

Externe Kontrolle und kollegiale Rückmeldung. Zu einer konstitutiven Spannung in institutionellen Evaluationen der Wissenschaft

Silke Gülker,¹ Dagmar Simon,¹ Marc Torka¹

Zusammenfassung: Evaluationen werden üblicherweise als ein Mechanismus zur externen Kontrolle angesehen. Im Fall von Wissenschaftsevaluationen wird damit allerdings vernachlässigt, dass Evaluationen von Kolleg(inn)en betrieben werden, sie damit Züge einer Selbstkontrolle tragen und eine doppelte Funktion übernehmen: Sie sollen nicht nur kontrollieren, sondern zugleich Rückmeldungen geben, wie Forschungen und deren Organisationsweisen zukünftig verbessert werden könnten. Beide Funktionen konstituieren eine strukturell unaufhebbare Spannung in Evaluationsprozessen. Am Beispiel von institutionellen Forschungsevaluationen wird in diesem Beitrag gezeigt, wie Evaluierende versuchen und Evaluierete hoffen, selbst im Kontext externer Kontrollen eine kollegiale Rückmeldung zu geben bzw. zu erhalten. Unsere Argumentation stützt sich auf die Untersuchung von drei verschiedenen strukturierten Evaluationsverfahren: das der deutschen Leibniz-Gemeinschaft, des niederländischen Standard Evaluation Protocols und der Research Assessment Exercises in Großbritannien.

Schlagwörter: Forschungsevaluation, Peer-Review, Universitäten, Außeruniversitäre Forschung

External Control and Collegial Feedback: Regarding a Constitutive Tension in Institutional Research Evaluation

Abstract: Evaluations are typically analyzed as external control mechanisms. In the case of science assessment, it is often left aside that colleagues operate these evaluations. Therefore, they are structured as self-control with two functions: next to ensuring control they provide evaluatees with feedback to facilitate future improvements for individual research and organizational development. These two functions produce a constitutive tension hardly to be brought into a stable equilibrium. We use the example of institutional research evaluations to illustrate how evaluators attempt to provide, and evaluatees hope to gain collegial feedback even within an external control context. The argument is backed up by examining three differently structured evaluation procedures, and accordingly, different degrees of tension: those of the Leibniz Association in Germany, the Standard Evaluation Protocol in the Netherlands and the Research Assessment Exercise in the United Kingdom.

Keywords: Research Evaluation, Peer Review, University, Non-University Research

¹ Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

1. Einleitung

Seit den 1980er Jahren finden im Bereich der Wissenschaft regelmäßige Evaluationen statt. Diese Entwicklung wird oftmals als Anzeichen einer verstärkten externen Kontrolle, eines Autonomieverlustes und einer fortschreitenden Deprofessionalisierung der Wissenschaft gedeutet. Tatsächlich aber nehmen Evaluationen in der Wissenschaft noch immer eine spezielle Form an. Sie sind zwar nicht freiwillig, aber sie werden von Kolleginnen und Kollegen vorgenommen, so dass die Grenzen zwischen externen Fremd- und internen Selbstkontrollen verschwimmen. Peer-Review als zentraler Selbststeuerungsmechanismus der Wissenschaft wird allerdings auf neue Gegenstände und Funktionen übertragen. Denn in regelmäßigen Verfahren der institutionellen Evaluation urteilen wissenschaftliche Gutachtergruppen über die Qualität und Produktivität von ganzen Forschungsorganisationen. Deren zukünftige Finanzierung hängt von der Teilnahme an diesen Verfahren und dem dort erzielten Ergebnis ab. Das wissenschaftliche Peer-Review wird also für das wissenschaftspolitische Ziel genutzt, über regelmäßige Leistungskontrollen Finanzierungsentscheidungen zu treffen. Wir bezeichnen ein Verfahren als ‚extern‘ initiiert, wenn eben solche politischen Entscheidungen zur Ressourcenallokation damit verbunden sind.²

Während es über das Peer-Review für Publikationen und Projekte einen breiten Diskurs gibt,³ sind die internen Prozesse von institutionellen Evaluationen wenig bekannt. Die damit befassten Arbeiten heben meist hervor, dass die Wissenschaft über diese Verfahren stärker extern kontrolliert werde, an Autonomie einbüße und nicht unbedingt wissenschaftsspezifische Leistungskriterien zur Anwendung kämen (vgl. Whitley/Gläser/Engwall 2010), da etwa quantitative Outputquoten (vgl. Kieser 2010) oder Fragen der angemessenen Organisationsweise von Forschung (vgl. Meier/Schimank 2010) eine bedeutende Rolle in diesen Verfahren spielten. Mit dieser Fokussierung, institutionelle Evaluationen primär unter dem Gesichtspunkt der externen Kontrolle und Bewertung zu betrachten, geht allerdings auch eine Vereinseitigung einher. Peer-Review hatte nämlich schon immer mehrere Funktionen, die man auch in institutionellen Evaluationen wiederfindet. Neben der Kontrolle von Standards und der Selektion von Publikationen oder Forschungsprojekten geht es vor allem darum, Lernprozesse anzustoßen (vgl. Blank 2008). Das Feedback an Kolleg(inn)en dient der „kollektiven Produktion“ (Gläser 2006: 347) und soll zur Verbesserung von Beiträgen führen. Kollegiales Feedback in Form von ‚Empfeh-

2 In jedem Fall muss man die Unterscheidung zwischen interner Selbst- und externer Fremdkontrolle reflexiv verwenden. Denn selbst eine von der Universitätsleitung selbstinitiierte Evaluation kann für die Fachbereiche zugleich eine Fremdkontrolle bedeuten.

3 Zur Überprüfung der Glaubwürdigkeit und Verlässlichkeit wissenschaftlicher Beiträge oder technischer Erfindungen ist das Peer Review bereits seit der Geburt der modernen Wissenschaft im 17. Jahrhundert zentral (vgl. Shapin 1994). Auf klassische Arbeiten zum Peer Review als wichtigem Selbststeuerungsmechanismus der Wissenschaft (vgl. Merton/Zuckerman 1971) folgten unzählige kritische Studien, die insbesondere die Reliabilität, Validität und Fairness des Peer Reviews in Frage stellen. Für Peer Reviews von Publikationen zum Beispiel Weller (2001) und von Projekten Cole/Cole (1981). Einen Überblick liefert Hirschauer (2004).

lungen‘ ist zwar nur in manchen Verfahren der institutionellen Evaluationen explizit vorgesehen und nicht alle sind zugleich mit Finanzierungsentscheidungen verbunden. Wir werden aber zeigen, dass kollegiales Feedback für Evaluierende und Evaluierte selbst dann eine bedeutende Rolle spielt, wenn Evaluationsverfahren dafür nicht entwickelt wurden. Deshalb kombinieren alle institutionellen Evaluationen, die auf einem Peer-Review-Prozess aufbauen, Elemente der externen Kontrolle und des kollegialen Feedbacks in einem spannungsreichen Prozess: „The inherent relation between learning and control in evaluation implies that both dimensions will always be present, even when one dominates the other“ (Hansson 2010: 240).

Auf der Basis einer Studie, in der die internen Prozesse von institutionellen Forschungsevaluationen in drei Verfahren analysiert wurden, werden wir zeigen, dass kollegiales Feedback sowohl für Evaluierende als auch für Evaluierte und in unterschiedlicher Ausprägung für die Verfahren selbst eine zentrale Orientierung ist. Zwischen externer Kontrolle und kollegialem Feedback besteht allerdings eine konstitutive Spannung, die kaum in ein stabiles Gleichgewicht gebracht werden kann. Wie Evaluierende unter der ungünstigen Rahmenbedingung externer Kontrolle aber dennoch versuchen, ein kollegiales Feedback zu geben, und Evaluierte hoffen, ein solches zu erhalten, ist die zentrale Frage unseres Beitrags. Damit wollen wir am Beispiel von Evaluationen in der Wissenschaft die zumeist ausgeblendete zweite Seite der spannungsreichen Doppelfunktion von Evaluationen wieder in den Diskurs zurückbringen.

Dieses konstitutive Spannungsverhältnis analysieren wir in zwei grundsätzlich verschiedenen Verfahrenstypen. In den Evaluationsverfahren der deutschen Leibniz-Gemeinschaft (WGL) und des niederländischen Standard Evaluation Protocol (SEP) sind interaktiv gewonnene und inhaltliche Rückmeldungen („Empfehlungen“) eine systematische Komponente. Hingegen sind in der britischen Research Assessment Exercise (RAE) der Austausch und die Interaktion mit den evaluierten Forschungseinheiten verboten und die Ergebnisse werden in Form eines numerischen Rankings kommuniziert, das zugleich die Basis für die selektive Forschungsfinanzierung der Politik ist. Mit unserer Analyse beziehen wir uns also auf maximal kontrastierende Verfahrenstypen und können damit zugleich Annahmen in Bezug auf eine große Spannbreite von unterschiedlichen Verfahren treffen. Gleichwohl ist die Vielfalt der international eingesetzten Verfahren zur institutionellen Forschungsevaluation weit größer, als wir sie in diesem Artikel repräsentieren könnten.⁴ Wir gehen allerdings davon aus, dass ein Spannungsverhältnis zwischen den unterschiedlichen Logiken von externer Kontrolle und kollegialem Feedback für jedes extern initiierte Verfahren – also für jedes Verfahren mit Budgetauswirkungen –

4 Für eine weitere Verallgemeinerung unserer Ergebnisse wäre deshalb ein Vergleich mit Verfahren interessant, die kein Peer Review vorsehen (vgl. Gläser/Laudel 2007) oder nicht unmittelbar mit externer Kontrolle im Zusammenhang stehen, weil manche institutionelle Audits keine direkten Budgetauswirkungen haben oder von wissenschaftlichen Einrichtungen sogar selbst initiiert werden.

von Bedeutung ist. Wie genau dieses Verhältnis je nach Verfahren austariert wird, ist eine empirische Frage.

Nach einer kurzen Beschreibung des methodischen Vorgehens (Kapitel 2) werden wir theoretisch explizieren, welche Spannungen auftreten müssten, wenn externe Kontrolle und kollegiales Feedback in einem Prozess aufeinander treffen (Kapitel 3). Nach einer kurzen Darstellung der wesentlichen Verfahrensregeln beider Evaluationstypen (Kapitel 4) zeigen wir schließlich, wie sich das Spannungsverhältnis bei Evaluierenden und Evaluierten im Verfahren der Leibniz-Gemeinschaft und des SEP (Kapitel 5) sowie in der RAE (Kapitel 6) empirisch zum Ausdruck bringt. Abschließend diskutieren wir, welche Bedeutung unsere Ergebnisse für den weiteren Diskurs über (Wissenschafts-)Evaluationen haben (Kapitel 7).

2. Methodisches Vorgehen

Die Untersuchung ist Teil eines Projektes, das danach fragt, wie Gutachterbewertungen in unterschiedlichen Verfahrenskontexten zustande kommen und welche Implikationen damit für die bewerteten Forschungsinstitute verbunden sind.⁵ Mit den Verfahren der WGL in Deutschland und nach dem Standard Evaluation Protocol (SEP) in den Niederlanden auf der einen Seite und der RAE in Großbritannien auf der anderen Seite wurden Prozesse in Verfahrenskontexten untersucht, die maximal kontrastieren. Die je Kontext ausgewählten Evaluationsfälle (ein Fall ist ein institutsbezogenes Verfahren) sind sowohl in Bezug auf Fächergruppen als auch in Bezug auf Institutstypen heterogen. Vertreten sind mono- wie auch interdisziplinäre Einrichtungen (vgl. Tabelle 1) aus außeruniversitären Forschungsorganisationen (WGL), Akademien (SEP)⁶ und Universitäten (RAE).

Weil Gutachtergespräche prinzipiell hinter verschlossenen Türen stattfinden,⁷ galt es eine Materialbasis zu schaffen, die dennoch eine weitgehende Rekonstruktion einzelner Begutachtungsprozesse ermöglicht. Um dem Problem geschönter Beschreibungen von wissenschaftspolitisch höchst brisanten Evaluationsprozessen zu begegnen, haben wir mehrfache Vorkehrungen getroffen. Evaluationsprozesse wurden multiperspektivisch, d.h. aus der Sicht mehrerer Akteure beschrieben. Die Beschreibungen konnten deshalb sowohl auf ihre Konsistenz innerhalb eines Inter-

5 Wir haben an anderer Stelle die Rollen und Rollenverständnisse von Gutachter(inne)n (vgl. Torcka 2011a), Genderspezifika (vgl. Matthies/Matthäus 2010), unterschiedliche Handlungsmodi (vgl. Gülker/Simon/Torka 2012) und institutionelle Folgen dieser Forschungsevaluationen (vgl. Gülker/Simon 2011, Torcka 2011b) behandelt.

6 Das SEP bezieht sich prinzipiell auf alle inner- und außeruniversitären Forschungseinrichtungen des Landes. Für diese Untersuchung wurden ausschließlich Institute der Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) ausgewählt. Diese außeruniversitären Einrichtungen sind mit denen der WGL vergleichbar. Mit der Evaluation wird auch die weitere Förderfähigkeit der Institute im Rahmen der KNAW überprüft.

7 Vgl. auch andere Untersuchungen zum Thema: Lamont (2009), Langfeld (2009), Travis/Collins (1991).

views als auch im Vergleich mit anderen geprüft werden. Schließlich zielten die Interviews weniger auf die Haltungen von Evaluierenden und Evaluierten ab, sondern auf möglichst detaillierte Beschreibungen von konkreten Prozessen. Mit diesem Vorgehen haben wir einzelne Begutachtungsprozesse möglichst umfassend erhoben: Nach explorativen Experteninterviews zur Entwicklung des Erhebungskonzeptes wurden in der Regel⁸ zeitnah vor der Begutachtung Interviews mit Verantwortlichen der Institute geführt, nach Abschluss der Begutachtung dann mit Gutachter(inne)n und nach Veröffentlichung der Ergebnisse noch einmal mit den Verantwortlichen der Institute. Im Falle der RAE allerdings erstreckt sich die Begutachtung auf ein gesamtes Jahr und unterliegt höchsten Vertraulichkeitsregeln, Gespräche mit Gutachter(inne)n konnten hier also erst mit Veröffentlichung der Ergebnisse zu Beginn 2009 geführt werden.

Inhalt der Gespräche war der gesamte Prozess einer Evaluation, von der ersten Ansprache über die Vorbereitung, die Durchführung bis hin zur Ergebnisformulierung und -rezeption. Der narrative Charakter der Interviews ermöglichte eine Thematisierung aller potenziell bedeutsamen Faktoren in den jeweiligen Phasen. Um eine möglichst präzise Rekonstruktion zu ermöglichen, wurden in den Instituten Gespräche mit Vertreter(inne)n unterschiedlicher Hierarchiestufen geführt.

Tabelle 1 illustriert, dass für jeweils drei Fälle pro Verfahrenskontext eine umfassende Erhebung in diesem Sinne möglich war. Die im Vergleich höhere Anzahl an Interviews im deutschen Kontext geht auf forschungspraktische Gründe zurück – die Erhebung hat in Deutschland begonnen und im Verlauf wurde eine Reduktion erforderlich.

Grundlage für den vorliegenden Beitrag sind sowohl Interviews mit Verantwortlichen der Institute als auch mit Gutachter(inne)n sowie die formalen Verfahrensvorgaben. Dies entspricht der Ausgangsannahme, dass die Antizipation der Verfahren sowohl der Gutachter(innen) als auch der Begutachteten vom Spannungsverhältnis zwischen kollegialem Feedback und Kontrolle geprägt ist. Aus Darstellungsgründen präsentieren wir hier Interviewsegmente, in denen dieses Spannungsverhältnis besonders deutlich zum Ausdruck kommt. Wenn wir im Folgenden die Thematisierung des kollegialen Feedbacks in den Vordergrund stellen, dann bedeutet das keinesfalls, dass Evaluierende und Evaluierte nicht auch politisch-strategisch handeln würden. In unseren Interviews berichten sie sogar freimütig, dass auf die Besetzung von Gutachterteams Einfluss ausgeübt wird, Evaluierende bereits im Vorfeld von der evaluierenden Organisation über erwünschte Ergebnisse instruiert würden oder Gutachter(innen) ihre eigenen Forschungslinien promoten. Institutionelle Evaluationen als *prima facie* wissenschaftspolitischer Handlungsräume lassen sich unserer These nach aber weder auf politisch-strategisches noch wissenschaftlich-kollegiales Handeln reduzieren, sondern sie konstituieren ein unauflösbares Spannungsverhältnis zwischen beiden.

⁸ Für einzelne Fälle der WGL musste der Vorbereitungsprozess in einem Interview rekonstruiert werden, das erst nach der Begehung stattfinden konnte.

Tabelle 1: Übersicht der im Gesamtprojekt geführten Interviews

Codename Institut	Fachgruppe	Interviews mit/im				gesamt
		Institut vorher	Gutach- ter(inne)n	Institut nachher	Koordi- nation	
Verfahren der WGL in Deutschland						
Alpha-D	Geisteswissenschaften	2	2	2	1	7
Beta-D	Natur- und Sozialwissenschaften	7	5	7	1	20
Gamma-D	Natur- und Sozialwissenschaften	7	5	2	1	15
Delta-D	Natur- und Ingenieurwissen- schaften	2	6	2	1	11
Epsilon-D	Naturwissenschaften	1	4	1	1	7
Verfahren nach dem SEP in den Niederlanden						
Alpha-NL	Geisteswissenschaften	2	5	2	1	10
Beta-NL	Sozialwissenschaften	2	2	2	1	7
Gamma-NL	Naturwissenschaften	2	4	1	1	8
Verfahren nach der RAE in Großbritannien						
Alpha-UK	Geisteswissenschaften	2		1		3
Beta-UK	Natur- und Sozialwissenschaften	3	3	1	1	8
Gamma-UK	Natur- und Ingenieurwissen- schaften	3	3	1	1	8
Epsilon-UK	Natur- und Sozialwissenschaften	2	3	1	1	7
Fallbezogene Interviews gesamt		35	42	23	11	111
Explorative Expertengespräche						13
Interviews gesamt						124

3. Kollegiales Feedback und externe Kontrolle – die mehrdeutigen Rahmungen von Evaluationen

Die hier untersuchten Evaluationsverfahren kombinieren Peer-Review mit einem wissenschaftspolitisch gesetzten Verfahrensrahmen. Diese Kombination führt zu Spannungsverhältnissen, die auf der Mikroebene austariert werden müssen – so der Ausgangspunkt unserer Analyse.

Peer-Review, also die Bewertung wissenschaftlicher Arbeit durch die Fachkolleg(inn)en, ist ein in der Wissenschaftsgemeinschaft weithin etabliertes und akzeptiertes Verfahren.⁹ Sie entspricht der Logik wissenschaftlicher Entwicklung, wie sie beispielsweise Popper (1973 [1934]) beschrieben hat: Erkenntnis entwickelt sich erst in der ständigen kritischen Prüfung durch die Fachgemeinschaft, eine Theorie ist immer Aufforderung zu ihrer Falsifizierung, der zugleich gewünschte und abzuwehrende Widerspruch sichert die Dynamik im Erkenntnisprozess. Merton (1990 [1942]) hat dies in seiner Formel vom organisierten Skeptizismus normativ gefasst, an dem sich die Wissenschaftsgemeinschaft orientieren soll. Kritik ist also nicht nur gewünscht, sondern auch notwendig. Allerdings sind an die Akzeptanz der Kritik Bedingungen geknüpft: Sie muss dafür geeignet sein, den Erkenntnisfortschritt zu unterstützen. Voraussetzung dafür ist zunächst, dass sie fachlich fundiert ist, also auf dem jeweils aktuellen Forschungsstand aufsetzen kann. Entsprechend kommen nur Fachkolleg(inn)en für eine akzeptable Kritik in Frage. Außerdem sind die Kritiker(innen) gefordert, sich intensiv mit dem jeweiligen Einzelfall der Forschungsarbeit auseinanderzusetzen, sie müssen Argumente und Analyse inhaltlich nachvollziehen. Mit der in diesem Sinne kollegialen Kritik ist auch ein spezifisches Rollenverständnis der beteiligten Wissenschaftler(innen) verbunden. Die Rückmeldung kommt ‚auf Augenhöhe‘, der Gutachter/die Gutachterin hat nicht qua Status eine Autorität. Vielmehr kann, wer heute Gutachter(in) ist, morgen selbst begutachtet werden (vgl. Neidhardt 1988). Kritik wird angenommen, wenn sie inhaltlich überzeugend ist. Der Autor/die Autorin geht mit der Kritik um, kann sie nicht ignorieren, muss sie aber auch nicht teilen. Diese Bedingungen stehen nun in einem Spannungsverhältnis zu Evaluation als wissenschaftspolitischem Instrument. Ein wesentliches Ziel der Verfahren aus dieser Perspektive ist es, Informationen für Entscheidungen zur Ressourcenallokation zu liefern. Damit haben wir es – abhängig vom Verfahrensrahmen – mit unterschiedlich stringenten Formen von Wettbewerben zu tun: Mehrere Anbieter wissenschaftlicher Leistung konkurrieren um knappe Ressourcen der öffentlichen Hand. Evaluation bedeutet dann nicht allein kollegialer Skeptizismus im Sinne des Erkenntnisfortschritts, sondern für die Bewerteten geht es darum, eigene Ressourcen und möglicherweise die Existenz eines Institutes zu sichern, für die Bewertenden darum, keine Ressourcen für schlechte Leistung zu verschwenden.

9 Die breite Akzeptanz als ‚bestmögliche‘ Form der Bewertung besteht trotz zahlreicher kritischer Untersuchungen zum Peer Review, etwa Cole/Cole (1981), Hirschauer (2004), Hirschauer (2006), Merton/Zuckerman (1971).

Dieser Handlungsrahmen braucht Logiken von Kontrolle, die mit den Logiken von kollegialem Feedback in Konflikt stehen können.

In den von uns untersuchten Evaluationsverfahren treffen nun beide Handlungsrahmen aufeinander. Tabelle 2 illustriert die Spannungsverhältnisse, die damit verbunden sind. Nimmt man jeweils (in Reinform so nie vorfindbare) Idealtypen an, dann ist Ziel einer *externen Kontrolle* eine möglichst eindeutige Bewertung vergangener Leistung, die für Dritte (hier wissenschaftspolitische Entscheidungsträger(innen)) transparent und mit der Leistung anderer vergleichbar sein muss und die deshalb eindeutige Kriterien braucht. Die Begutachteten sehen sich der Sanktionsdrohung ausgesetzt und verhalten sich deshalb strategisch, beschönigen Leistungen, vertuschen Probleme und schlechte Leistung oder versuchen, auf die Auswahl von Gutachter(inne)n Einfluss zu nehmen. Die Begutachtenden auf der anderen Seite verhalten sich misstrauisch, weil sie das strategische Verhalten aufzudecken suchen. Diesen Zielen und Haltungen entspricht ein Prozess, der jede persönliche Beeinflussung zu unterbinden sucht, Interaktionen zwischen Gutachter(inne)n und Begutachteten weitgehend untersagt und auch die Leistung entpersonalisiert, also zum Beispiel Autor(inn)en anonym hält. Das Ergebnis einer idealtypischen Kontrolle ist standardisiert, auch im Vergleich zwischen Anbietern leicht kommunizierbar und mit Signalwirkung. Eine Kontrolle endet eindeutig, hat bindende Wirkung und ist nicht diskutierbar.

Tabelle 2: Idealtypische Merkmale von Prozessen externer Kontrolle und kollegialem Feedback

Merkmal	Idealtyp ‚externe Kontrolle‘	Idealtyp ‚kollegiales Feedback‘
Ziele	Bewertung vergangener Leistung, Information an Dritte	Verbesserung laufender Arbeit, Information für Begutachtete
Charakter der Kriterien	Kriterien für viele Fälle	fallspezifische Kriterienentwicklung
(rationale) Haltung der Begutachteten	strategisch, Verdecken von potenziellen Problemen	offen, Probleme werden kommuniziert, Ratsuche
(rationale) Haltung der Gutachter(innen)	misstrauisch, Suche nach Fehlern und Täuschungen	konstruktiv, auf der Suche nach Lösungen
Prozess	Entpersonalisierung der Leistung, Vermeidung persönlicher Beeinflussung	direkte Interaktion zur inhaltlichen Verständigung
Ergebnis	standardisiert, mit Signalwirkung für Dritte	fallspezifisch, nutzbar nur/vor allem für Begutachtete
Konsequenzen	bindend, nicht diskutierbar	Rat, der nur angenommen wird, wenn er überzeugt

Demgegenüber ist das Ziel eines idealtypischen *kollegialen Feedbacks* die Verbesserung der laufenden Arbeit; der Inhalt des Feedbacks soll vor allem den Begutachteten selbst dienen. Nötig sind deshalb Kriterien, die die Besonderheiten des Einzelfalls berücksichtigen. Begutachtete in einer idealtypischen Situation kollegialen Feedbacks versprechen sich von den Gutachter(inne)n Rat, sie werden also gerade die Probleme kommunizieren, zu denen Lösungen am dringlichsten gefunden werden sollen. Die Gutachter(innen) teilen mit den Begutachteten das Interesse an der Verbesserung und geben deshalb konstruktives Feedback. Kollegiales Feedback braucht einen Prozess der Interaktion; erst im inhaltlichen Austausch können Probleme identifiziert und Lösungen gefunden werden. Das Ergebnis des Feedbacks ist fallspezifisch formuliert, dient den Begutachteten zur Verbesserung, ist aber nicht bindend. Angenommen wird nur der überzeugende Rat.

Die Gegenüberstellung macht deutlich, dass die beiden Handlungsrahmen ‚externe Kontrolle‘ und ‚kollegiales Feedback‘ in Spannung zueinander stehen. Beispielsweise wäre eine offene Problemmunikation seitens der Begutachteten zwar für den Rahmen des kollegialen Feedbacks, nicht aber für den Rahmen der externen Kontrolle, rational. Wenn wir weiter davon ausgehen, dass die von uns untersuchten Evaluationsverfahren Anteile beider Handlungsrahmen miteinander kombinieren, dann ist die Frage, wie die Spannungsverhältnisse auf der Mikroebene austariert werden.

Wie oben ausgeführt konzentrieren sich Untersuchungen zum Thema in der Regel auf die Kontrollfunktion von Evaluationen und vernachlässigen, dass Anteile des kollegialen Feedbacks auch bei wissenschaftspolitisch initiierten, aber von wissenschaftlichen Kolleg(inn)en spezifizierten und durchgeführten, Verfahren von Bedeutung sind. Wir fokussieren im Folgenden gerade auf diese Anteile und fragen, wie kollegiales Feedback unter Bedingungen von gleichzeitiger externer Kontrolle möglich ist. Dafür überprüfen wir zunächst die jeweils sehr unterschiedlichen Verfahrensregeln der WGL und SEP auf der einen Seite und RAE auf der anderen Seite daraufhin, welche Spielräume sie für kollegiales Feedback vorsehen. Im zweiten Schritt untersuchen wir jeweils für die Verfahrenskontexte die Interpretationen der Gutachter(innen) sowie der Begutachteten.

4. Verfahrensregeln als Rahmen für Kontrolle und Feedback

Das Spannungsverhältnis zwischen kollegialem Feedback und externer Kontrolle ist kennzeichnend für jede Evaluation, an der in irgendeiner Weise wissenschaftliche Peers beteiligt sind.¹⁰ Allerdings, so unsere Ausgangshypothese, setzen Verfahrensregeln den Rahmen dafür, wie dieses Spannungsverhältnis von den Beteiligten austariert werden kann. In unserer Analyse haben wir es mit zwei prinzipiell unterschiedlichen Typen von Evaluationen zu tun.

10 Zu Verfahren ohne Gutachter(innen) vgl. Gläser/Laudel (2007) und Espeland/Sauder (2007).

Bei den Verfahren der Leibniz-Gemeinschaft und nach dem SEP steht eine Begehung der Einrichtung im Zentrum, vorbereitet durch eine teilstandardisierte Selbstberichterstattung der Einrichtung. Ein Evaluationsverfahren befasst sich mit einem Institut, dafür wird eine spezialisierte Gutachterkommission eingesetzt. Im Ergebnis schreibt diese Kommission einen Bericht, der sowohl zur Qualitätsverbesserung in der Einrichtung beitragen soll als auch als Grundlage für weitere Förderentscheidung dient. Das Institut hat die Möglichkeit, zu diesem Bericht schriftlich Stellung zu nehmen (vgl. Tabelle 3).

Tabelle 3: Verfahrensmerkmale im Überblick

Verfahrensmerkmal	WGL/ SEP	RAE
Form der Evaluation	Interaktive Begehung	Gutachterkommissionen bewerten in Abwesenheit
Evaluationsgegenstand	Gesamte Institute	Von den Einrichtungen definierte Forschungseinheiten
Ziele und Folgen der Evaluation	a) Qualitätssicherung b) Entscheidungen über Fortführung der Förderung	a) Monitoring und Qualitätsbewertung (Controlling) b) Budgetentscheidungen
Informationsbasis der Evaluation	Teilstandardisierte Selbstevaluationsberichte der Institute	a) 4 Publikationen pro „forschungsaktivem Wissenschaftler/forschungsaktiver Wissenschaftlerin“ b) Selbstdarstellungen der Forschungsumgebung und Stellung des Instituts
Evaluationskriterien	Einheitliche und öffentlich zugängliche Kriterien bezogen auf Gesamtkonzept und Profil, Kooperation und Vernetzung, Personal- und Nachwuchsförderung, Qualitätssicherung	Einheitliche und öffentlich zugängliche Kriterien bezogen auf Originalität, Bedeutung und Präzision im Vergleich zu anderen – von national anerkannt bis weltweit führend
Evaluationsergebnisse	Bericht mit Empfehlungen an das Institut, formalisierter notenähnlicher Satz zur Gesamtbewertung, wissenschaftspolitische Empfehlung zur Förderung	Note zwischen 1* und 4* pro Publikation, aggregiertes Notenprofil für die Forschungseinheit als Basis zur Budgetberechnung
Reaktionsmöglichkeiten für die Institute	Institute können einen Kommentar zum Bericht schreiben, der veröffentlicht wird	Keine formale Reaktionsmöglichkeit

Das Verfahren der RAE basiert im Kern auf der Begutachtung von Publikationen. Zusätzlich werden Dokumente zu „esteem“ und „environment“ bewertet.¹¹ Die RAE ist mit einer landesweiten Kampagne vergleichbar, im Laufe eines Jahres werden alle Institute zeitgleich von (2008 insgesamt 67) Fachkommissionen bewertet. Die Inhalte der Beratungen werden streng vertraulich behandelt und im Ergebnis wird ein Notenprofil von „unclassified“ über 1* (national interessanter Beitrag) bis 4* (weltweit bahnbrechender Beitrag) erstellt, das am Ende unmittelbar in eine Formel zur Budgeterrechnung übertragen wird. Die Rückmeldung an die Institute beschränkt sich auf dieses Profil mit einer knappen, sehr allgemein gehaltenen Begründung. Für die Institute ist keine Stellungnahme dazu im Verfahren vorgesehen.

In diesem Überblick wird deutlich, dass die Verfahren der Leibniz-Gemeinschaft und nach dem SEP für kollegiales Feedback prinzipiell mehr Raum lassen als die RAE. Allein die direkte Interaktion zwischen Gutachter(inne)n und Begutachteten schafft überhaupt die Möglichkeit dazu. Im Bewertungsbericht wird fallspezifisches Feedback gegeben und es werden offene Empfehlungen formuliert, so wie es im Idealtyp kollegialen Feedbacks zu erwarten wäre. Gleichwohl geht es auch hier darum zu kontrollieren, ob und in welcher Höhe weiterhin eine Förderung (im Rahmen der Leibniz-Gemeinschaft bzw. der Niederländische Akademie der Wissenschaften) erfolgen soll. Die Handlungslogik der Kontrolle finden wir entsprechend auch in den Verfahrensregeln, wenn etwa die Selbstdarstellung der Institute nach standardisierten Kriterien zu erfolgen hat. Auch wird die Bewertung jeder Forschungseinheit im Bericht mit einem notenähnlichen Satz zusammengefasst, die Leistung ist beispielsweise „exzellent“, „sehr gut“ oder „gut“.

Das Verfahren der RAE betont generell deutlich stärker die Kontrollaspekte. Diese Begutachtung ohne direkten Kontakt, an deren Ende ein numerisches Profil steht, lässt für kollegiales Feedback wenige Möglichkeiten. Allerdings handelt es sich auch hier um ein Peer-Review-Verfahren und einzelne Verfahrensdetails entsprechen eher der idealtypischen Logik kollegialen Feedbacks denn reiner Kontrolle. Mit den Verfahrensregeln von 2008 wurde beispielsweise die Fallspezifität erhöht, indem erstmals auch die Bewertung der Forschungsumgebung und -strategie mit in die Gesamtformel eingegangen ist. Zudem wird die prozentuale Gewichtung der Leistungsbestandteile in den Sub-Panels entschieden, variiert also je nach Fächergruppe – auch dies ein Bemühen, den spezifischen Bedingungen der jeweiligen Einheiten besser gerecht zu werden.

Im Folgenden soll geprüft werden, wie die Beteiligten – Gutachter(innen) und Begutachtete – die Situation innerhalb dieser unterschiedlich gesetzten Rahmenbedingungen interpretieren.

11 Aktuell wird unter dem Titel *Research Excellence Framework* (REF) ein neues Bewertungsverfahren vorbereitet, das 2014 umgesetzt wird. Nachdem zwischenzeitlich auch eine Ablösung des Peer-Review-Prinzips durch rein numerische Verfahren in der Diskussion war, bleiben nun wesentliche Bestandteile der RAE auch im REF erhalten. Die Anzahl der Bewertungsgremien wurde reduziert und der Impact der Forschungsleistung wird zusätzlich bewertet (vgl. HEFCE 2011).

5. Interpretation der Evaluationen in der Leibniz-Gemeinschaft und im Standard Evaluation Protocol durch Evaluierende und Evaluerte

5.1 Die Gutachterperspektive

Wie wir gesehen haben, vereinen beide Evaluationsverfahren heterogene Ziele und prägen darüber das Handeln von Gutachtergruppen bei der Begehung. Sie sollen im Auftrag der Wissenschaftspolitik kontrollieren, ob die Leistungen von Instituten eine zukünftige Weiterfinanzierung rechtfertigen. Zugleich sollen sie den Instituten aber auch inhaltliche Empfehlungen geben, ob und wie sie sich verbessern könnten. Mit Rollen- und Zielkonflikten zwischen wissenschaftspolitischer Kontroll- und kollegialer Feedbackfunktion ist deshalb von vornherein zu rechnen (vgl. Schimank 2004).¹²

Dabei lassen die Verfahren offen, wie Gutachter(innen) die beiden Hauptziele „accountability“ and „improvement“ (KNAW 2009: 4) erreichen sollen. Zwar soll ein „standardisierter Kriterienkatalog“ die fachliche Bewertung orientieren, aber die „Auswahl, Gewichtung und ggf. Ergänzung der Kriterien [...] erfolgt durch die Fachgutachter“ (Leibniz-Gemeinschaft 2007). Die Kriterienkataloge umfassen zudem breite Deutungsdimensionen (z.B. Konsistenz, Produktivität, Effektivität von Forschungsprogrammen), die stets fallspezifisch interpretiert werden müssen. Die Gutachterteams verfügen somit über eine große Autonomie bei der Entscheidungsfindung, die von den formalen Verfahrensregeln kaum beschränkt wird. Deshalb muss man analysieren, wie Gutachter(innen) ihre Tätigkeit und Rolle als Evaluierende interpretieren und herausarbeiten, welchen „customary rules“ (vgl. Lamont 2009) sie verpflichtet sind. Im Folgenden werden drei auffällige Interpretationen näher beleuchtet, die in allen Gutachterinterviews regelmäßig auftauchen und in Spannung zur externen Kontrollfunktion stehen: Demnach sind Evaluationen (1) eine Arbeit an und für die Wissenschaftsgemeinschaft, sie bedürfen einer (2) bestimmten Form der Kritik, die auf die (3) Offenlegung, Diskussion und Bearbeitung von Problemen zielt.

In Evaluationen ist die Gutachterrolle undeutlich. Diese handeln zugleich im Auftrag wissenschaftspolitischer Organisationen und als wissenschaftliche Kolleg(innen). Fragt man aber Gutachter(innen), warum sie dennoch an Evaluationen teilnehmen, dann präsentieren sie ein spezifisches Rollenkonzept. Die wissenschaftlichen Organisationen und Gemeinschaften werden als primäre Adressaten ihrer Tätigkeit angesehen. Evaluationen sind „eine Dienstleistung, die wir sowieso auch erbringen sollten, weil das ja zur Tätigkeit von Wissenschaft oder von Professoren

12 Darüber hinaus kommt es vor und wird berichtet, dass auch seitens der Politik ein nicht allein kontrollierendes Interesse artikuliert wird – vielmehr kann auch der Erhalt oder die Stärkung eines spezifischen Instituts politisch gewünscht sein. Auch hier gilt es für die Gutachter(innen), mit diversen Selbst- und Fremdansprüchen umzugehen.

gehört [...] hier im Dienste der Wissenschaft etwas Positives zu leisten“ (Nunziger, 5-9, Gamma-D).¹³ Die Motive dafür, als Gutachter(in) tätig zu werden, sind natürlich vielfältig und nicht frei von Selbstzweck: Sie fühlen sich geehrt und anerkannt, entdecken darin einen Reputationsnachweis, sie interessieren sich für das Innenleben der Einrichtungen und die Einschätzungen anderer Gutachter(innen), sie wollen gegebenenfalls auch intervenieren und bevor Evaluationen unsachgemäß betrieben werden, steht man selbst in der Pflicht. Interessant an der Art der Antworten ist aber, dass selbst wissenschaftspolitisch initiierte institutionelle Evaluationen in wissenschaftsspezifische Motivlagen eingebettet und als Teil der professionellen Selbstkontrolle gedeutet werden. Diese in einem durch wissenschaftspolitische Ressourcen- und Verteilungsfragen geprägten Handlungsraum aufrechtzuerhalten, ist eine schwierige und keineswegs immer gelingende Herausforderung für die Gutachter(innen).

Ein wesentlicher Bestandteil der wissenschaftlichen Selbstkontrolle ist Kritik, die im Kontext von Evaluationen allerdings sehr folgenreich sein kann. Deshalb ist es für die Gutachter(innen) sehr wichtig, auf welche Weise und zu welchem primären Zweck Kritik im Kontext von Evaluationen geäußert werden soll. Die starke Ausrichtung an der scientific community legt zwei Formen Kritik zu üben nahe, die unter den Gutachter(inne)n aber beide als unangemessen gelten. Die in der Wissenschaft an sich übliche harte Kritik in der Sache ist schwierig, weil wissenschaftspolitische Instanzen und die evaluierten Institute dies als eine Aburteilung statt als einen Ausgangspunkt zur Bearbeitung von Problemlagen auffassen könnten. Deshalb gelten allzu harsche Kritiker(innen) oft als schlechte Evaluationskolleg(inn)en oder werden Evaluationsberichte trotz harter Kritik in einem lobenden, auf Verbesserungen hinweisenden Duktus verfasst.¹⁴ Aber auch das andere Extrem, nur Gefälligkeitsgutachten abzugeben und Evaluationen auf ein bloßes akademisches Ritual (vgl. Michaels 2010) zu reduzieren, entspricht nicht dem Selbstverständnis von Gutachter(inne)n, auch wenn sie vielleicht zu diesem Zweck von evaluierten Instituten vorgeschlagen wurden. Kritik dient ihnen als Mittel der Verbesserung und muss deshalb ausgewogen formuliert werden: „kritisch und sachlich, nicht zu freundlich gegenüber der Institutsleitung, auch nicht zu gehässig, also war völlig sachgerecht“ (Grimmer, 596-599, Epsilon-D). Wohlgesonnene Zuwendung und kritische Distanznahme bilden zwei Seiten eines Spannungsverhältnisses, das die Gutachter(innen) austarieren müssen (vgl. Neidhardt 2010), wenn sie ihrem (teils explizit geäußerten) Selbstverständnis als kollegiale Berater nachkommen. Denn

13 Die Quellennachweise sind wie folgt zu lesen: Pseudonym der Person, Zeilennummern im Transkript, evaluiertes Institut. Wenn die Gutachterperspektive thematisch ist, kommen ausschließlich Gutachter aus den jeweiligen Verfahren zu Wort, bei der Analyse der Institutsperspektive ausschließlich Mitarbeiter(innen) aus den evaluierten Instituten.

14 Es gibt allerdings Extremfälle drohender oder vollzogener Schließung. An den Bewertungsberichten solcher Fälle ist auffällig, dass selbst bei einem vernichtenden Urteil stets Verbesserungsvorschläge unterbreitet werden, auch wenn letztlich eine Schließung empfohlen wird. Diese erfolgt insbesondere dann, wenn Institute Empfehlungen aus früheren Evaluationen nicht nachkommen und sich damit als beratungsresistent erweisen.

während der Begehung „we are not sitting there to cause difficulties, to look for all that’s wrong“ (Umbreit, 226-227, Gamma-NL) und ihr Ziel „was not as much about finding faults but are they going in the right direction? Are there points of improvement? Can things be done in a better way?“ (Notter, 20-22, Beta-NL).

Um ihre Tätigkeit als kollegiale Berater(innen) wahrzunehmen, müssen Gutachter(innen) nach Problemlagen suchen, für die dann Verbesserungsvorschläge erarbeitet werden können. Allerdings stellt diese Suche im Kontext von potenziell folgenreichen Evaluationen eine besondere Herausforderung dar. Denn in den Dokumenten, Präsentationen und Interaktionen des Evaluationsverfahrens kehren die evaluierten Institute ihre Leistungen und nicht ihre Probleme hervor. Evaluierende sind in einer Situation der Kontrolle darauf vorbereitet, nur das „best self“ präsentiert zu bekommen und hinter die Fassade von Instituten gelangen zu müssen, um herauszufinden, „wie dort das Alltagsgeschäft läuft, ob es wirklich alles so perfekt ist oder nur dafür organisiert“ (Ehlert, 115-117, Epsilon-D). Da die offene Kommunikation von Problemen in Evaluationen nicht erwartet werden kann, fehlt den Gutachter(inne)n eine wesentliche Vorbedingung, um beratend tätig werden zu können. Der Evaluationsprozess besteht deshalb zu weiten Teilen aus der mühsamen Suche nach möglichen Problemlagen, für die dann Lösungen angeboten werden können.

Diese beginnt bereits bei der Durchsicht der eingereichten Akten. Die heterogenen Informationen werden auf ihre Konsistenz und Passung zu sehr allgemeinen Normalitätserwartungen im jeweiligen wissenschaftlichen Feld (z.B. Promotionszeiten, Publikationsarten und -mengen) geprüft. Auffälligkeiten und Widerstände deuten dann auf mögliche Problemlagen hin, denen zunächst im Gutachterteam und dann in der direkten Interaktion mit den evaluierten Institutsangehörigen nachgegangen wird. Erst beim Zusammentreffen der Gutachter(innen) verdichten sich subjektive Einzeleindrücke zu einer intersubjektiv geteilten Problemsicht der Gutachtergruppe. Dazu müssen sich hervorgebrachte Einwände im Gutachterteam bewähren, indem sie von anderen Gutachter(inne)n aufgegriffen werden und von den evaluierten Institutsangehörigen nicht ausgeräumt werden können. Die direkte Interaktion mit diesen bietet die Möglichkeit, die Problemsicht des Gutachterteams zu überprüfen oder auf weitere Problemlagen zu stoßen, die nicht unmittelbar aus den Akten hervorgehen. Die genaue Beobachtung der Reaktion (vor allem der Leitung) auf Fragen liefert ihnen Hinweise, ob Probleme vorliegen, reflektiert und bearbeitet werden: „Also das ist eben die Art, antworten die Leute auf Fragen, die man ihnen stellt, auch auf kritische Fragen und wie gehen sie mit diesen Fragen um? Weichen sie denen aus, beantworten sie die gar nicht, beantworten sie die klar? Wenn sie die klar beantworten und auch ein Problem eingestehen, ist das im Prinzip schon mal ein Indiz, dass es in die richtige Richtung geht. Und wenn sie ein realistisches Selbstbild auch haben, wie sie sich selber einschätzen, ist das auch ein positives Indiz. Also das sind Indizien, die was mit Glaubwürdigkeit zu tun haben“ (Fissler, 861-868, Gamma-D).

Wenn Gutachter(innen) das Eingestehen von Problemen aber generell positiv honorieren, dann stößt man auf eine spezifische Interpretation ihrer Handlungssituation und Rolle. Gutachter(innen) deuten Evaluation als eine nach Verbesserungs-

möglichkeiten suchende, kollegiale Beratung oder Supervision. Wie wir gesehen haben, ist diese Selbstbeschreibung allerdings im Kontext immer auch kontrollierender und mit wissenschaftspolitischen Interessen durchzogener Evaluationen mit erheblichen Durchsetzungsproblemen behaftet. Denn spätestens am Ende steht ein öffentlich einsehbares, potenziell folgenreiches und mit ‚Empfehlungen‘ versehenes Urteil, deren Einlösung spätestens bei der nächsten Evaluation überprüft wird. Unter diesen Bedingungen kann die Rolle als kollegialer Berater nicht stabil eingerichtet werden: „So it’s a combination, because advising is too, too optional, as if you can do with my advice what you want, you don’t have to“ (Notter, 117-119, Beta-NL).

Am Beispiel von Evaluierenden aus dem SEP- und dem WGL-Verfahren konnte man sehen, wie voraussetzungsreich es im Rahmen von unfreiwilligen, öffentlich einsehbaren und folgenreichen institutionellen Evaluationen dieses Typs ist, beratend tätig zu werden. Umso überraschender ist es, dass kollegiale Beratung dennoch ein wichtiges Orientierungsmuster für Evaluierende und, wie wir gleich sehen werden, auch für evaluierte Institute ist.

5.2 Die Institutsperspektive

An den bewerteten Instituten werden Evaluationen in eben dem Spannungsverhältnis antizipiert, wie es eingangs dargestellt wurde. Einerseits finden sich deutlich strategische Aspekte in der Vorbereitung und Durchführung von Begehungen. Sie können bis zur Anpassung an vorgegebene Kriterien gehen, wenn beispielsweise vornehmlich Artikel in international reputierten refereed journals veröffentlicht oder Drittmittel von besonders anerkannten Förderorganisationen eingeworben werden. Andererseits werden Rückmeldungen im Sinne kollegialer Hinweise auf Augenhöhe zugleich selbstbewusst und kritisch reflektiert.

Tatsächlich wird den Gutachter(inne)n die eigene Einrichtung vielfach als eine strategisch agierende präsentiert, die also in der Lage ist, Fokussierungen vorzunehmen, Schwerpunkte zu setzen, Kohärenz und Synergien zu entwickeln und Potenziale ihrer zukünftigen Entwicklung zu identifizieren. Die Institutsakteure nehmen zum Teil Einfluss auf die Gutachter(innen)wahl, versuchen im Vorfeld mit Gutachter(inne)n in Kontakt zu treten, oder versuchen zumindest, angenommene Gütekriterien und Bewertungsmaßstäbe der Evaluierenden zu bedienen und das Institut entsprechend zu präsentieren, damit sie in der Evaluation gut abschneiden. Dies bedeutet auch, Stärken herauszustellen und Schwächen möglichst zu verdecken oder so darzustellen, dass die Problembehandlung gleich mitgeliefert wird. Als strategisches Handeln im Kontext einer Bewertungssituation lassen sich auch die vielfachen Anstrengungen interpretieren, die darauf ausgerichtet sind, den Gutachter(inne)n einen angenehmen Aufenthalt zu organisieren, so zum Beispiel: „Und beim Zeitmanagement kommen halt jetzt diese vielen Kleinigkeiten, noch mal die Gutachter kontaktieren, wie sie anreisen, sicherstellen, dass sie auch wirklich bequem und pünktlich vom Bahnhof oder vom Flughafen zu uns finden usw. Also das

macht sehr viel Arbeit, aber wie gesagt, da muss man halt nur an alles denken wie bei der Vorbereitung einer Hochzeit [lacht], damit dann auch alles klappt“ (Dittmer, 401-406, Gamma-D).

Aber – und dies ist der eigentlich bemerkenswerte Befund – die Auseinandersetzung der Institute bleibt nicht auf der strategischen (Selbstdarstellungs-)Ebene stehen. Vielmehr werden die Evaluationen für interne Verständigungsprozesse als Anstoß zur Überprüfung eigener Routinen und zur Entwicklung von Zukunftsszenarien genutzt. Dies sind Handlungsmuster und -orientierungen, die typisch sind für eine Motivlage der Institutsakteure, die von der Evaluation Empfehlungen zur inhaltlichen und strukturellen Weiterentwicklung erwarten. Die Hinweise hierzu können direkt von den Gutachter(inne)n kommen, aber zumeist werden sie im Vorfeld einer Selbstevaluation gewonnen. Bei einer rein extern erzwungenen Kontrolle könnte die Auseinandersetzung auf der strategisch-taktischen Ebene verbleiben, es ginge dann ausschließlich darum, eine gute Note zu erreichen.

Für Selbstverständigungsprozesse auf Institutsebene werden beispielsweise die Diskussionen des Selbstdarstellungsberichts genutzt: „Das war insgesamt spannend und hat auch ich denke ne Menge Selbsterkenntnisse für das Institut gebracht“ (Müller, 130–131, Beta-D). Und: „I think it’s important to have the opportunity to show what you have reached ... because it’s only one in every five years and the daily work comes first“ (Vannata, 261-264, Beta-NL). Wichtig erscheint vor allem der (kritische) Blick auf die Gesamtorganisation, der im Alltagsgeschäft der Forschungsabteilungen nicht die entsprechende Aufmerksamkeit erreicht. Es ist ein „Vorteil von solchen Evaluierungen, dass die Evaluierung ja das macht, was eigentlich die Institutsleitung machen müsste, aber da versäumt hat, nämlich auf das ganze Haus zu schauen und von außen, nicht so dieser betriebsblinde Innenblick ...“ (Jürgens, 447-450, Alpha-D). Die Ausrichtung des Verfahrens auf die Gesamtorganisation und die breit angelegten Bewertungskriterien tragen demnach zu einer Gesamtreflexion bei. Die dominierenden Indikatoren für gute Wissenschaft, etwa Publikationen in refereed journals und Drittmittelinwerbung mit Begutachtungsverfahren, haben zwar ebenfalls einen hohen Stellenwert, stehen aber nicht allein im Fokus der Vorbereitung. Vielmehr werden auch Kriterien wie thematische Zusammenhänge und Kohärenz des Instituts, Wissenstransfer, Nachwuchsförderung und anderes gewichtet. Institute müssen ein einzigartiges Profil ausbilden und nachweisen, um zum Beispiel (im Fall der Leibniz-Gemeinschaft) ihre Existenz als außer-universitäres Institut neben den Hochschulen zu legitimieren. In diesem Sinne betonen die Institutsakteure die fallspezifische, nicht bloß an vorgegebenen Kriterien ausgerichtete Vorbereitung auf die Evaluation.

Die ‚Begehung‘ wird vielfach mehr als ein Gespräch unter Kolleg(inn)en als ein inquisitorischer Test wahrgenommen. Dies lässt gleichzeitig keinerlei Rückschlüsse auf ein naives Verständnis der Bewertungssituation zu; der Tragweite und Bedeutung sind sich die Evaluierten durchaus bewusst. Das Gespräch mit den Gutachter(inne)n wird in den Instituten einerseits geprobt, deren Wahrnehmungen und möglichen Urteile in der Selbstdarstellung einkalkuliert. Andererseits werden die Gutachter(innen) dafür geschätzt, dass es ihnen gelingt, „[...] sowohl Stärken als

auch Schwachstellen zu identifizieren [...] wo es nicht einfach nur um Publikationszahlen oder eingeworbene Drittmittel geht“ (Dagendorf, 27-28, Gamma-D).

Seitens der Institute wird also eine Auseinandersetzung mit dem Institut als Ganzes positiv wahrgenommen, standardisierte auditähnliche Bewertung dagegen ist negativ besetzt.

Sowohl Vorbereitung als auch Ergebnis der Evaluation werden in den Instituten ernsthaft reflektiert und zur Überprüfung der eigenen Handlungsrountinen genutzt. In dem Zusammenhang sind insbesondere die zum Teil ausführlichen Rückmeldungen der Gutachter(innen) zur künftigen Entwicklung der Institute von Bedeutung. Berichtsaussagen spielen eine wesentliche, vielfach unterstützende Rolle bei internen Veränderungsprozessen: „I think what, what we have done is that we take everything that they say, we take it serious, but we feel that it helps us to do what we already wanted to do, so that we don't have to, to go in another direction, so that's the good thing about this whole evaluation. So they were critical, some people said well, they could have been more critical, I say they were critical, and supportive“ (Ludwig, 472-477, Alpha-NL).

Die Institute schätzen die externen Bewertungen und Handlungsempfehlungen in zweierlei Hinsicht: Als Validierungsinstanz des Status quo und als Chance, für eigene Pläne und Veränderungsabsichten, Unterstützung und damit Sicherheit, zu bekommen: „Ja, das war eine Bestätigung der Pläne sich umzuorganisieren, das findet sich jetzt durchaus drin und das ist jetzt dann auch umgesetzt“ (Ulbricht, 12, Beta-D). Die Evaluationsergebnisse, sowohl die Bewertungen als auch die Empfehlungen, werden zur Unterstützung der Institute in ihrer inhaltlichen und auch organisatorischen Fortentwicklung genutzt. Die Institute nutzen die Evaluation als Organisationsberatung, auch wenn sie nicht selbst den Beratungsauftrag ausgesprochen haben. Die Empfehlungen werden dabei ernsthaft reflektiert und auch umgesetzt, jedoch nicht eins zu eins:¹⁵ „Das sind in der Regel keine Muss-Vorgaben, sondern Kann-Vorgaben, die wir prüfen werden, wie weit sind sie umsetzbar“ (Dagendorf, 43-45, Gamma-D).

Gerade weil die Institutsvertreter(innen) Evaluierung nicht ausschließlich als externe Kontrolle interpretieren, sondern die an interne Organisationsentwicklung anschlussfähigen Aspekte hervorheben und das Feedback der Gutachter(innen) suchen und schätzen, kritisieren sie die Art und Weise, wie Evaluationen ablaufen: Sie fänden zu oft statt, es stehe zu wenig Zeit für Präsentationen und Diskussionen mit den Gutachter(inne)n zur Verfügung, die teilweise nicht hinreichend vorbereitet seien. Im Vergleich mit standardisierten Auditierungsprozessen haben sie aus Sicht der Institutsvertreter(innen) allerdings einen deutlichen Mehrwert.

15 Hier scheint ein Entwicklungsprozess gegenüber früheren Evaluationen durch den Wissenschaftsrat und die Leibniz-Gemeinschaft stattzufinden: Empfehlungen waren in den vorherigen Runden quasi sakrosankt. Es wurde beispielsweise flächendeckend eine Matrixstruktur in der Forschungsorganisation eingeführt, um interdisziplinäres Arbeiten zu fördern. Jahre später sind bei nächstbestener Gelegenheit die Strukturen zum Teil wieder zurückgeführt worden, da sie sich für die Aufgaben des Instituts als nicht zielführend erwiesen haben (vgl. Röbbcke/Simon 2001, Simon 2007).

6. Interpretation der Research Assessment Exercise durch Evaluierende und Evaluierte

Wie wir in Kapitel 4 ausgeführt haben, bieten die RAE viel weniger Möglichkeiten für kollegiale Rückmeldungen. In diesem Verfahren ist die direkte Interaktion zwischen evaluierenden und evaluierten Wissenschaftler(inne)n untersagt, so dass kein fallspezifisches Feedback gegeben werden kann. Die evaluierten Institute erfahren nur, welchen Platz sie in einem standardisierten numerischen Ranking der Wissenschaftsgebiete einnehmen. Das Ergebnis ist unmittelbar an Finanzierungsentscheidungen gekoppelt und hat einschneidende Konsequenzen für die Institute. Wissenschaftler(innen), die mit ihren Forschungsarbeiten nicht zum Erfolg in diesem Wettbewerb beigetragen haben, werden entlassen oder in die Lehre abgeschoben (vgl. Lange/Gläser 2009). Mit der RAE hat die Wissenschaftspolitik ein „strong evaluation system“ (vgl. Whitley 2007) eingerichtet, das von disziplinären Gutachterpanels betrieben wird, so dass eine kleine „epistemic elite“ (vgl. Martin/Whitley 2010) in die Arbeitsbedingungen von Kolleg(inn)en eingreifen kann. Dieses Verfahren ist auf Kontrolle und Sanktionierung, aber nicht unmittelbar auf ein kollegiales Feedback zur Verbesserung von Forschungsleistungen ausgerichtet. Deshalb haben die Evaluierenden und Evaluierten auch hier von verschiedenen Formen der Einflussnahme berichtet. Diese umfassen Versuche, direkt in Panels mitzuarbeiten, um dort spezielle Entscheidungskriterien auszubilden und anzuwenden, vermeintlich weniger streng urteilende Gutachtergruppen für die Bewertung auszuwählen oder nur hervorstechende Beiträge zur Begutachtung einzureichen.

6.1 Die Gutachterperspektive

Betrachtet man die Interpretationen und Handlungsweisen von Gutachter(inne)n in diesem Verfahren, dann messen sie dem kollegialen Feedback aber dennoch eine zentrale Bedeutung bei. Wie in den zuvor analysierten deutschen und niederländischen Verfahren fassen auch die RAE-Gutachter(innen) ihre Tätigkeit als eine Unterstützungsleistung für die Wissenschaft auf: „The first reason [for the RAE] which is probably the only valid one, really is to help universities benchmark their research against their competitors“ (Rafael, 285-287, Epsilon-UK). Sie neigen sogar dazu, ihre Evaluatorenrolle als Verpflichtung gegenüber der Wissenschaftsgemeinschaft zu deuten, obwohl es sich bei diesem Verfahren um eine starke wissenschaftspolitische Intervention handelt: „I think part of my responsibility is not only to do my job in isolation, but to do my job in the context of the community that I relate to, and it's in the interest of that community“ (Quinn, 661-664, Epsilon-UK). Die Gutachter(innen) sind davon überzeugt, dass die RAE „had strengthened the quality of the science anyway in the UK“ (Urban, 1178) und ihre Intention ist eben, diese Qualitätsentwicklung zu unterstützen, auch wenn dieses Verfahren kaum dazu geeignet erscheint.

(Wie) gelingt es also den Gutachter(inne)n, der scientific community einen Dienst zu erweisen? Ihre primäre Aufgabe sehen sie darin, eine möglichst robuste Informationsbasis für verschiedenste Entscheidungsprozesse und -träger zu erstellen, weil sie selbst keine Kontrolle darüber haben, wie ihre Ergebnisse interpretiert und genutzt werden. Dafür arbeiten Gutachtergruppen in sogenannten „calibration sessions“ geteilte Bewertungsstandards für ihr Fachgebiet aus und setzen sie ggf. auch gegenüber abweichenden Panelmitgliedern durch, „to make sure we did the job properly but that we were also very fair to the community“ (Unruh, 140-141, Gamma-UK). Mit diesen Standards durchmustern die Gutachter(innen) jeweils mehrere hundert Publikationen auf ihre Qualität im Sinne von Neuheit, Originalität oder Konsistenz zwischen Anspruch und beigebrachter Evidenz.

Die Aufbereitung der Informationsbasis soll den Institutsleitungen und Kolleg(inn)en dazu dienen, ihren Stand innerhalb ihrer Wissenschaftsfelder zu überblicken und hieraus eigene Schlussfolgerungen für ihr zukünftiges Handeln zu ziehen. Insofern liefern die Gutachter(innen) mit ihren numerischen Qualitätsprofilen ein Mittel zur vergleichenden Selbstbeobachtung von wissenschaftlichen Einrichtungen, das aber in jedem Fall einer sachadäquaten Rückübersetzung bedarf: „I just hope they can make sense of the sort of deconstructed outputs that they get back from us“ (Rafael, 881-883, Epsilon-UK). Auch aus Sicht der Gutachter(innen) in der RAE bedarf es für unmittelbare Verbesserungsmaßnahmen die genaue Kenntnis des jeweiligen Falls, die eigentlich nur durch direkte Gespräche zu erlangen ist. Auch die sehr eingeschränkten Rückmeldemöglichkeiten im Rahmen der RAE „don't really fulfill that role of constructive engagement which would lead to improvements“ (Urban, 1173, Beta-UK). Allerdings regt die RAE Verbesserungen auf indirekte Weise an und bedient sich dabei zweier Mechanismen. Der erste Mechanismus, den man als Drang und Druck zur Selbstverbesserung bezeichnen könnte, setzt unmittelbar am unterstellten Ehrgeiz und der Ehre von Personen an, „because, people who don't do so well think ‚Well let's show these buggers what's going on‘ and let's do a better, [...] ‚Well let's really improve our act and learn from this‘“ (Urban, 1179-1183, Beta-UK). Der zweite Mechanismus zielt auf die Steigerung der Reflexions- und Handlungsfähigkeit von Institutsleitungen, die bereits im Vorfeld einen genaueren Einblick in ihre Einrichtung erhalten, „because you've got to make all your people write down what it is the hell they're doing and why and then try and glue it all together in some sort of cohesive whole“ (Rafael, 881-883, Epsilon-UK). Die Ergebnisse der RAE machen dann Stärken und Schwächen der Institution sichtbar und können weitere interne Diskussionen oder Interventionen anregen.

Am Beispiel der RAE konnten wir verfolgen, dass sich Gutachter(innen) selbst dann an der Norm orientieren, ein kollegiales Feedback zu Verbesserungszwecken zu geben, wenn das Verfahren weder dafür gemacht noch wirklich geeignet ist. Verbesserungen herbeizuführen impliziert auch hier, angemessen zu intervenieren. Genau das können die Gutachter(innen) in der RAE aber nicht garantieren, da sie nur eine abstrakte Informationsbasis liefern, aber deren Verwendung nicht kontrollieren können. Die zentrale Operation, dieses Wissen auf je spezielle Institute zu

beziehen und auf diesem Weg zu angemessenen Schlussfolgerungen zu gelangen, liegt nicht in den Händen der Gutachter(innen), sondern der evaluierten Organisationen. Deshalb müssen wir analysieren, ob Institutsvertreter(innen) die Ergebnisse des Verfahrens als kollegiales Feedback oder externe Kontrolle deuten.

6.2 Die Institutsperspektive

Das Verfahren der RAE zielt – wie oben ausgeführt – auf die Bewertung individueller Leistungen von Wissenschaftler(inne)n der jeweiligen Forschungseinheit ab, gemessen in erster Linie an der Qualität der Veröffentlichungen. Die institutionellen und organisatorischen Rahmenbedingungen machen nur einen kleineren Teil des Gesamtergebnisses aus. Dennoch sind sowohl die Vorbereitung auf die RAE als auch deren Ergebnisse vielfach Anlass für umfassende Umstrukturierungen in den Instituten. Diese sind vielfach strategisch-taktischer Natur und zielen auf die Verbesserung der Evaluationsergebnisse durch die Ausnutzung und Manipulation der Verfahrenslogik („playing the game“). Man stößt aber auch auf Umgangsweisen, die weit darüber hinaus reichen und qualitative Verbesserungen zum Ziel haben.

Bereits in der Vorbereitung gilt es eine Reihe von strategischen Entscheidungen zu treffen. Schon die Frage, bei welchem Sub-Panel man die Unterlagen einreicht, kann einen Unterschied ausmachen, da manche Panels im Ruf stehen, strenger zu beurteilen als andere. Weiterhin kann als eine Art grundlegende strategische Entscheidung angesehen werden, wie viele Mitarbeiter(innen) als ‚forschungsaktiv‘ gemeldet werden. Dahinter stehen taktische Überlegungen: Geht man davon aus, dass die Budgetformel nicht linear ist, dass also sehr gute Gesamtnoten ein Mehrfaches an Budget versprechen, kann es rational sein, eher weniger und dafür nur die als Spitzenforscher(innen) geltenden Mitarbeiter(innen) anzugeben. In dem Zusammenhang ist es auch durchaus üblich, zusätzliche publikationsstarke Wissenschaftler(innen) von anderen Universitäten und Forschungseinrichtungen zu rekrutieren. Seitens des RAE-Managements wurde auf diese Taktik bereits im Laufe der Verfahren reagiert und festgelegt, dass alle gemeldeten Wissenschaftler(innen) eine bestimmte Mindestvertragslaufzeit und ein Mindestvolumen am jeweiligen Institut nachweisen müssen.

Strategie und Taktik spielen also in der Vorbereitung der RAE eine große Rolle – wie auch durch die Verfahrensregeln zu erwarten. Wie vielfach berichtet (z.B. Martin/Whitley 2010) führt die Differenzierung zwischen „active“ und „non-active researchers“ einerseits zu einer verstärkten Konkurrenz und Neubewertung von Mitarbeiter(inne)n des Instituts: „So I’m talking about guerilla researchers where maybe only ten percent of staff was submitted, or twenty percent [...], but you know you may end up with some perverse consequences whereby these eighty percent of people who weren’t submitted actually get some reward in some way from the exercise“ (Gomati, 362–367, Epsilon-UK). Auffällig und nicht in gleicher Weise erwartbar ist allerdings, dass auch im Rahmen der RAE die Auseinandersetzung der Institute keinesfalls auf dieser Ebene verbleibt. Die bevorstehende Evaluation wird

vielmehr gleichzeitig für interne Selbstverständigungsprozesse über die Verbesserung der Leistungsfähigkeit der zu bewertenden Einheiten genutzt. Neben der Einrichtung interner Monitoring-Systeme zur Feststellung der Publikationsleistungen von Mitarbeiter(inne)n werden auch zukünftige Publikationspläne eruiert und Möglichkeiten zur Verbesserung der Arbeitsbedingungen diskutiert. Um Publikationen (rechtzeitig) realisieren zu können, werden Unterstützungsstrukturen, z.B. in Form einer Reduktion des Lehrdeputats oder der Einstellung von Hilfskräften, geschaffen. Die Ergebnisse der RAE werden also keineswegs nur auf dem Weg der Manipulation zu verbessern versucht, sondern ebenfalls durch eine Qualitätssteigerung in der wissenschaftlichen Produktion: „I think that the (last) RAE has been good for this faculty. It has made us wake up and to reorganize ourselves in a way which means that we are producing more and higher quality stuff“ (Jäger, 545-548, Beta-UK). Dazu wurde häufig die RAE intern simuliert und der Austausch über Publikationen formalisiert und kollektiviert: „Like in some faculties, individuals were asked to assess, self-assess their own publications, rather than have panels assess. [...] I insisted that it was independent assessment. But by and large, the individuals were asked to recommend their own four best publications and then that was looked over by a research committee in each school and also by the faculty research executive and also by the university-wide executive“ (Fredo 206-212, Alpha-UK). In diesem Prozess spielen institutionelle Aspekte eine herausragende Rolle, die zu erheblichen Veränderungen der Organisation und Struktur von Fakultäten führen können: „we had now tried to organise ourselves as a research institution, that we weren't just an institution which organised its teaching and left individuals to do their own research, we tried to give support and structure to the whole research enterprise“ (Jäger, 356-359, Beta-UK).

Vielfach werden im Vorfeld der RAE Umstrukturierungen vorgenommen und Instrumente einer Selbstevaluation eingeführt. Der Anlass ist zweifelsohne die RAE und insofern wird alles daran gesetzt, auf den Bewertungsprozess vorab zu reagieren. Jenseits des Anlasses der externen Evaluation wird der darüber hinausreichende Wert der initiierten internen Prozesse für die Weiterentwicklung des Instituts oder der Fakultät herausgestellt.

Obwohl also allen Beteiligten vorab die RAE als kontrollierendes Verfahren bekannt ist, nutzen auch die Institutsvertreter(innen) die wenigen Hinweise als Anstoß für Veränderungen. Die Hinweise, also konkret die Notenprofile, lassen einen sehr großen Interpretationsspielraum zu. Die Institute erhalten eben keine explizierete inhaltliche Rückmeldung, sie erhalten lediglich das auf eine Nummer kondensierte Ergebnis einer ausführlichen Auseinandersetzung der Gutachter(innen). Auf Institutsebene werden Erklärungen für dieses Ergebnis entwickelt, so dass schließlich der Umgang mit dem Ergebnis von der jeweiligen Leitungsphilosophie eines Instituts abhängig ist. Im einen Fall werden kleinere Forschungseinheiten und engerer Austausch implementiert, im anderen Fall werden Mitarbeiter(innen) als Reaktion auf die RAE von Forschungsaufgaben entbunden und fortan nur noch in der Lehre eingesetzt. In beiden Fällen gehen Umgang mit und Reaktion auf die RAE weit über strategisch-taktisches Verhalten im Sinne der Verfahrensregeln hinaus.

Wie sehr seitens der Institute die RAE auch als kollegiale Unterstützung zur Weiterentwicklung der eigenen Arbeit angesehen wird, zeigt auch die enorme Unzufriedenheit mit der Ergebniskommunikation: „... there is a contrast because the panels have written [...] excellent texts for the whole subject area [...] and very little time on the institution specific, [...] it isn't worth the paper it's written down“ (Gomati, 486-490, Epsilon-UK). Im Sinne einer kollegialen Rückmeldung wünschen sich die Institutsvertreter(innen) eine ausführliche und inhaltliche Auseinandersetzung der Gutachter(innen) mit den evaluierten Einheiten, um von ihren Einschätzungen und Empfehlungen lernen zu können. Wie schon für die Gutachter(innen) gezeigt, so antizipieren auch die Vertreter(innen) der Institute die RAE als ein Verfahren, das die Qualität der eigenen Forschung verbessern soll. Mit diesem Anspruch stoßen sie zum Teil an die Grenzen der Verfahrensregeln und formulieren entsprechende Kritik.

7. Fazit und Ausblick

Aus der Tatsache, dass institutionelle Wissenschaftsevaluationen politisch initiierte Rechtfertigungsordnungen sind, werden oftmals weitreichende Schlüsse hinsichtlich eines Autonomieverlustes der Wissenschaft durch externe Kontrollen gezogen, ohne die Interpretationen und Handlungsweisen von Wissenschaftler(inne)n innerhalb von Evaluationsverfahren hinreichend zu berücksichtigen. Deshalb haben wir die Innenwelt von Evaluationsverfahren analysiert, die ein starkes Kontrollmoment (Ressourcenallokation) mit einer starken Beteiligung von Wissenschaftler(inne)n (Peer-Review) kombinieren. Insbesondere sind wir der Frage nachgegangen, wie die beteiligten Gutachter(innen) und Begutachteten in einem Evaluationsverfahren die Handlungssituation Evaluation deuten und wie sie auf diese Deutung handelnd reagieren. Dabei war auffällig, dass – ebenfalls im Kontrast zum dominanten Diskurs – die Handlungssituation Evaluation keineswegs ausschließlich als ‚externe Kontrolle‘ interpretiert wurde. Immer wieder stießen wir parallel dazu auf eine Deutung von Evaluationen als kollegiale Rückmeldung. Beide Orientierungsmuster standen in den von uns analysierten Evaluationsverfahren in steter Spannung zueinander und durchzogen die Evaluationsprozesse auf verschiedenen Ebenen. So haben Gutachter(innen) ihre Aufgabe nahezu durchgängig als einen Dienst an der Fachgemeinschaft beschrieben. Ihr Selbstverständnis sei dabei weniger, das jeweilige Forschungsfeld zu kontrollieren, als es mit Hilfe von Anregungen voranzubringen. Diesem Selbstverständnis sind allerdings bereits durch den Verfahrensrahmen enge Grenzen gesetzt. Deshalb kritisieren die Gutachter(innen) im RAE die geringen Rückmeldungsmöglichkeiten und berichten die Evaluierenden des WGL- und SEP-Verfahrens von der Schwierigkeit, hinter die Erfolgsberichte der Institute zu gelangen und produktive Kritik zu üben. Denn bei den Instituten tritt das Moment externer Kontrolle deutlicher zu Tage, da ihre Finanzierung und öffentliche Reputation vom Ergebnis der Evaluation abhängt. Gleichwohl finden wir auch seitens der

Institutsvertreter(innen) Deutungen, die Evaluationen als einen Anlass zur kollegialen Mobilisierung (in der Vorbereitungsphase) und Rückmeldung verstehen.

Die von uns untersuchten Evaluationstypen polarisieren, weil sie mit Kontrolle und Feedback (vgl. Blank 2008) zwei unterschiedliche Funktionen zugleich erfüllen wollen. Die Folge sind strukturell angelegte, kaum auflösbare, sondern allenfalls bearbeitbare Spannungen in den Situationsdefinitionen, Rollenverständnissen und Handlungsweisen von evaluativen Prozessen. Um genauer herauszuarbeiten, welche Verwerfungen diese konstitutive Spannung empirisch hervorbringt und unter welchen Bedingungen diese produktiv bearbeitet werden können, bedarf es tiefschärfere Analysen von Praktiken in weiteren Evaluationsverfahren. Denn Evaluationen werden trotz dieser internen Spannungen weder von den Evaluierenden noch von den Evaluierten prinzipiell abgelehnt. Kontrollen seien nötig, sofern Fachkolleg(inn)en sie vornehmen, und Evaluationen sogar hilfreich, sofern sie als kollegiales Feedback mit dem Ziel Lernprozesse und Verbesserungen herbeizuführen verstanden werden. Unsere Ergebnisse fügen sich somit unmittelbar in einen Diskurs ein, der im Sinne einer „qualitative future of research evaluations“ (vgl. Donovan 2007) nach Anschlussfähigkeiten von Evaluationen an die Wissenschaft sucht. Gefordert wird hier ein „modified peer review model“ (vgl. Hansson 2010), dem es unter Bedingungen der Kontrolle gelingt, dialogische Elemente zu stärken und organisationale Lernprozesse zu ermöglichen. Nimmt man die Akzeptanz von Evaluationsverfahren durch alle Beteiligten zum Maßstab für ihr Gelingen, dann muss man auf Basis unserer Studie vermuten: Je mehr Kommunikation und Rückmeldung Evaluationsverfahren zulassen, desto mehr Akzeptanz haben sie.

8. Literatur

- Blank, Stephan (2008): Evaluation und Lernprozesse: Eine systemtheoretische Analyse. In: *Zeitschrift für Evaluation*, 7 (1), S. 75-95.
- Cole, Jonathan R./Cole, Stephen (1981): *Peer Review in the National Science Foundation*. Washington, DC: National Academy of Sciences.
- Donovan, Claire (2007): The Qualitative Future of Research Evaluation. In: *Science and Public Policy*, 34 (8), S. 585-597.
- Espelund, Wendy N./Sauder, Michael (2007): Ranking and Reactivity: How Public Measures Recreate Social Worlds. In: *American Journal of Sociology*, 113 (1), S. 1-40.
- Gläser, Jochen (2006): *Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung*. Frankfurt a.M.: Campus.
- Gläser, Jochen/Laudel, Grit (2007): Evaluation without Evaluators: The Impact of Funding Formulae on Australian University Research. In: Whitley, Richard/Gläser Jochen (Hg.): *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*. Dordrecht: Springer, S. 127-151.
- Gülker, Silke/Simon, Dagmar (2011): Nach der Evaluation ist vor der Evaluation. Institutionelle Folgen von Forschungsbewertungen im internationalen Vergleich. In: Hornbostel, Stefan/Schelling, Anna (Hg.): *Evaluation: New Balance of Power? iFQ Working Paper No. 9*. Dezember 2011. Berlin: iFQ Institut für Forschungsinformation und Qualitätssicherung, S. 93-93.
- Gülker, Silke/Simon, Dagmar/Torka, Marc (2012): Evaluation of Science as Consultancy. In: *Quaderni: Research Evaluation: Reopening the Controversy*, 1/2012, No. 77, S. 41-54.

- Hansson, Finn (2010): Dialogue in or with the Peer Review? Evaluating Research Organizations in Order to Promote Organizational Learning. In: *Science and Public Policy*, 37 (4), S. 239-251.
- Higher Education Funding Council (HEFCE) (2011): REF 2014. Research Excellence Framework. Assessment Framework and Guidance on Submissions. Verfügbar unter: http://www.hefce.ac.uk/research/ref/pubs/2011/02_11/02_11.pdf [21.12.2011].
- Hirschauer, Stefan (2004): Peer Review auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation. In: *Zeitschrift für Soziologie*, 33 (1), S. 62-83.
- Hirschauer, Stefan (2006): Wie geht Bewerten? Zu einer anderen Evaluationsforschung. In: Flick, Uwe (Hg.): *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. Hamburg: Rowohlt, S. 405-423.
- Kieser, Alfred (2010): Unternehmen Wissenschaft? In: *Leviathan*, 3/2010, S. 347-367.
- KNAW (2009): Standard Evaluation Protocol 2009-2015. Protocol for Research Assessment in the Netherlands. Amsterdam.
- Lamont, Michèle (2009): *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge: Harvard University Press.
- Lange, Stefan/Gläser, Jochen (2009): Performanzsteigerung durch Selektivität? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen. In: *der moderne staat – Zeitschrift für Public Policy, Recht und Management*, 2 (2), S. 411-432.
- Langfeld, Liv (2009): The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome. In: *Social Studies of Science*, 31 (6), S. 820-841.
- Leibniz-Gemeinschaft (2007): Kriterien für die Evaluierung von Einrichtungen der Leibniz-Gemeinschaft (SEN 0057/07, 18.07.2007).
- Martin, Ben/Whitley, Richard (2010): The UK Research Assessment Exercise: A Case of Regulatory Capture? In: Whitley, Richard/Gläser, Jochen/Engwall, Lars (Hg.): *Reconfiguring Knowledge Production. Changing Authority Relationships in the Science and their Consequences for Intellectual Innovation*. Oxford: Oxford University Press, S. 51-80.
- Matthies, Hildegard/Matthäus, Sandra (2010): Science Between Organization and Profession: Opportunities for Neutralizing Gender? – Reflections on Research Assessment. In: Riegraf, Birgit/Aulenbacher, Brigitte/Kirsch-Auwärter, Edit/Müller, Ursula (Hg.): *Gender Change in Academia. Re-Mapping the Fields of Work, Knowledge, and Politics from a Gender Perspective*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 87-100.
- Meier, Frank/Schimank, Uwe (2010): Mission Now Possible: Profile-Building and Leadership in German Universities. In: Whitley, Richard/Gläser, Jochen/Engwall, Lars (Hg.): *Reconfiguring Knowledge Production. Changing Authority Relationship in the Science and their Consequences for Intellectual Innovation*. Oxford: Oxford University Press, S. 211-236.
- Merton, Robert K. (1990 [1942]): The Normative Structure of Science. In: Alexander, Jeffrey/Seidman, Steven (Hg.): *Culture and Society*. Cambridge: Cambridge University Press, S. 67-74.
- Merton, Robert K./Zuckerman, Harriet (1971): Patterns of Evaluation in Science. In: *Minerva*, 9 (1), S. 66-100.
- Michaels, Axel (2010): Evaluation als akademisches Ritual. In: Hornbostel, Stefan/Schelling, Anna (Hg.): *Evaluation: New Balance of Power?* iFQ Working Paper No. 9. Dezember 2011. Berlin: iFQ Institut für Forschungsinformation und Qualitätssicherung, S. 25-32.
- Neidhardt, Friedhelm (1988): *Selbststeuerung in der Forschungsförderung. Das Gutachterwesen der DFG*. Opladen: Westdeutscher Verlag.
- Neidhardt, Friedhelm (2010): Selbststeuerung der Wissenschaft: Peer Review. In: Simon, Dagmar/Knie, Andreas/Hornbostel, Stefan (Hg.): *Handbuch Wissenschaftspolitik*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 280-292.
- Popper, Karl R. (1973 [1934]): *Die Logik der Forschung*. Tübingen: Mohr Siebeck (Fünfte Aufl. Nachdruck der vierten, verbesserten Aufl.).
- Röbbecke, Martina/Simon, Dagmar (2001): *Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten*. Berlin: edition sigma.
- Schimank, Uwe (2004): Leistungsbeurteilung von Kollegen und Politikberatung am Beispiel von Evaluationen im Hochschulsystem. In: Schützel, Rainer/Brüsemeyer, Thomas (Hg.): *Die*

- beratende Gesellschaft: Zur gesellschaftlichen Bedeutung von Beratung. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 39-56.
- Shapin, Steven (1994): *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: University of Chicago Press.
- Simon, Dagmar (2007): Als Konsequenz mehr Kohärenz? Intendierte und nicht intendierte Wirkungen von institutionellen Evaluationen. In: Matthies, Hildegard/Simon, Dagmar (Hg.): *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 178-192.
- Torka, Marc (2011a): Der Evaluator. Positionierungen, Strategien, Deutungs-, Handlungs- und Urteilsweisen von Wissenschaftlern in Prozessen der Wissenschaftsevaluation. In: *Sozialer Sinn – Zeitschrift für hermeneutische Sozialforschung*, 1/2011, S. 115-138.
- Torka, Marc (2011b): Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen. In: Hornbostel, Stefan/Schelling, Anna (Hg.): *Evaluation: New Balance of Power?* iFQ Working Paper No. 9, Dezember 2011. Berlin: iFQ Institut für Forschungsinformation und Qualitätssicherung, S. 69-81.
- Travis, G.D.L./Collins, Harry M. (1991): New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System. In: *Science, Technology & Human Values*, 16, No. 3, S. 322-341.
- Weller, Ann C. (2001): *Editorial Peer Review. Its Strengths and Weaknesses*. New Jersey: Information Today Inc.
- Whitley, Richard (2007): Changing Governance of the Public Sciences. In: Whitley, Richard/Gläser, Jochen (Hg.): *The Changing Governance of the Sciences*. Dordrecht, Netherlands: Springer, S. 3-27.
- Whitley, Richard/Gläser, Jochen/Engwall, Lars (2010): *Reconfiguring Knowledge Production. Changing Authority Relationships in the Science and their Consequences for Intellectual Innovation*. Oxford: Oxford University Press.