

Der Einfluss von Evaluierungen auf Forschungseinrichtungen am Beispiel der Leibniz-Gemeinschaft

Michelle Mallwitz, Philipp Adler, Guido Benzler

Begutachtungen sind fester Bestandteil des Wissenschaftssystems. In Form von institutionellen Evaluationen werden sie zudem immer stärker als Steuerungsinstrument im Sinne des New Public Managements eingesetzt. Die Forschungseinrichtungen der Leibniz-Gemeinschaft beispielsweise müssen sich spätestens alle sieben Jahre einer Begutachtung ihrer inhaltlichen und strukturellen Entwicklung stellen. Die Beurteilung entscheidet über die Weiterförderung der Einrichtung. Im vorliegenden Beitrag wird auf Basis einer Inhaltsanalyse dargestellt, welche Aspekte in der Bewertung besondere Beachtung finden und wie sich dies im Verlauf der letzten zwei Evaluationsrunden (2003–2007 und 2008–2015) verändert hat. Hiermit soll ein Beitrag zur Diskussion über Evaluation als Steuerungsinstrument geleistet und diskutiert werden, welche weiterführenden Fragen sich am Beispiel der WGL-Einrichtungen ergeben.

1 Einführung

Institutionelle Evaluationen spielen eine immer stärkere Rolle als Steuerungsinstrument an außeruniversitären Forschungseinrichtungen (vgl. u. a. *Heinze 2002*). Die Institute der Leibniz-Gemeinschaft (WGL) etwa müssen sich spätestens alle sieben Jahre einer Begutachtung ihrer inhaltlichen und strukturellen Entwicklung stellen. Die Beurteilung entscheidet über die finanzielle Weiterförderung. Bei erstmaliger negativer Beurteilung folgt die nächste Evaluierung in einem kürzeren Zeitabstand. Fällt diese ebenfalls negativ aus, droht der Einrichtung das Ende der Finanzierung. Der Evaluation kommt somit eine existenzielle Rolle zu.

Im vorliegenden Beitrag wird am Beispiel der WGL-Einrichtungen herausgearbeitet, welche Aspekte in der Evaluierung eine besondere Relevanz für die Überprüfung der wissenschaftlichen Exzellenz und der gesellschaftlichen Relevanz haben und wie sich diese im Verlauf der letzten zwei Evaluationsrunden verändert haben. Als Basis dient eine Inhaltsanalyse der Förderempfehlung des Senats in der veröffentlichten Senatsstellungnahme aller WGL-Einrichtungen, die in zwei aufeinanderfolgenden Evaluationsrunden (2003–2007 und 2008–2015) begutachtet wurden. Auf Basis dieser ersten Annäherung werden weiterführende Fragen hinsichtlich der eingesetzten Kriterien in

der Evaluation außeruniversitärer Forschungseinrichtungen diskutiert und der weitere Forschungsbedarf aufgezeigt. Bewusst ausgespart wird eine normative Einschätzung von Evaluationen als Steuerungsinstrument (siehe hierzu u.a. *Heinze/Arnold 2008, S. 711ff.*).

2 Institutionelle Evaluation

Angesichts dessen, dass der wissenschaftliche Erkenntnisgewinn einerseits hochspezifisch, andererseits von Unsicherheit geprägt ist, sind Peer-Review-Verfahren als Modi der Bewertung dem Wissenschaftssystem inhärent. *Evaluationen* definiert *Wilhelm Krull (2011)* als Betrachtung von Institutionen und Forschungsprogrammen auf einer „höher aggregierten Ebene“ (ebd., S. 15). *Institutionelle Evaluationen* lassen sich als Instrument im Sinne des New Public Managements (NPM) verstehen. NPM bezeichnet die Tendenz, „[ö]ffentliche Dienstleistungen [...] mit Blick auf die Arbeitsorganisation effizienter, hinsichtlich des Preis-Leistungs-Verhältnisses ökonomischer und mit Blick auf die gesellschaftliche Wirkung effektiver zu erbringen“ (*Lange 2008, S. 239*). Es wird ein „Quasi-Markt“ (ebd.) geschaffen, der den Wettbewerb und – so die Annahme – die Leistung der Forschungseinrichtungen stärken soll. Im Zuge des NPM werden staatliche Regulierungen zurückgefahren und die Eigenverwaltung der Forschungsinstitute gestärkt. Gleichzeitig wird die Mittelvergabe an wettbewerbliche Bedingungen geknüpft, und externe Kontrollgremien erhalten größere Bedeutung. Mit einer stärkeren Wettbewerbsorientierung fordern „*Politik und Gesellschaft Rechenschaft über Leistungen, Qualität und Nutzen staatlich geförderter Wissenschaft und Forschung*“ ein (*Heinze 2002, S. 14*). Rechtfertigung und Legitimierung werden zu zentralen Elementen. Die institutionelle Evaluation als Betrachtung einer gesamten Einrichtung dient hierfür als Messinstrument und betrachtet (vermeintlich) messbare Indikatoren (vgl. u.a. *Kromrey 2003*). Die Evaluation von Forschungsleistungen beruht in diesem Sinne auf dem Verständnis eines „*Input-Throughput-Output-Modells*“ (*Heinze 2002, S. 15*) von Forschung, in dem Ressourcen eingesetzt werden und am Ende eine Leistung steht.

Kritik am Evaluationsverfahren setzt häufig dort an, wo eine (zu) starke Output-Orientierung in der Betrachtung vermutet wird. Exemplarisch soll hier auf *Margit Osterloh (2010)* verwiesen werden, die auf die Heterogenität unterschiedlicher Disziplinen hinweist, auf Studien zur Subjektivität vermeintlich objektiver Peer-Review-Verfahren sowie auf die Gefahr, nicht intendierte, negative Effekte zu generieren, beispielsweise indem sich die Forscherinnen und Forscher stärker auf die Indikatoren als auf die eigentliche Leistung ausrichten (siehe hierzu u.a. *DFG 2010*).

3 Forschungsgegenstand: Leibniz-Gemeinschaft

Die Leibniz-Gemeinschaft (WGL) wurde 1997 gegründet und geht auf die *Einrichtungen der Blauen Liste*¹ zurück. Aktuell setzt sie sich aus 91 selbstständigen Forschungseinrichtungen zusammen, die von den Natur-, Ingenieur- und Umweltwissenschaften über die Wirtschafts-, Raum- und Sozialwissenschaften bis zu den Geisteswissenschaften reichen. Die WGL hat etwa 18 700 Beschäftigte, darunter 9 500 Wissenschaftlerinnen und Wissenschaftler, und einen gemeinsam von Bund und Ländern finanzierten Etat von insgesamt etwa 1,8 Milliarden Euro (vgl. *Gülker/Simon 2011, S. 86ff.*; *WGL 2017a*). Die Einrichtungen werden in der Regel alle sieben Jahre evaluiert. Ziel dieser Evaluation ist es „zu überprüfen, ob die Voraussetzungen der gemeinsamen Förderung eines Leibniz-Instituts (überregionale Bedeutung und gesamtstaatliches wissenschaftspolitisches Interesse) weiterhin gegeben sind“ (vgl. *WGL 2016*).

Als ehemalige *Einrichtungen der Blauen Liste* haben die WGL-Institute seit den 1980ern eine für deutsche Verhältnisse lange Tradition in der institutionellen Evaluation (vgl. *Krull 2011, S. 21*). 1994 beauftragte die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung den Wissenschaftsrat (WR) mit einer Systemevaluation der *Einrichtungen der Blauen Liste*. Die Systemevaluation wurde 2000 mit der Stellungnahme des WR abgeschlossen. Die Empfehlungen dieser Stellungnahme bilden die Grundlage für das heutige Evaluationsverfahren (vgl. *WR 2000*).

Die Verantwortung für das Evaluationsverfahren liegt beim Senat der WGL, bestehend aus den „für die gemeinsame Forschungsförderung zuständige[n] Bundes- und Landesminister[n], Präsidenten und Vorsitzende[n] deutscher Wissenschaftsorganisationen sowie weitere[n] Wahlmitglieder[n]“ (*WGL 2017b*). Der Senatsausschuss Evaluierung der WGL stellt für die Evaluation ein Gremium aus Fachgutachterinnen und Fachgutachtern zusammen. Dieses erhält zunächst einen umfangreichen Selbstbericht der Einrichtungen. Es folgt eine Begehung der Einrichtung, an der neben der Gutachtergruppe auch Vertreterinnen und Vertreter von Bund und Land beteiligt sind. Sowohl der Inhalt des Selbstberichts als auch die Kriterien für die Evaluierung sind in den *Grundsätzen des Evaluierungsverfahrens des Senats der Leibniz-Gemeinschaft* geregelt (vgl. *WGL-Senat 2014*).

Die Ergebnisse der Evaluation werden in Form eines Berichts der Einrichtung und dem Zuwendungsgeber zugeleitet (vgl. *Gülker/Simon 2011, S. 84ff.*). Der WGL-Senat erarbeitet auf Basis des Selbstberichts und der Begehung eine wissenschaftspolitische Stellungnahme, die eine Förderempfehlung beinhaltet. Eine erstmalige negative

¹Der Namen geht auf eine „in ihrer ursprünglichen Form auf blauem Papier gedruckten Anlage zur Ausführungsvereinbarung Forschungseinrichtungen der Rahmenvereinbarung Forschung (RV-Fo) aus dem Jahr 1975“ (Heinz u. Arnold 2008, S. 291) zurück.

Evaluation führt zu einem verkürzten Zyklus von etwa dreieinhalb Jahren bis zur nächsten Begehung. Eine zweimalige negative Evaluation kann zur Schließung einer Einrichtung führen, „wenn die Qualität der wissenschaftlichen Arbeit nicht gegeben ist, und/oder das gesamtstaatliche Interesse an einem Forschungsgebiet nicht mehr besteht“ (WR 2000, S. 36). Der Druck auf die Einrichtungen, ihre wissenschaftliche Exzellenz und Relevanz nachzuweisen, ist auch deshalb hoch, weil Landesinstitute sich Chancen ausrechnen, in die WGL und damit eine gemeinsame Bund-Länder-Förderung aufgenommen zu werden. Dieser „Aufstieg“ von Landesinstituten ist wiederum nur möglich, wenn er mit einem gleichzeitigen „Abstieg“ bestehender WGL-Einrichtungen einhergeht. Von den im Jahr 2016 veröffentlichten zehn Evaluierungen wurde bei acht Instituten zur vollen Förderdauer geraten. Ein weiteres Institut allerdings wird bereits 2020 wieder evaluiert und eines der Institute wird nicht mehr gefördert (WGL 2016).

4 Methodisches Vorgehen

Grundlage für die Untersuchung sind die über die WGL-Homepage veröffentlichten Förderempfehlungen innerhalb der Senatsstellungen der beiden bisherigen Evaluationszyklen (2003–2007 und 2008–2015).² Insgesamt sind 52 Institute der WGL in die Auswertung eingeflossen, die in beiden bisherigen Evaluationszyklen betrachtet wurden. In einem explorativen Vorgehen wurde in einem ersten Schritt mittels einer qualitativen Inhaltsanalyse herausgearbeitet, welche wiederkehrenden Empfehlungen zur Weiterentwicklung der Einrichtung über die veröffentlichten 104 Förderempfehlungen in der Betrachtung sind (vgl. Mayring 2010). Ausgewertet wurden Textstellen, die Kritik üben oder Verbesserungen nahelegen. In einem zweiten Schritt wurde untersucht, wie häufig die herausgearbeiteten Empfehlungen vorkommen, wobei zwischen der ersten und der zweiten Förderrunde differenziert wurde, um diese Werte schließlich vergleichen zu können. Einschränkend sei darauf hingewiesen, dass das Untersuchungsdesign lediglich die Häufigkeit der Codes betrachtet und damit implizit die Hypothese trifft, dass dies Aussagen über die Relevanz zulässt.

5 Darstellung der Ergebnisse

Im ersten Schritt der Analyse konnten zwölf Codes erhoben werden, die in den Senatsstellungen besonders häufig genannt werden. Als Orientierung dienen dabei die Bewertungskriterien aus den *Grundsätzen des Evaluierungsverfahrens des Senats der Leibniz-Gemeinschaft* (vgl. WGL-Senat 2014). Diese zwölf Codes konnten im Zuge

²Die Senatsstellungen sind unter <https://www.leibniz-gemeinschaft.de/ueber-uns/evaluierung/das-evaluierungsverfahren-des-senats/senatsstellungen/> (Zugriff: 01.10.2017) zugänglich. Im Folgenden wird auf die jeweilige Stellungnahme unter Angabe des Kürzels der Einrichtung und des jeweiligen Jahres Bezug genommen.

einer Clusteranalyse zu vier Kategorien zusammengefasst werden. Die Kategorie *Gesamtkonzept und Profil* greift die strategische Ausrichtung auf, während die anderen drei Kategorien *Publikationsleistungen*, *Drittmiteleinnahmen* und *Transfer* die erbrachten Leistungen der vergangenen drei Jahre in Betracht ziehen.

Die nachstehende Tabelle zeigt die entsprechenden Codes und Kategorien mit beispielhaften Ankerzitaten (vgl. Tabellen 1–4).

Tabelle 1: Codes und Ankerzitate Gesamtkonzept und Profil

Gesamtkonzept und Profil	
Code	Ankerzitat
Profilbildung	„Auf der Grundlage seiner Themen, Methoden und Infrastrukturen hat das ZALF ein Gesamtkonzept erarbeitet, dessen Kohärenz und Fokussierung zukünftig weiter gesteigert werden muss.“ (ZALF 2013, S. 2)
Strategische Weiterentwicklung	„Damit die GESIS für die Zukunft gerüstet ist, müssen die methodische Forschung gestärkt und vor allem integrierte Serviceangebote adressatengerecht mit der sozialwissenschaftlichen Klientel entwickelt werden.“ (GESIS 2005, S. 4)
Institutsübergreifende Zusammenarbeit	„Nach wie vor ist eine Verbesserung auf der Ebene der Abteilungs- und Forschungsgruppenleitungen nötig, um das gemeinsame Forschungspotential am HPI so weit wie möglich ausschöpfen zu können.“ (HPI 2010, S. 3)

Tabelle 2: Codes und Ankerzitate Publikationsleistungen

Publikationsleistungen	
Code	Ankerzitat
Quantitative Steigerung	„Kritisch ist vor allem, dass ein Teil der Wissenschaftler wenig oder gar nicht publiziert.“ (ATB 2007, S. 3)
Qualitative Steigerung	„Insbesondere die Anzahl der Veröffentlichungen in referierten Zeitschriften [...] sollten weiter gesteigert werden.“ (IfW 2005, S. 3)
Publikationsstrategie	„Empfehlenswert sind die Erarbeitung eines Publikationskonzeptes [...]“ (FIS 2006, S. 3)

Tabelle 3: Codes und Ankerzitate Drittmiteleinnahmen

Drittmittel	
Code	Ankerzitat
Steigerung der gesamten Drittmittel	„Der Senat erwartet, dass das ZPID die Drittmittelerträge in den kommenden Jahren im Rahmen des Ausbaus von Forschung und Entwicklung deutlich steigert.“ (ZPID 2011, S. 4)
Steigerung der DFG-Mittel	„Allerdings empfiehlt der Senat, den Anteil von DFG-Mitteln innerhalb des Drittmittelportfolios, zu dem derzeit nur einzelne Institute und Beschäftigte beitragen, erheblich zu steigern.“ (ZALF 2013, S. 3)
Steigerung der EU-Mittel	„Insbesondere die Einnahmen aus Drittmitteln von der DFG und der EU muss das ATB deutlich erhöhen.“ (ATB 2014, S. 3)

Tabelle 4: Codes und Ankerzitate Transfer

Transfer	
Code	Ankerzitat
Qualität der Transferleistungen	„Hinsichtlich der Konjunkturanalysen und der Branchenforschung hat die Bewertungsgruppe festgestellt, dass sich die Qualität der zugrundeliegenden ökonomischen Modelle, statistischen Ansätze und Umfragemethoden stärker an dem heutigen Stand der Forschung orientieren sollte.“ (ifo 2006, S. 3)
Kontakte zu Politik und Gesellschaft	„Darüber hinaus wäre es wünschenswert, wenn das IDS versuchte, auf die Gestaltung des Urheberrechts mit dem Ziel einzuwirken, der Wissenschaft den Zugriff auf KobRA zu Forschungszwecken zu ermöglichen.“ (IDS 2004, S. 3)
Patente und Ausgründungen	„Der Senat empfiehlt, die einzelnen Aktivitäten zur Verwertung von Arbeitsergebnissen in einer übergreifenden Verwertungsstrategie zusammenzuführen.“ (IKZ 2012, S. 3)

Im zweiten Schritt wurde eine Zählung dieser Bewertungsaspekte vorgenommen und beide Evaluationsrunden einander gegenübergestellt. Eine Betrachtung der Nennungen der Schwachstellen und Kritikpunkte ohne eine vorherige Aufteilung auf die beiden Evaluationsperioden verdeutlicht, dass der Bereich *Drittmiteleinahmen* am häufigsten genannt wird (123 Nennungen). An zweiter und dritter Stelle stehen *Publikationsleistungen* (106 Nennungen) beziehungsweise die Aspekte rund um das *Gesamtkonzept und Profil* (97 Nennungen). *Transfer* spielt mit 43 Nennungen für die Evaluationen laut den vorliegenden Ergebnissen eine eher untergeordnete Rolle (vgl. Tabelle 5).

Tabelle 5: Anzahl der Nennungen nach Codes

Kategorie	Codes	Anzahl der Nennungen	Anzahl der Nennungen
Gesamtkonzept und Profil	Profilbildung	56	97
	Strategische Weiterentwicklung	18	
	Institutsübergreifende Zusammenarbeit	23	
Publikationsleistung	Quantitative Steigerung	46	106
	Qualitative Steigerung	54	
	Publikationsstrategie	6	
Drittmiteleinahmen	Steigerung	39	123
	DFG-Mittel	51	
	EU-Mittel	33	
Transfer	Qualität der Transferleistungen	23	43
	Kontakte zu Politik und Gesellschaft	11	
	Patente und Ausgründungen	9	
		369	

Bei genauerer Betrachtung der einzelnen Themenbereiche erscheinen manche Kritikpunkte besonders entscheidend. So wurden beispielsweise die Profilbildung, die Steigerung der DFG-Mittel und die qualitative Steigerung der Publikationen jeweils etwa bei der Hälfte der untersuchten Evaluationen als Kritikpunkt genannt. Ebenfalls von Bedeutung für viele der Einrichtungen aus Sicht der Förderempfehlung ist die quantitative Steigerung von Publikationsleistungen. Betrachtet nach Evaluationsrunden ist die Profilbildung das meistgenannte Kriterium in den Empfehlungen.

Tabelle 6: Anzahl der Nennungen nach Evaluationsrunden (2003–2007, 2008–2015)

Kategorie	Codes	Anzahl der Nennungen (2003–2007)	Anzahl der Nennungen (2008–2015)
Gesamtkonzept und Profil	Profilbildung	32	24
	Strategische Weiterentwicklung	10	8
	Institutsübergreifende Zusammenarbeit	17	6
Publikationsleistung	Quantitative Steigerung	27	19
	Qualitative Steigerung	30	24
	Publikationsstrategie	2	4
Drittmiteleinnahmen	Steigerung	22	17
	DFG-Mittel	29	22
	EU-Mittel	20	13
Transfer	Qualität der Transferleistungen	13	10
	Kontakte zu Politik und Gesellschaft	4	7
	Patente und Ausgründungen	6	3
		212	157

Insgesamt ist die Zahl der kritischen Nennungen der verschiedenen Aspekte zwischen den beiden Bewertungsperioden zurückgegangen (vgl. Tabelle 6). Besonders auffällig ist dies bei den Punkten *institutsübergreifende Zusammenarbeit* (17 Nennungen → 6 Nennungen), *quantitative Steigerung der Publikationsleistungen* (27 Nennungen → 19 Nennungen) und *Steigerung von DFG- und EU-Mitteln* (jeweils sieben Nennungen weniger als in der vorherigen Evaluationsrunde). Nur zwei Codes wurden in der zweiten Evaluationsrunde häufiger genannt als in der ersten. Dabei ist auffällig, dass es sich im Verhältnis zu den Gesamtnennungen um eher weniger relevante Punkte zu handeln scheint (*Kontakte zu Politik und Gesellschaft* und *Publikationsstrategie*). Überraschend scheint, dass die Aspekte in der Kategorie *Transfer* einen so geringen Stellenwert einnehmen, obwohl der Austausch mit Gesellschaft, Politik und Wirtschaft zum Leitbild der WGL gehört.

6 Diskussion der Ergebnisse

Die besonders häufige Nennung des Aspekts *Profilbildung* deutet daraufhin, dass neben den quantitativ messbaren Leistungen die strategische Ausrichtung der Einrichtungen eine wichtige Rolle für ihre Bewertung spielt. *Silke Gülker* und *Dagmar Simon* kamen in einer qualitativen Betrachtung der Evaluationen ausgewählter WGL-Einrichtungen zu dem Schluss, dass die institutionellen Evaluationen eine Reflektion der strategischen Ausrichtung mit sich bringen, da auf inhaltlicher Ebene eine Einschätzung der ganzen Einrichtung vorgenommen wird (ders. 2011, S. 90). Die häufige Bezugnahme auf das Thema *Profilbildung* in der Analyse scheint dies zu stützen. *Heinze* und *Arnold* (2008) weisen in diesem Zusammenhang darauf hin, dass die WGL gegenüber den anderen großen außeruniversitären Forschungseinrichtungen aufgrund der Zusammenstellung und der internen Konkurrenzsituation eine weitaus größere Binnendifferenzierung aufweist. (vgl. ebd.) Die Profilierung erfolgt damit stärker auf Einrichtungsebene als auf Ebene der Gesamtorganisation, was diesem Merkmal zusätzlich Gewicht verleiht.

Die Bedeutung quantitativ messbarer Daten, die *Gülker* und *Simon* (2011) noch als sehr hoch einschätzten, scheint in der vorliegenden Betrachtung in der zweiten Evaluationsrunde leicht abzunehmen. Es zeigt sich, dass innerhalb der Evaluation als NPM-Instrument weitere Instrumente im Sinne des NPM gefördert werden. Deutlich macht dies beispielsweise dieser Auszug aus einer Senatsstellungnahme: *„Der Senat empfiehlt den Zuwendungsgebern, dem Institut nach Einführung von Kosten-Leistungsrechnung und Programmbudgets nun auch die benötigte Flexibilität bei der Bewirtschaftung der Ressourcen zu gewähren, indem die Stellenpläne flexibilisiert werden und eine überjährige Bewirtschaftung des Haushaltes ermöglicht wird.“* (WGL-Senat 2006, S. 2)

Die Verbindung aus der Betrachtung der Strategie der Einrichtung sowie der Forschungsleistungen kann als Versuch verstanden werden, sowohl die Vergangenheit als auch die Zukunft in die Bewertung miteinzubeziehen. Dabei können sinnvolle und ausformulierte strategische Ziele und ein klares Profil eine hohe Leistungsfähigkeit in den kommenden Jahren zwar nicht garantieren, jedoch ermöglichen sie eine Einschätzung der Ausrichtung und der Ziele der Einrichtung.

Interessant ist, dass der Transfer wissenschaftlicher Erkenntnis – obwohl elementarer Teil des WGL-Selbstverständnisses – in der Evaluation eine untergeordnete Rolle zu spielen scheint. Hier spielt es sicherlich eine Rolle, dass die außeruniversitären Forschungseinrichtungen – trotz größerer Ressourcenausstattung pro Professorin/Professor – gegenüber den Universitäten *„nur etwa ein Fünftel zur Gesamtmenge aller vom staatlich finanzierten Forschungssektor produzierten wissenschaftlichen*

Publikationen“ beitragen (Heinze und Arnold, 2008, S. 693) und damit der Fokus in den Empfehlungen stärker auf den Publikationsleistungen als auf dem Transfer liegt.

Die Ergebnisse zeigen, dass die Nennung kritischer Aspekte in den Senatsstellungen im zweiten Durchlauf insgesamt abnimmt. Für den Rückgang können zwei mögliche Erklärungsansätze herangezogen werden: Es ist möglich, dass die Institute auf die Empfehlungen der ersten Runde der Evaluierungen reagiert und Veränderungen angestrebt haben. Diese Hypothese könnte in einer weiterführenden Untersuchung mit Blick auf die Selbstberichte der Einrichtung in der zweiten Evaluationsrunde untersucht werden. Heinze und Arnold (2008) weisen nach, dass die WGL sich einem „*seitens der Forschungspolitik stark gewachsenen Druck nach effizientem Ressourceneinsatz und Anwendungsorientierung*“ (ebd., S. 711) stellen muss, im Zuge dessen die WGL-Einrichtungen deutlich höhere Zuwachsraten bei Publikationen und Patentmeldungen verzeichnen als beispielsweise die Max-Planck- oder die Fraunhofer-Gesellschaft (vgl. ebd.). Gülker und Simon (2011) weisen darauf hin, dass Empfehlungen aus der Evaluation heraus entweder direkt umgesetzt oder aber die entsprechenden Daten im Selbstbericht in der nächsten Runde positiver präsentiert werden (vgl. ebd., S. 91). Stefan Hornbostel (2011) spricht in diesem Zusammenhang zynisch von „*mechanistischen Umsetzungen*“ und „*ressourcenintensive(n) Inszenierungen*“ (ebd., S. 9). Auffällig ist dennoch, dass in beiden Runden genau gleich viele Institute mit einer verkürzten Förderdauer belegt und damit „schlecht“ bewertet wurden.

Eine zweite Schlussfolgerung könnte sein, dass sich aufgrund übergeordneter wissenschaftspolitischer Diskurse der Fokus des Senats gewandelt hat und nun neben den Leistungen auch die weniger gut quantitativ abbildbaren Aspekte wie Governance der Einrichtung oder die Vernetzung mit anderen Instituten oder Hochschulen stärker in den Blick genommen werden. Hier stellt sich also die Frage, inwieweit die Einrichtungen sich in der zweiten Runde – gemessen an den in den Förderempfehlungen gewählten Kriterien – tatsächlich verbessert haben oder inwieweit sie gelernt haben, sich besser zu präsentieren.

Es ist zu vermuten, dass eine reduzierte Förderdauer höheren Veränderungsdruck erzeugt. Gülker und Simon (2011) deuten in ihrer Untersuchung darauf hin, dass die institutionelle Evaluation der Einrichtungsleitung ebendieser eine neue Möglichkeit der Steuerung und der Legitimation ihres Handelns gibt (vgl. ebd., S. 91). In beiden Evaluationsrunden ist der Fokus auf das Profil und die Weiterentwicklung der Einrichtung hoch. Aspekte dieses Clusters beziehen sich ganz wesentlich auf die Aufgabenbereiche der Einrichtungsleitung. Hinzu kommt, dass im Rahmen der Senatsstellungen immer wieder auch Punkte genannt werden, die als NPM-Instrumente die Leitung der Einrichtung stärken.

7 Ausblick

Die Diskussion der Ergebnisse bietet im Sinne der Research Note eine Basis für weiterführende Hypothesen: Dem Vorwurf, dass Evaluationen keinen Einfluss hätten, kann auf Basis der vorliegenden Erhebung widersprochen werden. Ob allerdings die gewünschten Wirkungen erzielt werden oder die Einrichtungen beispielsweise „nur“ ihre Selbstinszenierungsfähigkeiten trainieren oder aber andere, nicht intendierte Wirkungen mit der Evaluation einhergehen, kann diese Untersuchung nicht beantworten. *Torka (2011)* weist generell auf die Schwierigkeit hin, kausale Mechanismen zwischen Evaluationen und wissenschaftlichem Handeln empirisch zu untersuchen (vgl. ebd. S. 70).

Aufbauend auf dieser ersten Betrachtung schließen sich daher weiterführende interessante Fragen an: In Hinblick auf die quantifizierbaren Kriterien wie etwa die Publikationsleistungen ließe sich beispielsweise untersuchen, ob sich zwischen Einrichtungen, die positiv bewertet wurden und solchen, die mit einer kürzeren Frist bis zur nächsten Evaluation belegt wurden, Unterschiede zeigen. Durch die stete Zunahme der Datengrundlage in Form von zusätzlichen Evaluierungen samt Evaluierungstexten sowie die immer transparentere Aufbereitung von Publikationsleistungen ist dies ein Ansatz, der sich auf Basis der guten Datengrundlage mittelfristig weiterverfolgen lässt. Erste statistische Auswertungen zeigen durchaus eine Veränderung der Publikationsleistungen nach negativen Evaluationen auf, doch ermöglicht die geringe Zahl der negativen Evaluierungen noch keine signifikante Aussage.

Auch wäre es für die weitere Betrachtung der Evaluation als Steuerungsinstrument fruchtbar zu untersuchen, inwieweit die Empfehlungen der Senatsstellungen sich auf die Position der Einrichtungsleitung auswirken und wie sie von dieser genutzt werden. Dieser Ansatz könnte mit qualitativen Interviews mit den Leitungen und der mittleren Führungsebene der Einrichtung vorangetrieben werden, auch im Hinblick auf Unterschiede bei sehr positiven und eher negativen Bewertungen.

In der Senatsstellungnahme wurde ein besonderer Fokus auf die Empfehlung des Senats gelegt (in Abgrenzung zum Bericht der unabhängigen Bewertungsgruppe, der neben dem Bericht der evaluierten WGL-Einrichtung ebenfalls Bestandteil der Senatsstellungnahme ist). Diese abschließende Empfehlung greift die Einschätzungen der Fachgutachterinnen und Fachgutachter auf, ist darüber hinaus aber auch aufgrund der Aufgabe und Zusammensetzung des Senats in seiner Einschätzung stärker wissenschaftspolitisch geleitet. In einem weiteren Schritt wäre es daher lohnenswert, die Kriterien und deren Häufigkeit in den Berichten der Fachgutachterinnen und Fachgutachter den abschließenden Empfehlungen des Senats gegenüberzustellen, um zu vergleichen, wie viel Einfluss die Gutachtergruppe tatsächlich auf die finale Entscheidung hatte. Eine Auffälligkeit, die sich bei der Untersuchung des Materials zeigte, ist, dass seitdem sich die DFG für ein Kaskadenmodell in der Frauenförderung aus-

gesprachen hat, dieser Begriff öfter in den Senatsstellungen auftaucht. Interessant könnte es dementsprechend sein, die Codes und Kategorien in den Senatsstellungen mit anderen einflussreichen wissenschaftspolitischen Papieren zu vergleichen und in diesem Kontext zu deuten.

Literatur

DFG (2010): DFG-Pressemitteilung Nr. 7, 23. Februar 2010 http://www.dfg.de/service/presse/pressemitteilungen/2010/pressemitteilung_nr_07/index.html (Zugriff: 04.11.2016)

Gülker, Silke; Simon, Dagmar (2011): Nach der Evaluation ist vor der Evaluation? Institutionelle Folgen von Forschungsbewertungen im internationalen Vergleich. In: Hornbostel, Stefan; Schelling, Anna (Hrsg.): *Evaluation: New Balance of Power?* iFQ-Working Paper No. 9. Berlin, S. 83–93

Heinze, Thomas; Arnold, Natalie (2008): Governanceregimes im Wandel. Eine Analyse des außeruniversitären, staatlich finanzierten Forschungssektors in Deutschland. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 60, 2008, S. 686–720

Heinze, Thomas (2002): Evaluation von Forschungsleistungen: Konzeptionelle Überlegungen und Situationsbeschreibung für Deutschland. In: *Wissenschaftsmanagement*, 2, 2002, 6, S. 14–22

Hornbostel, Stefan (2011): Resonanzkatastrophen, Eigenschwingungen, harmonische und chaotische Bewegungen. In: Hornbostel, Stefan; Schelling, Anna (Hrsg.): *Evaluation: New Balance of Power?* iFQ-Working Paper No. 9. Berlin, S. 7–14

Kromrey, Helmut (2003): Evaluation in Wissenschaft und Gesellschaft. Vortrag bei der Feierstunde zur Eröffnung des Centrums für Evaluation am 21.2.2003 in Saarbrücken. http://www.hkromrey.de/Kromrey_EvalWissG.pdf (Zugriff: 04.11.2016)

Krull, Wilhelm (2011): Bewertung, Begutachtung und Evaluation in Wissenschaft und Forschung. In: Hornbostel, Stefan; Schelling, Anna (Hrsg.): *Evaluation: New Balance of Power?* iFQ-Working Paper No. 9. Berlin, S. 15–23

Lange, Stefan (2008): *New Public Management und die Governance der Universitäten*. In: *dms – der moderne staat – Zeitschrift für Public Policy, Recht und Management*, 1, 2008, S. 235–248

Mayring, Philipp (2010): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim

Osterloh, Margit (2010): *Governance by Numbers*. Does it really work in research? In: *Analyse & Kritik*, 2, 2010, S. 267–283

Torka, Marc (2011): Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen. In: Hornbostel, Stefan; Schelling, Anna (Hrsg.): *Evaluation: New Balance of Power?* iFQ-Working Paper No. 9. Berlin, S. 69–82

WGL (2017a): Die Leibniz-Gemeinschaft. <http://www.leibniz-gemeinschaft.de/ueberuns/> (Zugriff: 01.10.2017)

WGL (2017b): Senat. <https://www.leibniz-gemeinschaft.de/ueber-uns/organisation/senat/> (Zugriff: 01.10.2017)

WGL (2016): Das Evaluierungsverfahren des Senats der Leibniz-Gemeinschaft. <http://www.leibniz-gemeinschaft.de/ueber-uns/evaluierung/das-evaluierungsverfahren-des-senats/> (Zugriff: 04.11.2016)

WGL-Senat (2014): Grundsätze des Evaluierungsverfahrens des Senats der Leibniz-Gemeinschaft. http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Evaluierung/Austausch_August_2014/Anlage_3_-_Kriterien_fuer_die_Evaluierung_von_Einrichtungen_der_WGL.pdf (Zugriff: 04.11.2016)

WGL-Senat (2006): Stellungnahme zum Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK). https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Evaluierung/Senatsstimmungen/Senatsstimmung-IPK-2006.pdf (Zugriff: 01.10.2017)

WR (2000): Systemevaluation der Blauen Liste – Stellungnahme des Wissenschaftsrates zum Abschluß der Bewertung der Einrichtungen der Blauen Liste. <http://www.wissenschaftsrat.de/download/archiv/4703-00.pdf> (Zugriff: 04.11.2016)

Artikel eingereicht: 01.10.2017
Artikel angenommen: 09.01.2018

Anschriften der Autoren:

Michelle Mallwitz M.A.
Referentin für Studiengangsentwicklung
Goethe-Universität Frankfurt am Main
Theodor-W.-Adorno-Platz 1
60323 Frankfurt am Main
E-Mail: mallwitz@em.uni-frankfurt.de

Jan Philipp Adler M.Sc.
Berater für Hochschulen und außeruniversitäre Forschungseinrichtungen
rheform EntwicklungsManagement GmbH
Kasernenstraße 27
40213 Düsseldorf
E-Mail: philipp.adler@rheform.de

Prof. Dr. Guido Benzler
Geschäftsführer der rheform EntwicklungsManagement GmbH, Beratungsunternehmen für Hochschulen und außeruniversitären Forschungseinrichtungen
rheform Entwicklungsmanagement GmbH
Kasernenstraße 27
40213 Düsseldorf
E-Mail: guido.benzler@rheform.de