingesetzten Expertenkommission. Online unter <http://www.tf.uni-kiel.de/de/dateien/fakultaet/strukturberichte/erichsen-gutachten-von-2003>, (letzter Zugriff am 22.08.2016)*

*Expertenkommission Wissenschaftsland Bayern 2020 (2005): Wissenschaftsland Bayern 2020. Empfehlungen einer internationalen Expertenkommission. München*

*Fischer, Klaus (2011): Soziale und kognitive Aspekte des Peer Review-Verfahrens. In: Fischer, Klaus; Parthey, Heinrich (Hrsg.): Evaluation wissenschaftlicher Institutionen. Berlin (Wissenschaftsforschung, Jahrbuch 2003). 2. Aufl., S. 23–62*

*Grande, Edgar; Jansen, Dorothea; Jarren, Otfried; Schimank, Uwe; Weingart, Peter (2014): Die neue Governance der Wissenschaft. Zur Einleitung. In: Grande, Edgar et al. (Hrsg.): Neue Governance der Wissenschaft. Reorganisation – externe Anforderungen – Medialisierung. Bielefeld, S. 15–48*

*Hazelkorn, Ellen (2013): Reflections on a Decade of Global Rankings: What we've learned and outstanding issues. Beiträge zur Hochschulforschung 35 (2), S. 8–33.*

*Hazelkorn, Ellen; Ryan, Martin (2013): The Impact of University Rankings on Higher Education Policy in Europe: A Challenge to Perceived Wisdom and a Stimulus for Change. In Zgaga, Pavel; Teichler, Ulrich; Brennan, John (Hrsg.): The globalisation*

challenge for European higher education. Convergence and diversity, centres and peripheries (Higher education research and policy, Bd. 4), Frankfurt/M., S. 55–78

*Hessisches Hochschulgesetz (HHG HE) in der Fassung vom 14. Dezember 2009*

*Hochschulrektorenkonferenz (2014):* Empfehlungen der 16. Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt/Main. Zum Forschungsrating des Wissenschaftsrates. Bonn

*Hornbostel, Stefan (2007):* Theorie und Praxis von Hochschulrankings. In Statistisches Bundesamt (Hrsg.): Amtliche Statistik und Hochschulrankings (Statistik und Wissenschaft, Bd. 11). Wiesbaden, S. 6–13

*Hornbostel, Stefan; Klingsporn, Bernd; Ins, Markus von (2009):* Messung von Forschungsleistungen – eine Vermessenheit? In: Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen (Diskussionspapiere der Alexander von Humboldt-Stiftung). Bonn, S. 14–35

*Hüther, Otto; Krücken, Georg (2016):* Hochschulen. Fragestellungen, Ergebnisse und Perspektiven der sozialwissenschaftlichen Hochschulforschung. Wiesbaden

*In der Smitten, Susanne; Jaeger, Michael (2012):* Ziel- und Leistungsvereinbarungen als Instrument der Hochschulfinanzierung. Ausgestaltung und Anwendung. Hannover (HIS: Forum Hochschule, 16)

*Jaeger, Michael (2009):* Steuerung durch Anreizsysteme an Hochschulen. Wie wirken formelgebundene Mittelverteilung und Zielvereinbarungen? In: Bogumil, Jörg; Heinze, Rolf G. (Hrsg.): Neue Steuerung von Hochschulen. Eine Zwischenbilanz. Berlin (Modernisierung des öffentlichen Sektors. Sonderband, 34), S. 45–66

*König, Karsten (2009):* Hierarchie und Kooperation. Die zwei Seelen einer Zielvereinbarung zwischen Staat und Hochschule. In: Bogumil, Jörg; Heinze, Rolf G. (Hrsg.): Neue Steuerung von Hochschulen. Eine Zwischenbilanz. Berlin (Modernisierung des öffentlichen Sektors. Sonderband, 34), S. 29–44

*Krempkow, René; Landrock, Uta; Neufeld, Jörg; Schulz, Patricia (2013):* Intendierte und nicht-intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultätsinternen leistungsorientierten Mittelvergabe in der Medizin. Abschlussbericht des Projektes GOMED – Governance Hochschulmedizin. Hrsg. v. Institut für Forschungsinformation und Qualitätssicherung (IFQ). Berlin

*Lange, Rainer (2010):* Benchmarking, Rankings und Ratings. In: Simon, Dagmar et al. (Hrsg.): Handbuch Wissenschaftspolitik. Wiesbaden, S. 323–334

*Lenzen, Dieter (2012):* Ranking, Rating – Steuerung und Motivation. In: Erziehungswissenschaft 4, S. 13–25

*Leistungsbewertung in der Wissenschaft. Förderlinie des Bundesministeriums für Bildung und Forschung.* <http://www.hochschulforschung-bmbf.de/de/1333.php> (letzter Zugriff am 16. Februar 2017)



*Münch, Richard; Pechmann, Max (2009):* Der Kampf um Sichtbarkeit. Zur Kolonisierung des wissenschaftsinternen Wettbewerbs durch wissenschaftsexterne Evaluationsverfahren. In: Bogumil, Jörg; Heinze, Rolf G. (Hrsg.): Neue Steuerung von Hochschulen. Eine Zwischenbilanz. Berlin (Modernisierung des öffentlichen Sektors. Sonderband, 34), S. 67–92

*Matthies, Hildegard; Simon, Dagmar; Knie, Andreas (2008):* „Gefühlte“ Exzellenz. Implizite Kriterien der Bewertung von Wissenschaft als Dilemma der Wissenschaftspolitik. In: Matthies, Hildegard; Simon, Dagmar (Hrsg.): Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. Leviathan, Sonderheft 24. Wiesbaden, S. 331–343

*Neidhardt, Friedhelm (2010):* Selbststeuerung der Wissenschaft: Peer Review. In: Simon, Dagmar et al. (Hrsg.): Handbuch Wissenschaftspolitik. Wiesbaden, S. 280–292

*Neidhardt, Friedhelm (2006):* Forschungsevaluation. Soziologie 35 (4), S. 419–425.

*Reichert, Sybille (2013):* Jenseits der Leistungsüberprüfung – Diskussionspapier zur Suche nach einem neuen Umgang mit Qualitätssicherung an Hochschulen. In: Leistungsmessung und Qualitätssicherung in der Wissenschaft (SWTR Schrift, 3/2013), S. 22–47

*Sächsisches Hochschulgesetz (SächsHSG) in der Fassung vom 1. Januar 2011*

*Schedler, Kuno; Proeller, Isabella (2011):* New public management. 5., korrigierte Aufl. Bern, Stuttgart, Wien

*Torka, Marc (2011):* Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen. In: Hornbostel, Stefan; Schelling, Anna (Hrsg.): Evaluation: The new balance of power? Berlin (ifQ-Working-Paper, 9), S. 69–82

*Wissenschaftliche Kommission Niedersachsen (2006):* Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Bewertung des Evaluationsverfahrens. Hannover

*Wissenschaftsrat (2016):* Empfehlungen zur Weiterentwicklung der MINT-Bereiche an den Hochschulen des Landes Hamburg. Berlin

*Wissenschaftsrat (2014):* Empfehlungen zur Weiterentwicklung des Hochschulsystems des Saarlandes. Berlin

*Wissenschaftsrat (2013a):* Empfehlungen zur Weiterentwicklung des Hochschulsystems des Landes Bremen. Mainz

*Wissenschaftsrat (2013b):* Empfehlungen zur Weiterentwicklung des Hochschulsystems des Landes Sachsen-Anhalt. Braunschweig

*Wissenschaftsrat (2012):* Leitfaden der Evaluation universitätsmedizinischer Einrichtungen. Berlin

*Wissenschaftsrat (2011):* Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Halle

*Wissenschaftsrat. (2004):* Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung. Hamburg

Manuskript eingereicht: 15.10.2016  
Manuskript angenommen: 23.02.2017

**Anschrift des Autors:**

Dr. Axel Oberschelp

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)

Goseriede 9

30159 Hannover

E-Mail: [oberschelp@dzhw.eu](mailto:oberschelp@dzhw.eu)

Axel Oberschelp ist wissenschaftlicher Mitarbeiter/Projektleiter in der Abteilung Hochschulforschung, Arbeitsbereich Steuerung, Finanzierung, Evaluation des DZHW.