

Gülker, Silke; Matthies, Hildegard; Matthäus, Sandra

Working Paper

Evaluationsverfahren aus laborkonstruktivistischer Perspektive

WZB Discussion Paper, No. SP III 2009-601

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Gülker, Silke; Matthies, Hildegard; Matthäus, Sandra (2009) : Evaluationsverfahren aus laborkonstruktivistischer Perspektive, WZB Discussion Paper, No. SP III 2009-601, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:

<https://hdl.handle.net/10419/50774>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WZB – discussion paper

Silke Gülker, Hildegard Matthies, Sandra Matthäus

Evaluationsverfahren aus laborkonstruktivistischer Perspektive

SP III 2009-601

guelker@wzb.eu

matthies@wzb.eu

matthaeus@wzb.eu

Suggested Citation/Zitierweise:

Gülker, Silke & Hildegard Matthies & Sandra Matthäus (2009): Evaluationsverfahren aus konstruktivistischer Perspektive. *WZB Discussion Paper SP III 2009-601*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung.

**Social Science Research Center Berlin
(WZB)**

Research Area Society and Economic
Dynamics

Research Group Science Policy Studies

**Wissenschaftszentrum Berlin für Sozial-
forschung gGmbH (WZB)**

Forschungsschwerpunkt Gesellschaft und
wirtschaftliche Dynamik

Forschungsgruppe Wissenschaftspolitik

Reichpietschufer 50, D-10785 Berlin
Telefon: +49 30 25491-0, Fax: +49 30 25492-684
www.wzb.eu

ABSTRACT

Today, it is increasingly common to evaluate scientific organisations. Consequently, the old debate on adequate procedures and criteria for the assessment of scientific quality is relevant again. The paper emphasizes the impact of the *codes of practice* on the process of judgement. Following Luhmann (1969), evaluations are conceptualized as procedures – as independent social systems they develop a certain momentum of their own. At the same time, they are limited by a number of previous selections such as roles and required techniques. The analysis focuses on these selections. The main question is to which extent there are different codes of practices in different regimes of evaluation and to which extent we can assume them to have a structuring impact on the process of assessment. Subjects of our analysis are the codes of practice of the Wilhelm-Leibniz-Gemeinschaft in Germany, the “Standard Evaluation Protocol for Public Research“ in the Netherlands and the „Research Assessment Exercises“ (RAE) in Great Britain. A constructivist perspective as Karin Knorr Cetina has offered, serves as an heuristic foil to reveal the social components of rules and techniques.

ZUSAMMENFASSUNG

Einrichtungen der Wissenschaft werden mit zunehmender Selbstverständlichkeit evaluiert. Damit bekommt auch die alte Debatte um angemessene Verfahren und Kriterien der Qualitätsbewertung eine neue Aktualität. Das Papier hebt die Bedeutung von *Verfahrensregeln* für den Urteilsprozess hervor. Evaluationen werden als Verfahren im Sinne Luhmanns (1969) angenommen. Als eigenständige Systeme entwickeln sie ein gewisses Maß an Eigendynamik. Gleichzeitig sind die durch eine Vielzahl von vorgelagerten Selektionsleistungen begrenzt. Die Analyse fokussiert auf eben diese Selektionsleistungen. Untersucht wird, wie sich Verfahrensregeln in unterschiedlichen Evaluationsregimes unterscheiden und welche strukturierende Wirkung auf den Bewertungsprozess jeweils damit angenommen werden kann. Gegenstand der Analyse sind die Verfahrensregeln der Wilhelm-Leibniz-Gemeinschaft in Deutschland, des “Standard Evaluation Protocol for Public Research“ in den Niederlanden und des „Research Assessment Exercises“ (RAE) in Großbritannien. Als heuristische Folie wird der laborkonstruktivistische Ansatz von Knorr-Cetina verwendet, weil damit auch die soziale Komponente von Regeln und Techniken gut herausgearbeitet werden kann.

Inhalt

1	EINFÜHRUNG	1
2	EINSCHLÜSSE UND AUSSCHLÜSSE DURCH VERFAHRENSREGELN	4
3	VERFAHRENSREGELN VON EVALUATIONEN IM INTERNATIONALEN VERGLEICH	10
3.1	Evaluationsverfahren der Wilhelm-Leibniz-Gemeinschaft in Deutschland	11
3.2	Evaluationsverfahren des “Standard Evaluation Protocol for Public Research“ in den Niederlanden	20
3.3	Evaluationsverfahren des „Research Assessment Exercises“ (RAE) in Großbritannien	24
4	FAZIT	32
	LITERATUR	36

1 Einführung

Die Einrichtungen der Wissenschaft werden mit zunehmender Selbstverständlichkeit evaluiert und gleichzeitig nehmen Intensität und Aufwand der Bewertungsverfahren zu. Damit bekommt auch die alte Debatte um angemessene Verfahren und Kriterien der Qualitätsbewertung eine neue Aktualität. Die Frage der Angemessenheit ist allerdings nicht von den Erwartungen zu trennen, die an Evaluationen gerichtet werden und die fallen je nach Interessenlage und Zielen ganz unterschiedlich aus. Während beispielsweise die Wissenschaft Evaluationen eher als Instrument zur Qualitätssicherung von Forschung und Lehre versteht, sollen sie für die Institutionen der staatlichen Politik vor allem eine effiziente Mittelallokation unterstützen. Der allgemeine Trend zur Suche nach der „golden goose of effectiveness“ (Stanley Cohen, zit. nach: Power 1997: 141) scheint sich damit auch auf die Wissenschaft auszudehnen und Evaluationen werden als eine Art Controlling- und Benchmarkinginstrument von Wissenschaftseinrichtungen eingesetzt (vgl. etwa Power 1994; Power 2008; Strathern 2000).

Wenn auch international unterschiedlich institutionell verankert, wird die „Gesamtperformanz“ wissenschaftlicher Einrichtungen mittlerweile in regelmäßigen Abständen nahezu flächendeckend überprüft. In der Regel geschieht dies durch ein Gremium, das sowohl aus wissenschaftlichen FachgutachterInnen als auch aus RepräsentantInnen staatlicher Instanzen zusammengesetzt ist. Dabei beruhen die Bewertungen entweder allein auf der Sichtung schriftlicher Dokumente der zu bewertenden Einrichtung (z.B. in Großbritannien) oder sie werden durch ein- bis zweitägige Vor-Ort-Besuche ergänzt, in dessen Rahmen die wissenschaftliche Arbeit von VertreterInnen der Einrichtung präsentiert und mit den GutachterInnen diskutiert wird. Vorbereitet werden diese „Begehungen“ in der Regel durch eine umfassende schriftliche Selbstdarstellung der Institute, die auch eine Reihe quantitativer Daten enthält. Am Ende der Begutachtung wird ein Bewertungsbericht erstellt, der sowohl der bewerteten Einrichtung als auch den wissenschaftspolitischen Auftraggebern der Evaluation ausgehändigt wird. Neben dem eigentlichen Gutachten enthält dieser Bericht zumeist Empfehlungen, die sich sowohl an das evaluierte Institut als auch an den politischen Auftraggeber richten. Sie betreffen vor allem die künftige programmatische Orientierung der Einrichtung, die Organisationsstruktur und das Management sowie die finanzielle Förderung. Der Ausgang einer Evaluation kann also erhebliche Konsequenzen für die betreffende

Einrichtung haben, bis hin zum Umfang des künftigen Budgets. Zugleich sind diese Empfehlungen Meßlatte für die nächste Evaluation.

Aufgabe der Evaluation von wissenschaftlichen Einrichtungen ist folglich vor allem die Generierung von Informationen, auf dessen Basis Entscheidungen der alimentierenden Institution, zum Beispiel des Staates, legitimiert werden können. Sie unterscheiden sich damit auch substantiell von anderen Bewertungsverfahren, die seit jeher selbstverständlicher Teil der wissenschaftlichen Selbststeuerung sind. Anders als etwa bei einem Peer Review von Zeitschriftenartikeln ist das Urteil der FachgutachterInnen im Rahmen von Organisationsevaluationen weniger an der Frage nach dem wissenschaftlichen Wahrheitsgehalt orientiert. Vielmehr geht es darum, Merkmale zu identifizieren, von denen angenommen wird, dass sie den Prozess der Wissensproduktion in einer spezifischen Einrichtung positiv (oder negativ) beeinflussen und aus denen Schlussfolgerungen zur zukünftigen wissenschaftlichen Leistungsfähigkeit gezogen werden können. Es geht also vorrangig um normative oder präskriptive Fragen, über die sich die GutachterInnen im Prozess der Bewertung zu verständigen und nach Möglichkeit einvernehmliche Antworten zu finden haben. Entsprechend wird im Bewertungsprozess nicht allein entlang des Codes „wahr“ oder „falsch“ differenziert, sondern auch entlang des Codes „gut“ oder „schlecht“ im normativen Sinn. Die Ergebnisse dieser Bewertungen sind schlussendlich für einen politischen Entscheidungsprozess kompatibel zu machen, das heißt, sie müssen in den Code „förderungswürdig/nicht förderungswürdig“ übersetzt werden können.

Das Projekt „Urteilsbildung im Peer Review – Internationale Fallstudien zur Evaluation von wissenschaftlichen Einrichtungen“¹ hat diese spezifische, von der Peer Review Forschung bislang weitgehend unberücksichtigte Form von Wissenschaftsbewertung zum Gegenstand. Die Frage „Wie geht eigentlich bewerten?“

¹ Das Projekt wird in der Forschungsgruppe Wissenschaftspolitik am Wissenschaftszentrum Berlin für Sozialforschung umgesetzt (für Details siehe <http://www.wzb.eu/gwd/wipo/projects.de.htm>). In dem Vorhaben werden über ein methodisch aufwändiges Verfahren der Fallrekonstruktion zu den drei genannten Verfahrensbeispielen jeweils mehrere ausgewählte Evaluationsfälle analysiert. Hauptmaterialkorpus sind dabei möglichst zeitnah zum Evaluationsfall durchgeführte leitfadengestützte Interviews mit den am Verfahren beteiligten Akteuren, VertreterInnen der Evaluationsagenturen, der evaluierten Einrichtungen sowie der GutachterInnen, als auch die im Verfahren erzeugten Dokumente (Selbstdarstellung der Institute und Bewertungsbericht).

(Hirschauer 2006) soll hier für den Prozess von Organisationsevaluationen spezifiziert werden.

Das vorliegende Papier thematisiert eine konzeptionelle Teilfrage des Projektes, die für die empirische Arbeit von zentraler Bedeutung erscheint: Welche Bedeutung haben Verfahrensregeln für den Urteilsbildungs- und Entscheidungsprozess im Evaluationsverfahren?

Evaluationen werden als Verfahren im Sinne Luhmanns (2001 [1969]) angenommen, die temporär den Charakter von eigenständigen Systemen entwickeln und somit eine gewisse Eigendynamik entwickeln. Gleichzeitig sind sie allerdings durch eine Vielzahl von dem konkreten Verfahren vorgelagerten Selektionsleistungen – nämlich Verfahrensregeln – begrenzt, die eine gewisse Vergleichbarkeit der Evaluationspraxis erzeugen sollen. Dazu zählen beispielsweise Regeln, die den Gegenstand der Evaluation festlegen und eingrenzen, Regeln zur Zulassung von Personen und zu deren Rollen im Verfahren – etwa wer evaluieren darf –, Regeln zur Form des Ergebnisses, zum Beispiel dass ein konsensuales Ergebnis erwartet wird und dass es schriftlich zu verfassen ist, bis hin zu Regeln, die festlegen, was aus den Ergebnissen einer Bewertung folgt. Die Entscheidungen über den Inhalt solcher Regeln sind, zumindest in Teilen, schon getroffen, bevor das eigentliche Evaluationsverfahren beginnt.

Wir konzentrieren die folgende Analyse auf eben diese Selektionsleistungen und untersuchen, wie sich Verfahrensregeln in unterschiedlichen Evaluationsregimes unterscheiden und welche strukturierende Wirkung auf den Bewertungsprozess damit angenommen werden kann. Empirischer Gegenstand unserer Analyse sind dabei die Verfahrensregeln der Wilhelm-Leibniz-Gemeinschaft in Deutschland, des “Standard Evaluation Protocol for Public Research“ in den Niederlanden und des „Research Assessment Exercises“ (RAE) in Großbritannien.

Im folgenden konkretisieren wir zunächst die Spezifika von Evaluationen als Verfahren und entwickeln anhand des laborkonstruktivistischen Ansatz von Knorr-Cetina eine heuristische Folie zur Gegenüberstellung der ausgewählten Verfahren (Kapitel 2). Kapitel 3 präsentiert die vergleichende Analyse, deren Ergebnisse anschließend in einem Fazit zusammengefasst werden (Kapitel 4).

2 **Einschlüsse und Ausschlüsse durch Verfahrensregeln**

Angesichts ihres verfahrensförmigen Charakters weisen Evaluationen von wissenschaftlichen Einrichtungen starke Parallelen zu juristischen Urteilsfindungen auf (Hornbostel 2008: 72). Wie Gerichtsverfahren haben sie einen mehr oder minder klar definierten politischen Auftrag in einer bestimmten Angelegenheit, nämlich der kooperativen Suche nach der angemessenen Bewertung der Leistung einer wissenschaftlichen Einrichtung. Mit Luhmann kann man ihnen somit den Status sozialer Systeme zuweisen, „die eine spezifische Funktion erfüllen, nämlich eine einmalig verbindliche Entscheidung zu erarbeiten“ (Luhmann 2001 [1969]: 41). Damit sind sie zugleich „von vornherein in ihrer Dauer begrenzt“ (ebd.). Ihre Funktion leitet sich zwar aus einem übergeordneten, das Verfahren überdauernden System² ab, doch für die Dauer ihrer Existenz weisen sie wie Gerichtsverfahren typische Merkmale eines selbstreferenziell operierenden sozialen Systems auf. Sie stellen eine Differenz zwischen innen und außen her, generieren ihre eigenen Regeln und erzeugen dadurch eine Reduktion von Komplexität. Dieses heißt etwa, dass die an einer Evaluation beteiligten Akteure nicht auf vorfixierte Auslöser reagieren, sondern nur auf je spezifische Selektionsleistungen anderer, die selbst Teil des Verfahrens sind. Bedingt durch das Verhalten der Beteiligten entwickelt jede Evaluation somit eine je spezifische Matrix möglicher Ereignisse, in deren Verlauf die Handlungsmöglichkeiten der Beteiligten sich sukzessive zu wenigen Alternativen verdichten bis schließlich eine Entscheidung gefunden ist (Luhmann 2001 [1969]: 45): das gemeinsame Urteil der Begutachtungsgruppe. Die Verfahrensstruktur schafft somit eine Offenheit für Kritik und Alternativen, die einfacheren sozialen Systemen verschlossen bleibt, „namentlich Funktionen der *kooperativen* Wahrheitssuche von divergierenden Standpunkten aus und Funktionen des Darstellens und Austragens von Konflikten“ (ebd.: 50).

Gleichwohl sind auch Evaluationen Teil eines übergeordneten sozialen Systems, durch das ihre Autonomie begrenzt wird. Dies geschieht vor allem in Form von an die Evaluation gerichteten generalisierten Verhaltenserwartungen (Normen und generelle Verfahrensvorschriften). Regeln dieser Art sind die Grundlage für zahl-

² In unserem Fall handelt es sich hierbei um eine Art „Grenzorganisation“ (Braun 2004 in Anlehnung an Guston 2000) der Art, dass die Rahmenbedingungen für Evaluationen durch Organisationen des Wissenschaftssystems in Abstimmung mit staatlichen Organen bestimmt werden.

reiche Selektionsentscheidungen im Evaluationsverfahren selbst, etwa was evaluiert wird oder wer evaluiert, und konstituieren somit stets auch in spezifischer Weise den Gegenstand der Bewertung. Darüber hinaus begrenzen sie den Raum, in welchem die Kommunikation innerhalb eines Bewertungsverfahrens variieren kann, so dass auch der Prozess des Bewertens bereits auf ihm vorgängigen Strukturen aufsetzt.

In dem vorliegenden Papier werden die Verfahrensregeln für drei Bewertungsverfahren daraufhin überprüft, welche Selektionsentscheidungen jeweils explizit oder implizit mit ihnen verknüpft sind und welche Konsequenzen dieses sowohl im Hinblick auf den Bewertungsgegenstand als auch den Prozess der Bewertung als kooperativen Akt der Urteilsbildung hat. Was wird in Folge der Verfahrensregeln beispielsweise mehr und was weniger beleuchtet? Was wird möglicherweise ganz aus der Bewertung ausgeklammert? Wer bewertet? Mit welchen Mitteln verschaffen sich diejenigen, die bewerten, einen Eindruck vom Gegenstand? Die Betrachtung zeigt, dass in jede Verfahrensregel eine Fülle weiterer Entscheidungen über Selektionen eingelagert sind, die in ihrer Gesamtheit das Regelwerk einer Evaluation weitaus komplexer machen, als es auf den ersten Blick erscheinen mag.

Um die vielschichtigen Selektions- und Konstruktionsleistungen von Verfahrensregeln aufzuspüren, haben wir im laborkonstruktivistischen Ansatz von Knorr-Cetina (Knorr-Cetina 1988; Knorr-Cetina 1991) eine geeignete heuristische Folie gefunden. Dies mag auf den ersten Blick verwundern, liegt der Fokus dieser Arbeiten doch auf der Analyse wissenschaftlicher Praxis und nicht deren Bewertung. Knorr-Cetinas Erkenntnisinteresse geht allerdings über den konkreten Analysegegenstand hinaus. Ebenso wie in späteren Arbeiten zu den Finanzmärkten fragt Knorr-Cetina schon in ihren Laborstudien nach den (Produktions-)Techniken von Sinn und Bedeutung, also Wissen, und inwiefern die Konfiguration dieser Techniken ihre Produkte beeinflusst (Knorr-Cetina 2005a; Knorr-Cetina 2005b).

Auch Verfahrensregeln lassen sich im weiteren Sinn als Techniken oder Werkzeuge verstehen, die dazu dienen, die Erzeugung eines damit angestrebten Ergebnisses, im Fall von Evaluationen etwa einer konsensualen Beurteilung, zu vereinfachen. Somit trifft auch für sie zu, was Knorr-Cetina in ihren Laborstudien und Finanzmarktanalysen herausgearbeitet hat: Mit jeder Entscheidung für ein bestimmtes Werkzeug geht eine Entscheidung für eine spezifische Betrachtungs-

weise eines Gegenstandes einher, die sowohl bestimmte Möglichkeiten eröffnet als auch bestimmte Ausschlüsse vornimmt.

Die Kernidee der Laborstudien von Knorr-Cetina besteht jedoch darin, sich der Analyse wissenschaftlicher Erkenntnisproduktion von einem anthropologischen Ansatz her zu nähern (vgl. Harré 1991 in Knorr-Cetina 1991: 10). Aus dieser Perspektive ist nichts a priori gesetzt, sondern gestaltet sich im sozialen Handeln. Auch wissenschaftliche Tatsachen sind demnach nicht losgelöst vom Forschungshandeln und den Erzeugungspraktiken im Labor zu denken, das heißt, sie haben stets eine soziale Komponente (vgl. Knorr-Cetina 1988). Somit unterliegt jegliche Art von Wissen Konstruktionsakten, die als grundsätzlich analysierbare Handlungsprozesse angesehen werden können und von denen ein Großteil in den Laboren zu beobachten ist (vgl. Knorr-Cetina 1991: 22). Es ist anzunehmen, dass diese Beobachtung für Evaluationen als Prozess der kooperativen Urteilsbildung umso mehr zutrifft, da Werturteile in der Regel ohnehin nur eine Annäherung an Wahreitsaussagen darstellen und sich insofern zusätzlicher Begründungen bedienen müssen.³

Labore werden in Knorr-Cetinas Ansatz als Orte mit einer eigenständigen sozialen Ordnung angesehen, welche die Gestalt wissenschaftlicher Erkenntnisse beeinflussen (vgl. Harré 1991: 11). Erst durch die sozialen und alltäglichen Praktiken wird den Objekten im Labor ein (wissenschaftlicher) Sinn gegeben (Knorr-Cetina 1988: 85). Labore stellen in dieser Lesart lokale, domestizierte und damit artifizielle „Handlungskontexte dar, die in eingegrenzter Weise auf bestimmte Erzeugungsvorgänge spezialisiert sind. Die Erzeugbarkeit des gesuchten Produkts – und nicht etwa dessen Übereinstimmung mit irgendeiner unabhängigen ‚Natur‘ validiert hierbei das Behandlungsprogramm des Labors sowie die Wirklichkeit des Produkts. (...) Untersuchungsobjekte werden, sofern sie aus einer ‚natürlichen‘ oder ‚alltäglichen‘ Umwelt stammen, im Labor mit einer neuen Lebenswelt konfrontiert, innerhalb derer sie als Objekt neu konstituiert werden “ (Knorr-Cetina 1988: 87). Diesen Prozess des Herauslöses der Untersuchungsobjekte aus ihrer ‚Natur‘ und deren Neukonstituierung im Labor wird als Laboratorisierungsprozess

³ Inwiefern Evaluationen auch als Prozesse wissenschaftlicher Wissensproduktion interpretiert werden können, das heißt, ob und inwiefern sie auch dem Modus der Wahrheitssuche folgen, wird von uns derzeit konzeptionell und empirisch geprüft und wird Gegenstand weiterer Publikationen sein.

beschrieben: Die Untersuchungsobjekte werden dabei für das Labor ‚passend gemacht‘, für die Untersuchung ‚zugerichtet‘ (vgl. ebd.: 87f.) und damit an die kulturellen Akteure, also z.B. die WissenschaftlerInnen angepasst und in deren kulturelle Interaktion eingebunden.

Mit der Laboratorisierung einer Wissenschaft einher geht der Übergang zu einer spezifischen Erzeugungslogik, der die Produktion und v.a. Prozessierung von Zeichen zu Eigen ist (Knorr-Cetina 1988: 91). Die Neukonstituierung der ‚natürlichen Untersuchungsobjekte‘ im Labor erfolgt über und resultiert in Zeichen, wie z.B. Impulsen und Reaktionskomponenten (vgl. ebd.: 87). Diese wiederum werden von spezifischen Apparaturen hervorgebracht, wie z.B. der Gesprächsapparatur in der Sexualwissenschaft. Sie können in Form von visuellen, verkörperten und szenischen Zeichen auftreten (vgl. ebd.: 92). Dabei sind Zeichen jedoch vielfach noch *ohne* Sinn. Dieser muss erst hergestellt werden, so dass Zeichen *mit* Sinn, also Objekte, entstehen können (vgl. ebd.: 93). Genau bei dieser Bedeutungskonstitution von Zeichen stoßen die technischen Apparaturen und Maschinen des Labors jedoch an ihre Grenzen und werden durch soziale Erzeugungsinstrumente von Sinn wie Mündlichkeit, Schriftlichkeit und Körperlichkeit ergänzt. Einzelne Laborformen können folglich entlang der Frage, ob und in welcher Weise Handlungsformen wie Mündlichkeit, Schriftlichkeit und Körperlichkeit „in Erzeugungsinstrumente für wissenschaftliche Ergebnisse“ (ebd.: 93) transformiert werden, voneinander unterschieden werden. Insbesondere für die mündlichen und körperlichen Erzeugungsinstrumente gilt, dass sie „im naturwissenschaftlichen Labor nahtlos [dort ansetzen], wo die Arbeit technischer Geräte endet“ (Knorr-Cetina 1988: 93). Dabei kommt dem Körper gleichsam eine übersinnliche Funktion zu, denn ihm wird zugetraut, das zu identifizieren, was Gesprächsapparatur und Verstand nicht antizipieren können, was also dem reflexiven Bewusstsein nicht zugänglich ist, sondern nur unterschwellig wahrgenommen wird. Möglich wird dies, weil der Körper neben seiner Funktion als „Messinstrument“ (ebd.: 97) zugleich als „Archiv und Depot von Erfahrung“ (ebd.) eingesetzt werden kann.⁴ Er „misst“ somit nicht nur, sondern interpretiert das Gemessene zugleich und stattet es mit Sinn aus, indem er es im Lichte seines gesammelten körperlichen Wissens interpretiert. Die Nähe zum therapeutischen

⁴ Hier lassen sich auch interessante Bezüge zu Polanyis Begriff des „tacit knowledge“ herstellen.

Empfindungsvokabular in den Beschreibungen zahlreicher Laborversuche ist ein Ausdruck davon.

Auch die Verfahrensregeln zur Evaluation von wissenschaftlichen Einrichtungen erzeugen – im übertragenen Sinn – einen Laboratorisierungsprozess, denn wie im wissenschaftlichen Labor konstituieren sie das zu evaluierende Objekt in einer ganz bestimmten Art und Weise. Sie beinhalten beispielsweise Entscheidungen darüber, was bewertet werden soll – und was nicht – und legen außerdem fest, mit welchen Methoden und durch wen dieses geschehen soll. Dabei können Verfahrensregeln danach unterschieden werden, ob sie direkte oder indirekte Formen von Laboratorisierungsprozessen erzeugen. Zur direkten Form lassen sich formalisierte Praktiken zum Verfassen von Selbstberichten der zu bewertenden Organisation, die Einteilung der Organisation in spezifische Bewertungseinheiten, die Auswahl der zu bewertenden Produkte, z.B. bestimmte Projekte und Publikationen sowie die Auswahl des Zeitpunktes der Evaluation sowie die Auswahl der Evaluationsintervalle zählen. Der indirekten Form dagegen können formal festgelegte Bewertungskriterien zugerechnet werden. Auch sie lösen eine Zurichtung des Gegenstandes aus, denn ihre Funktion ist, das Augenmerk der FachgutachterInnen auf ganz bestimmte Aspekte zu lenken. Entsprechend konnten wir in unseren Interviews, die wir mit VertreterInnen der Institute geführt haben, feststellen, dass sich die zu bewertenden wissenschaftlichen Einrichtungen bei der Vorbereitung auf die Evaluation strategisch an den Bewertungskriterien ausrichten und sich schon beim Verfassen ihrer Selbstberichte daran orientieren. Auch für informelle Kriterien kann hierbei in Gestalt von ‚Erwartungserwartungen‘ die Wirkung einer indirekten Laboratorisierung identifiziert werden.

Auch in einem Evaluationsverfahren geht es somit um eine selektive Produktion und Prozessierung von Zeichen und Objekten. Und wie im wissenschaftlichen Labor kommen dabei auch soziale Erzeugungsinstrumente von Wissen wie Mündlichkeit, Schriftlichkeit und Körperlichkeit vor allem da zum Einsatz, wo bloßes Zählen und standardisiertes Messen, etwa das Berechnen bibliometrischer Daten, nicht ausreicht.

Mündlichkeit spielt in Evaluationen als Prozess der kooperativen Wahrheitssuche eine zentrale Rolle. So ist der Dialog zwischen den GutachterInnen ohne Mündlichkeit kaum denkbar. Darüber hinaus ist Mündlichkeit in einigen Verfahren auch in Form von Gesprächen mit den Begutachteten vorgesehen. Das laboratori-

sche Feld für diese Instrumentierung ist der Vor-Ort-Besuch der Einrichtung. Die Bedeutung der Gespräche vor Ort für die Urteilsbildung wird insbesondere von den GutachterInnen hervorgehoben. Vielfach zeigen sich dabei auch starke Bezüge zur Bedeutung der Körperlichkeit, indem etwa die GutachterInnen die hohe Bedeutung des Vor-Ort-Gewesenseins hervorheben und ihre Eindrücke vom Bewertungsgegenstand häufig unter Verwendung von Begrifflichkeiten schildern, die eine starke Nähe zum therapeutischen Empfindungsvokabular aufweisen. Der Körper fungiert somit auch in Evaluationsverfahren in seiner Doppелеigenschaft, sowohl ‚Messinstrument‘ als auch ‚Depot und Archiv von Erfahrung‘ zu sein. Letztere Funktion übernimmt der GutachterInnenkörper durch die in ihm deponierten Erfahrungen zu vergleichbaren Gegenständen, die bei den Vor-Ort-Begehungen oder Diskussionsrunden abgerufen werden und beim Messen als Interpretationsfolie dienen. Demgegenüber hat Schriftlichkeit vor allem die Aufgabe der Dokumentation, sowohl der zu bewertenden Leistung als auch des Bewertungsprozesses, zum Beispiel in Gestalt von Protokollen. Auch Bewertungsvordrucke, die entweder numerische oder wörtliche Urteile oder Stellungnahmen der GutachterInnen vorsehen, können dazu gezählt werden. Je nach Stellenwert dieser Dokumente stellt Schriftlichkeit somit vor allem eine Verbindung zwischen dem Gemessenen und dessen Bewertung her und macht sie auch nachträglich intersubjektiv zugänglich. Insofern wird über den Stellenwert von Schriftlichkeit in einem Verfahren unter Umständen auch dessen Verbindlichkeit hergestellt.

Analog zu Wissenserzeugungsprozessen verorten wir im Folgenden Evaluationen von Wissenschaftseinrichtungen in einem von vorhergehenden Selektionen konstituierten virtuellen Raum, dessen Format durch Verfahrensregeln bestimmt wird. Es wird analysiert, wie sich die Laboratorisierung von Evaluationsobjekten je nach Verfahrensregeln unterschiedlich gestaltet, welchen Stellenwert dabei Mündlichkeit, Schriftlichkeit und Körperlichkeit haben und welche Auswirkungen der je spezifische Mix dieser Erzeugungsinstrumente von Zeichen und Objekten auf die Begutachtungspraxis haben könnte.

3 Verfahrensregeln von Evaluationen im internationalen Vergleich

Unsere Untersuchung konzentriert sich wie bereits erwähnt auf die Verfahrensregeln von Evaluationen. Sowohl das übergeordnete Evaluationssystem, das über eine jeweilige Evaluation hinaus Bestand hat, als auch das konkrete, auf den einmaligen Zweck der Entscheidungsfindung stattfindende Verfahren selbst – die Evaluationspraxis – bleiben vorerst unbeobachtet. Wir räumen den Regeln von Evaluationsverfahren also die Funktion ein, einen Laboratorisierungsprozess auszulösen und gewinnen damit eine Folie, auf der wir die unterschiedlichen Rahmenbedingungen einer Evaluationspraxis ordnen und gegebenenfalls divergierende Konstruktionsleistungen bei der Bewertung von Wissenschaft interpretieren können. Dabei geschieht auch die folgende empirische Auseinandersetzung mit den Verfahrensregeln der Leibniz-Gemeinschaft in Deutschland, des Standard Evaluation Protocol for Public Research in den Niederlanden und die Research Assessment Exercises in Großbritannien in dem Bewusstsein, dass eine solche Darstellung immer unvollständig bleiben muss. Auch in der Beschreibung von Laboratorisierungsprozessen werden wiederum Auswahlentscheidungen getroffen. Wir konzentrieren uns auf eine Darstellung der folgenden Aspekte:

- *Verfahrensregeln zur Anpassung des Gegenstands an die kulturellen Akteure:* Wie und inwiefern wird der Gegenstand der Bewertung, also die Komplexität wissenschaftlicher Arbeit, für das Evaluationsverfahren so „zugerichtet“, das es den GutachterInnen möglich ist, sich in einem begrenzten Zeitraum ein gemeinsames Urteil zu bilden? Welche Selektionsentscheidungen werden dabei getroffen, welche Objekte bzw. Leistungen ein- oder ausgeschlossen?⁵
- *Mündlichkeit:* Welche (formelle und informelle) mündliche Interaktion ist zwischen wem vorgesehen beziehungsweise wird durch die Verfahrensregeln ermöglicht?
- *Schriftlichkeit:* Welche Verschriftlichung durch wen ist vorgesehen? Welche (Standardisierungs-)Vorgaben gibt es für Dokumente? Welche Möglichkeiten

⁵ Wir können dabei davon ausgehen, dass sich diese Zurichtung des Bewertungsgegenstandes nur zum Teil als bewusster intentionaler Prozess vollzieht. Vielfältige Selektionen finden unausgesprochen, ungerichtet, vielfach zufällig statt.

informeller/bilaterer schriftlicher Korrespondenz lassen die Verfahrensregeln offen?

- *Körperlichkeit*: Welchen Stellenwert hat der Einbezug von körperlichen/sinnlichen Eindrücken für die Bewertungen in den jeweiligen Verfahrensregeln?

3.1 Evaluationsverfahren der Wilhelm-Leibniz-Gemeinschaft in Deutschland

In Deutschland existiert keine einheitliche Regelung zur Evaluation von Wissenschaftseinrichtungen, da die wissenschaftspolitischen Hauptkompetenzen bei den Bundesländern liegen. Entsprechend werden auch in der Forschungsförderung sehr heterogene Konzepte verfolgt. Darüber hinaus ist das deutsche Wissenschaftssystem durch ein lang tradiertes Prinzip der autonomen Selbststeuerung gekennzeichnet – inhaltliche Eingriffe in die Forschungsplanung und -bewertung sind in hohem Maße rechtfertigungsbedürftig (Braun 1997; Neidhardt 1988; Ritter 1992; Weingart 1998). Für die Evaluierung der Forschungsarbeit an Hochschulen finden sich entsprechend bis heute nur wenig übergreifende Verfahren, obwohl auf Ebene einzelner Bundesländer Konzepte zur kombinierten Lehr- und Forschungsevaluierung längst etabliert sind⁶.

Bereits seit längerer Zeit von bundesweiter Bedeutung und über die Jahre systematisch weiter entwickelt worden sind jedoch Evaluationsverfahren für außeruniversitäre Forschungseinrichtungen. Eine gewisse Tradition haben hier insbesondere die Verfahren zur Evaluierung von Instituten der Wissenschaftsgemeinschaft Wilhelm Gottfried Leibniz (WGL). Die WGL wurde 1997 als Nachfolgeorganisation der „Wissenschaftsgemeinschaft Blaue Liste“⁷ gegründet und ihr gehören derzeit 82 Forschungsinstitute an, die von Bund und Ländern gemeinsam finanziert werden. Einrichtungen der Blauen Liste werden bereits seit 1979 im Auftrag von Bund und Ländern regelmäßig evaluiert, zunächst vom Wissen-

⁶ Beispielsweise wurde im Jahre 1995 für Niedersachsen die Zentrale Evaluations- und Akkreditierungsagentur Hannover (ZEVA) eingerichtet, die einen regelmäßigen Evaluationszyklus für alle niedersächsischen Hochschulen umsetzen soll (<http://www.zeva.org>; vgl. übergreifend auch Barz 1998, zur indikatorbasierten Finanzierung von Hochschulen auch Leszczensky et al. 2004a; Orr et al. 2007)

⁷ Der Name geht zurück auf eine auf blauem Papier veröffentlichte Liste von damals 46 Instituten, für die im Jahre 1977 eine gemeinsame Förderung von Bund und Ländern verabredet wurde.

schaftsrat, der das Verfahren im Jahre 2000 ausgewertet und anschließend einem durch WissenschaftlerInnen besetzten „Senat der Leibniz-Gemeinschaft“ übertragen hat. Dieser hat wiederum einen Senatsausschuss Evaluierung (SAE) eingerichtet, der seither für die Gestaltung der Evaluationsverfahren zuständig ist. Neben einigen Mitgliedern des Senats gehören dem SAE Wissenschaftlerinnen und Wissenschaftler an, die nicht an Leibniz-Einrichtungen arbeiten, sowie Vertreterinnen und Vertreter von Bund und Ländern als Finanzgeber der Einrichtungen. Administrativ werden Senat und Senatsausschuss Evaluierung bei Fragen der Evaluierung durch das ihnen unmittelbar unterstellte „Referat Evaluierung“ unterstützt.

Jedes Mitgliedsinstitut der WGL wird in der Regel⁸ alle sieben Jahre evaluiert. Für das Bewertungsverfahren wird vom SAE ein Gremium von FachgutachterInnen zusammengestellt, deren Mitglieder nach den Verfahrensregeln keinen Kooperationskontakt zu dem Institut haben dürfen und gegen die keine anderweitigen Befangenheitsbedenken bestehen. Folglich kann auch das zu bewertende Institut potentielle GutachterInnen aus Befangenheitsgründen ablehnen. Die Leitung der Begutachtungsgruppe wird von einem Mitglied des SAE wahrgenommen, der/dem formal auch die Auswahl der GutachterInnen obliegt. Weitere Kriterien für die Berufung sind nicht schriftlich fixiert. Aus Gesprächen mit verantwortlichen ReferentInnen der WGL erfahren wir, dass versucht wird, die fachliche Breite eines Instituts durch eine entsprechende Kompetenzverteilung in der Kommission abzudecken und dass außerdem Erfahrung in der Evaluation eine Rolle spielt. Vielfach erkundigen sich die zuständigen ReferentInnen daher bei der DFG oder beim Wissenschaftsrat nach geeigneten Kandidaten oder Kandidatinnen.

Kern des Bewertungsprozesses ist ein Vor-Ort-Besuch des Instituts, die so genannte „Begehung“. Zu dessen Vorbereitung werden den FachgutachterInnen Unterlagen über die zu evaluierende Einrichtung zur Verfügung gestellt, die Letzere auf Basis eines von der WGL vorgegebenen Fragerasters erstellt hat. Die Unterlagen enthalten Angaben zu den Aufgaben des Instituts, seiner Größe und Struktur, zur Mittelausstattung, zu den Arbeitsschwerpunkten sowie zu den zentralen Ar-

⁸ Das Ergebnis eines Bewertungsverfahrens kann allerdings vorsehen, dass eine nächste Evaluation bereits früher stattfinden muss.

beitsergebnissen (Publikationen, Konferenzen, Drittmittel). Die Begehung dauert eineinhalb Tage und beinhaltet eine Präsentation der Institutsleitung gegenüber der gesamten Begutachtungsgruppe, arbeitsteilig organisierte Abteilungsbesuche, ein Gespräch mit den MitarbeiterInnen unter Ausschluss der Führungskräfte, ein Gespräch mit der Verwaltungsleitung und Gespräche mit vom Institut ausgewählten KooperationspartnerInnen. Neben den FachgutachterInnen sind an der Begehung auch VertreterInnen von Bund und Ländern beteiligt.

Spezifische Formen der Zurichtung

Die Laboratorisierung des Bewertungsgegenstands erfolgt auf verschiedenen Ebenen. Eine Selektionsebene betrifft die Festlegung der zu bewertenden Einheit, die in diesen Verfahren ein Institut sein soll. Vorstellbar wären genauso gut kleinere oder größere Einheiten, etwa könnte die wissenschaftliche Arbeit von Individuen zum Bewertungsgegenstand gemacht werden oder die von Arbeitsteams, die aufgrund von Kooperationen aber auch über die Grenzen eines Instituts hinausgehen könnten. Die Entscheidung für ein einzelnes Institut als Bewertungseinheit entspricht der Logik der institutionellen Förderung von Wissenschaftseinrichtungen und scheint daher aus politischer Sicht plausibel zu sein; für die wissenschaftliche Bewertung ist sie jedoch keineswegs zwingend.

Eine weitere Ebene der Zurichtung betrifft die wissenschaftliche Arbeit, deren Komplexität für das Bewertungsverfahren so reduziert werden muss, dass sie von der Bewertungsgruppe innerhalb von eineinhalb Tagen erfassbar und außerdem mit den Leistungen anderer Einrichtungen vergleichbar ist.⁹ Nicht die ganze Arbeit wird folglich evaluiert, sondern es werden Auswahlentscheidungen getroffen. In der Regel finden diese Entscheidungen anhand der Organisationsstruktur eines Institutes statt. Es präsentieren sich einzelne Abteilungen und in der konkreten Darstellung wird oftmals ein besonderer Fokus auf Projekte gelegt. Dann gilt es auszuwählen, welche Projekte präsentiert werden sollen und welche Schwerpunkte dabei zu setzen sind.

Mit der Entscheidung für den eineinhalbtägigen Vor-Ort-Besuch in den Verfahrensregeln der WGL-Evaluationen ist also gleichzeitig eine Entscheidung darüber

⁹ So haben Vorgaben bezüglich der Bewertungseinheit sowie auch die Abfrage von diversen statistischen Daten wie Stellen in Vollzeitäquivalenten auch den Zweck, ein Institut in Relation zu anderen zu bewerten.

getroffen worden, dass nur Teile der wissenschaftlichen Arbeit des Instituts näher betrachtet werden. Des weiteren ist davon auszugehen, dass diese Selektionen antizipierten Erwartungen folgen und somit wissenschaftliche Arbeiten, die diesen ‚Erwartungserwartungen‘ nicht oder weniger entsprechen, nicht zur Begutachtung vorgelegt werden. Diese Selektionsleistung beginnt bereits mit der schriftlichen Selbstdarstellung. Deren Aufbau und Schwerpunkte sind zwar durch die Verfahrensregeln vorgegeben, doch in der Umsetzung erfolgt stets eine inhaltliche Akzentuierung durch das Institut. Weitere Selektionsleistungen finden sich in der gewählten Form der mündlichen Präsentation. So macht es einen Unterschied, ob man sich beispielsweise für eine Poster-Präsentation entscheidet, bei der lediglich eine komprimierte Darstellung der Arbeit möglich ist, oder für einen Vortrag, der mehr Raum für Ausschmückungen lässt. Auf der anderen Seite ermöglicht das Poster eine längere ‚Sichtbarkeit‘ der Arbeit, während das gesprochene Wort schneller vergänglich ist. Doch für welche Form der Präsentation man sich entscheidet: So oder so muss die Arbeit mehrerer WissenschaftlerInnen über mehrere Jahre in wenigen Sätzen zusammengefasst werden, für die man sich eine Vielfalt von unterschiedlichen Akzentuierungen vorstellen kann. Auszuwählen ist überdies, *wer* die ausgewählten Inhalte während der Begehung präsentiert und gegebenenfalls auf Rückfragen der GutachterInnen antwortet.

Auch wenn das Verfahren der WGL die Selektionen auf der Ebene der wissenschaftlichen Arbeiten weitgehend dem Institut überlässt, ändert dies nichts am prinzipiellen Zurichtungscharakter der erforderlichen Selektionen, mit denen gleichsam eine operationalisierbare Bewertungsmasse hergestellt werden soll. Denn mit jeder dieser Entscheidungen wird die „Natur“ des wissenschaftlichen Produktionsalltages im Sinne des Laboratorisierungskonzepts so in ihrer Komplexität reduziert, dass sie im Rahmen der eineinhalbtägigen Begehung in Augenschein genommen und bewertet werden kann. Zugleich können wir annehmen, dass für jede dieser Entscheidungen alternative Optionen zur Verfügung stehen, die zu anderen Gewichtungen und damit möglicherweise zu anderen Urteilen führen würden.

Eine weitere Form der Zurichtung ist die Auswahl des Bewertungsgremiums. Aufgrund der Vorschrift, dass die FachgutachterInnen keine Kooperationsbeziehung zum Institut haben dürfen, setzt sich dieses Gremium in den Verfahren der WGL zumeist aus Personen zusammen, die nicht direkt vom Fach sind, sich zum Teil im Rahmen des Verfahrens erstmals begegnen und die wissen, dass die ge-

meinsame Arbeit in dieser Zusammensetzung zeitlich eng begrenzt ist. Ein weiterer Teil der Bewertungsgruppe hat den Mitgliedschaftsstatus „qua Amt“; VertreterInnen des Bundes- und Landesministeriums, durch die das Institut gefördert wird. Durch ihre Teilnahme fließen unter Umständen auch staatliche Interessen in die Bewertung ein. Die Vorsitzenden von Beiräten des Instituts sind ebenfalls Teil der Runde. Die divergente Interessenlage im Bewertungsgremium sowie die unterschiedliche Nähe der GutachterInnen zum Bewertungsgegenstand lässt sehr heterogene Wahrnehmungen der Institutsleistungen erwarten – auch ein Aspekt, der für die Dynamik der kooperativen Urteilsbildung von Bedeutung sein dürfte.

Mündlichkeit, Körperlichkeit und Schriftlichkeit

Auch wenn alle GutachterInnen sich anhand der schriftlichen Unterlagen individuell vorbereiten, stellt der Vor-Ort-Besuch des Instituts das Kernstück des Bewertungsprozesses dar. Dieses wird auch seitens der GutachterInnen so gesehen:

„Ne, die Begehung ist sehr wichtig. Also ich denke, das-, ich sage mal 50 Prozent der Institut werden vorher schlechter eingeschätzt.“

Interviewerin: „Ah ja, und was macht es dann-, was macht es dann vor allem aus?“

„Naja erstens sind die Gutachter, das ist doch auch alles ein menschlicher Prozess. Das muss man ja alles genau im Detail lesen, ob die jetzt die Zeit haben, das alles zu machen, ist die Frage. Ich nehme mal an, die machen's, aber trotzdem ist manch-, wird öfters gesagt, dass anhand der schriftlichen Lage das anders eingeschätzt wurde. Also das spielt schon-, das Verfahren ist schon gut. Man kriegt ja auch Meinungen auch von anderen Gutachtern, manchmal ist man sich doch unsicher, wenn man jetzt ein Fachgutachter ist, die kriegt man dort, dann kriegt man-, geht man im Institut rum, sieht, wie das aussieht, sieht die Leute, redet mit denen. Das ist schon sehr wichtig die Begehung.“ (Gutachter Institut B-D)

Die Aussage dieses Gutachters zeigt, dass damit auch Mündlichkeit in den Verfahren der WGL in mehrfacher Weise ein wichtiges Instrument darstellt. Denn mit der herausgehobenen Rolle des Vor-Ort-Besuches bekommt sowohl der Dialog zwischen GutachterInnen und Begutachteten als auch die Interaktion im Kreis der GutachterInnen eine erhebliche Bedeutung. So besteht im Rahmen der Begehung die Möglichkeit, sowohl mit der Leitung als auch den MitarbeiterInnen des Instituts unmittelbar zu kommunizieren. Sachverhalte können somit im Gespräch vertieft werden, Unverständliches geklärt oder auch Themen zusätzlich in die Betrachtung aufgenommen werden, die in der Selbstdarstellung des Instituts bisher nicht vorgesehen waren. Wir können uns vorstellen, wie dabei neben der sachorientierten Mündlichkeit auch Stimme und sprachliche Eloquenz und Souveränität einen Einfluss auf den Urteilsprozess haben kann.

Eine unvermeidliche Folge des hohen Stellenwerts von Mündlichkeit im Bewertungsverfahren der WGL sind jedoch weitere Beschränkungen hinsichtlich der möglichen Bewertungsobjekte. So können nicht alle GutachterInnen mit allen MitarbeiterInnen gleichzeitig kommunizieren, sondern es muss eine Auswahl getroffen werden. Folglich findet nur das Gespräch mit der Leitung des Instituts im Beisein aller GutachterInnen statt. Der Besuch der einzelnen Arbeitseinheiten erfolgt hingegen arbeitsteilig, so dass jeweils nur ein Teil der GutachterInnen mit einem Teil der MitarbeiterInnen ins Gespräch kommt. Ebenso können die GutachterInnen in der so genannten „kopflosen“ Runde, in der sich die MitarbeiterInnen ohne ihre Vorgesetzten dem Gespräch stellen, jeweils nur mit einer kleinen Auswahl an Mitarbeiterinnen oder Mitarbeitern sprechen.

Die zweite Ebene der Mündlichkeit betrifft die Interaktion zwischen den GutachterInnen. Dazu sieht das Verfahren mehrere Runden vor. Der Bewertungsvorgang startet und endet mit einer internen Besprechung der FachgutachterInnen und Bund-/LändervertreterInnen, die von der oder dem Vorsitzenden der Bewertungsgruppe moderiert wird. In der Eröffnungssitzung werden von dieser oder diesem oder der Geschäftsstelle Evaluierung die Bewertungskriterien und spezifische Anforderungen an den Bewertungsprozess vorgestellt. In der Abschlussbesprechung wird eine erste Bilanz gezogen, indem alle GutachterInnen ihren Eindruck von der eineinhalbtägigen Begehung mitteilen. Am ersten Abend der Begehung wird außerdem im Kreis der Bewertungsgruppe eine Zwischenbilanz gezogen. Dabei werden individuelle Eindrücke über die Begehung ausgetauscht und zu ersten Bewertungsaussagen verdichtet, die dann unter Umständen von den anderen kommentiert werden. Darüber hinaus ermöglichen zahlreiche informelle Situationen mündliche Interaktionen vielfältiger Art. Auffällig sind etwa zahlreiche Schilderungen der GutachterInnen über die Gespräche beim gemeinsamen Abendessen, das den ersten Begehungstag abschließt. Hier werden ohne festgelegte Struktur Einschätzungen ausgetauscht und die eigene Positionierung im Verhältnis zu den anderen GutachterInnen überprüft:

"Ja, ja ich meine man bespricht schon bei so einem Abendessen die ganze Stimmung noch mal und auch, was einem gut gefallen hat, was weniger gut und dadurch vertiefen sich natürlich auch die Dinge noch einmal in anderer Art und Weise, weil man noch mal auch andere Standpunkte dann hören kann oder hört dann einfach, die sind wie gesagt nicht so schriftreif dann sind und das hilft glaub' ich, das ist schon sehr hilf-, für jeden hilfreich, einmal um noch mal zu reflektieren, anders zu reflektieren. Also ist so ne Art Reflektion und ich meine dann ist es natürlich-, das ist superschwer zu fassen, dass man dann mit jemandem, wenn man dann nach-, von dem Restaurant sind wir dann in ganz kleinen Gruppen zum Hotel gegangen, natürlich, dass man das nicht so ganz offen sagt

oder ist das-, aber ich meine, man kann-, so zwischenmenschliche Sachen kann man ja nicht, also so ein kleines Gespräch, auch was dann nach-, so zwischendurch mal geführt hat. Es ist natürlich-, offiziell geht das nie in die Bewertung ein, aber es ist trotzdem, man fragt dann noch mal nach, wie ist denn das, kennst du nicht doch noch mehr Sachen oder was-, was hast du sonst für einen Eindruck von denen oder so was, dann kriegt man natürlich auch noch mehr Informationen darüber." (Gutachterin, Institut D-D)

Schon dadurch, dass im WGL-Verfahren das Instrument des Peer Review verwendet wird und nicht rein metrische Verfahren über die Güte eines Instituts entscheiden, erscheint es evident, dass im Bewertungsprozess immer auch spezifisches an die Person gebundenes Wissen aufgerufen wird. Neben dem hohen Stellenwert der Mündlichkeit ist damit auch der Gutachterkörper als Messinstrument sowie als Depot und Archiv von Erfahrung gefragt. Durch die Vor-Ort-Begehung wird der Gutachterkörper zudem *in situ* zum Einsatz gebracht und an den Ort verfrachtet, dessen Güte durch ihn beurteilt werden soll. Indem die Bewertung etwa in den Räumen des Instituts stattfindet, werden alle sinnlichen Eindrücke dieses Tages davon zum „Gegenstand“ der Bewertung; Gebäude, Raum, Licht, Farben – all dies spielt ebenso eine Rolle wie der Eindruck, den die MitarbeiterInnen des Instituts vermitteln. Weil sie ihre Arbeiten und deren Ergebnisse nicht allein schriftlich einreichen, sondern persönlich präsentieren, werden sie zu körperlich erfahrbaren Objekten – ein Sachverhalt, der in der Vorbereitung einer Evaluation auch seitens der zu evaluierenden Einrichtung reflektiert wird, wie uns folgendes Zitat zeigen kann:

„Also ich habe dann z.B. auch aufgeschrieben, das war aber jetzt nichts, was irgendwie die Gutachtergruppe gesagt hat, also dass vielleicht an dem Tag nicht unbedingt alle Fahrräder vor der Haustür angelehnt sein sollten, sondern dass man reingehen kann oder irgendwie solche Sachen oder dass man meinetwegen, manche Leute kommen ja auch aus anderen Häusern jetzt zu diesem Platz, dass man rechtzeitig da sein sollte auf jeden Fall. Also da gibt's bei uns auch immer so ein paar Pappenheimer [lacht] und dass man die Jacken dann irgendwie in den Büros lassen sollte und nicht irgendwie so ne Bahnstimmung erzeugen sollte, ne gewisse Gemütlichkeit ist irgendwie, also schafft schon auch ne andere Atmosphäre und lauter solche Dinge.“ (Mitarbeiterin Institut B-D)

Und seitens eines Gutachters wird bestätigt, wie sinnliche Aspekte – gewollt oder nicht – im Rahmen einer Begehung wahrgenommen werden:

„(...) und daher kenne ich diese Drucksituation, die damit (mit einer Evaluation, die Verf.) verbunden ist und weiß eben auch, wie man sozusagen diese-, dass ein Teil des Spiels ist die schöne Fassade, keine Druckfehler, schön formatiert, einheitliche Poster, einheitlich Corporate Identity, *keiner redet länger als er soll* [betont], alle sind furchtbar dankbar und freundlich. Es gibt so einen positiven Grundton und bis hin zur Serviettenfarbe und weiß der Teufel, über was sich die Leute dann anfangen Gedanken zu machen. Und *das ist* [betont] die Fassade, die aufgebaut wird, die unheimlich viel Energie kostet, die *aber erwartet wird, wenn man das nicht* [betont] tut, gilt man als der Stoffel, das weiß ich, aber das ist natürlich nicht das, die Fähigkeit, eine Fassade aufzustellen, ist nicht das,

was ich ja eigentlich bewerten soll als Evaluierer. Natürlich die Tatsache, dass die Leute das tun, zeigt, dass sie die Sache insgesamt ernst nehmen, aber wenn ihnen die Frage gestellt wird, überlebt ihr ja oder nein, natürlich nehmen sie das dann ernst.“ (Gutachter Institut D-C)

Neben Mündlichkeit und Körperlichkeit spielt allerdings auch Schriftlichkeit im WGL-Verfahren eine wichtige Rolle. Den ersten Eckpfeiler bildet die schriftliche Selbstdarstellung des Instituts, die anhand eines von der WGL vorgegebenen Fragenkatalogs zu Beginn des Verfahrens erstellt wird. Sie dokumentiert die Leistungen des Instituts aus der Perspektive der Selbstwahrnehmung und dient zur Vorbereitung der GutachterInnen auf den Vor-Ort-Besuch. Dabei ist davon auszugehen, dass einzelne GutachterInnen auf dieser Basis bereits eine Vorbewertung des Instituts vornehmen; zumindest bilden sie sich, wie die Interviews zeigen, bei der Sichtung des Materials eine erste Meinung und treffen eine Vorauswahl, worauf sie bei der Begehung besonders achten wollen. Von der Art und Weise wie die schriftliche Selbstdarstellung erstellt ist hängt somit bereits zu einem gewissen Ausmaß der Ausgang des gesamten Verfahrens ab.

Ein weiteres wichtiges schriftliches Dokument sind die Protokolle der Besprechungen der GutachterInnen vor Ort, die von einer Mitarbeiterin bzw. einem Mitarbeiter des Evaluationsreferats der WGL angefertigt werden. Sie sind eine wichtige Grundlage des Bewertungsberichts. Zweite Grundlage sind die schriftlichen Statements der GutachterInnen, die diese innerhalb von drei Wochen im Anschluss an die Begehung abzugeben haben. Erstellt wird der Bewertungsbericht von dem bzw. der zuständigen MitarbeiterIn der WGL. Anschließend wird er mit dem bzw. der Vorsitzenden der Bewertungskommission abgestimmt und dann an das Begutachtungsgremium versandt. Die GutachterInnen haben noch die Möglichkeit, Änderungsbedarfe zu formulieren, bevor der Bericht vom Senatsausschuss der Leibnizgemeinschaft zusammen mit einer eigenen Stellungnahme zur Frage der Weiterförderung des Instituts verabschiedet und an die Bund-Länderkommission gegeben wird. Neben der generellen Empfehlung zur weiteren Förderfähigkeit der Einrichtung kann der Bericht darüber hinaus Empfehlungen zur inhaltlichen Ausrichtung, zur Arbeitsweise und Struktur und gegebenenfalls auch zu Budgetveränderungen enthalten. Vor der offiziellen Verabschiedung des Berichts durch den Senat der Leibnizgemeinschaft bekommt die Institutsleitung den Bericht vertraulich zugesandt und wird aufgefordert, eine Stellungnahme zur Bewertung und zu den Empfehlungen abzugeben.

Schriftlichkeit hat somit im Verfahren der WGL eine Dokumentations- und eine Kontrollfunktion. Dokumentiert werden zum einen die Entwicklung des Instituts in Gestalt der Selbstdarstellung, zum anderen die einzelnen Statements der GutachterInnen im Verlauf des Verfahrens in Form von Protokollen sowie mittels ihrer eigenen schriftlichen Stellungnahme nach der Begehung. Zum anderen dokumentiert der Bewertungsbericht nach offizieller Verabschiedung durch die Gremien der WGL die Gesamtbewertung des Instituts und bildet somit auch einen wichtigen Maßstab für die nächste Evaluation. Zugleich haben sowohl die schriftliche Selbstdarstellung des Instituts als auch der Bewertungsbericht eine gewisse Kontrollfunktion. Anhand der schriftlichen Selbstdarstellung können die GutachterInnen beispielsweise erkennen, wie sich das Institut selber sieht und seine Leistungen darstellt, und der Bewertungsbericht ermöglicht ihnen zu kontrollieren, ob ihr Statement sich in der Gesamtbewertung angemessen wieder findet. Ebenso können die Institute anhand des Bewertungsberichtes überprüfen, inwieweit sich ihre Selbstsicht mit der Fremdwahrnehmung der GutachterInnen deckt.

Gleichzeitig können wir uns auf der Grundlage dieser Regeln zur Bedeutung von Schriftlichkeit mehrfache Übersetzungsleistungen vorstellen, die in der Dynamik der einzelnen Verfahren entstehen und eine je eigene Konstruktionsleistung hervorbringen. Besonders gilt dies für den Bewertungsbericht. So haben die MitarbeiterInnen der WGL aus den heterogenen Einzelbewertungen der GutachterInnen ein Gesamtwerk zu komponieren, das unter den Gremiumsmitgliedern möglichst keinen Widerspruch mehr erzeugt. Denn in dieser Phase des Verfahrens ist nur noch ein schriftlicher Abstimmungsprozess vorgesehen, was sich für die Klärung von Dissensen innerhalb einer Gruppe nicht gut eignet. Außerdem muss der Bericht so verfasst sein, dass er dem begutachteten Institut ebenso wie den Adressaten in der Politik handhabbare Empfehlungen gibt, was wiederum zu Irritationen führen kann, weil derlei Kompromissformeln nicht unbedingt eins zu eins in die Logiken des jeweiligen Systems übersetzt werden können. Während der Inhalt eines Bewertungsberichts in den Verfahren der WGL somit einerseits als Ausdruck eines künstlich erzeugten Konsenses gelesen werden kann, scheint er andererseits das strukturelle Potential des Widerspruchs in sich eingeschrieben zu haben – je nach dem, von wem er gelesen wird.

3.2 Evaluationsverfahren des “Standard Evaluation Protocol for Public Research“ in den Niederlanden

Das Standard Evaluation Protocol for Public Research (SEP) wurde im Jahre 2003 gemeinsam von der Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), der Vereniging van Universiteiten (VNSU) und der Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) verfasst und bildet seither die Grundlage für die alle sechs Jahre stattfindende Evaluation der öffentlich finanzierten Forschungseinrichtungen (KNAW et al. 2003). Das SEP fokussiert hauptsächlich die Forschung und im Vorwort wird darauf hingewiesen, dass den unterschiedlichen Ausrichtungen der Institute zum Beispiel durch die Auswahl der GutachterInnen Rechnung getragen werden soll (ebd.: 6). Hauptziel des SEP ist es, durch Evaluationen „Improvement and accountability“ (ebd.: 7) von wissenschaftlichen Einrichtungen zu erhöhen – insbesondere die Rechenschaftspflicht von öffentlich finanzierter Forschung wird betont. Mit dem SEP wird nicht allein die *wissenschaftliche Leistung*, sondern auch die Qualität der Institutsleitung (Führung, Strategie, Instrumente und Forschungsorganisation) bewertet. Für die Bewertung der Forschungseinrichtungen sind insgesamt vier Kriterien vorgesehen: Qualität, Produktivität, Relevanz sowie Dynamik und Realisierbarkeit. Das zu evaluierende Institut wird auf einer Skala von 5 (excellent) bis 1 (unsatisfactory) nach den genannten Kriterien bewertet. In der Anwendung dieser Kriterien soll disziplinären Unterschieden Rechnung getragen werden. Betont wird, dass der Arbeitsaufwand für die Evaluation insgesamt gering gehalten und Synergien genutzt werden sollen.

Ausgangspunkt der Bewertung der einzelnen Forschungseinrichtungen ist ein Selbstevaluationsbericht, den die Institute selber zu erstellen haben. Die Selbstevaluation (mid-term) soll jeweils drei Jahre vor oder nach der externen Evaluation stattfinden und einerseits die Umsetzung der Empfehlungen der letzten Evaluation überprüfen sowie andererseits die kommende Evaluation vorbereiten. Außerdem sind die Forschungseinrichtungen angehalten, ein Monitoringsystem zu etablieren, das auch online verfügbar ist und im jährlichen Rhythmus wesentliche Kennziffern über die Entwicklung des Instituts bereitstellt. Die Dokumente des Selbstevaluationsberichtes sind nach Maßgabe des SEP in zwei Teile unterschieden: Der erste Teil (A) charakterisiert das Institut als Ganzes und beschreibt die

strategische Ausrichtung des Institutes, deren organisatorische Umsetzung und Ergebnisse im Überblick.¹⁰ Der zweite Teil (B) stellt die Forschungsprogramme und Forschungsgruppe vor. Beide Teile des Berichts sollen jährlich verfasst werden, so dass auch Veränderungen über den Berichtszeitraum deutlich werden.

Neben der Textexegese des Selbstevaluationsberichts findet ähnlich den WGL-Verfahren eine in der Regel ein bis zweitägige Begehung (site visit) des Institutes durch eine Begutachtungsgruppe statt. Diese besteht aus einem oder einer Vorsitzenden (Chair), drei GutachterInnen und einem Sekretär bzw. einer Sekretärin des KNAW. Bei ihrer Auswahl ist im SEP im Unterschied zu den Regularien der WGL ein Vorschlagsrecht durch die zu evaluierende Einrichtung explizit vorgesehen. So sind die Trägereinrichtungen z.B. aufgefordert, jeden Vorschlag zur Besetzung der Bewertungskommission sorgfältig zu prüfen. In der Regel wird dazu von den Instituten eine „long list“ an Vorschlägen erstellt, welche die Verantwortlichen der KNAW zu einer „short list“ verdichten. Nur in Ausnahmefällen werden zu der Liste zusätzliche Personen hinzugefügt und eingeladen. Die Einladung der GutachterInnen selbst erfolgt durch das KNAW.

Die Begehung beginnt wie in dem Verfahren der WGL in der Regel mit einer nichtöffentlichen Sitzung der GutachterInnen. Der weitere Ablauf der Begehung kann allerdings stark variieren. Anders als im Falle der WGL wird kein einheitliches Protokoll umgesetzt, sondern in einem Gespräch zwischen KNAW und dem zu begutachtenden Institut ein paar Monate vor dem Termin der genaue Zeit- und Ablaufplan entwickelt. Am Ende der Begehung steht jedoch stets eine gemeinsame nicht öffentliche Runde der GutachterInnen, in der die Eindrücke der einzelnen Mitglieder der Bewertungsgruppe gesammelt und ein erstes Bewertungsergebnis entwickelt wird. Seitens der KNAW ist für diesen Prozess eine Art Checkliste entwickelt worden, in der für die genannten vier Kriterien Indikatoren aufgelistet sind, die in dieser Runde systematisch abgearbeitet und dann zu einem Gesamtergebnis zusammengefasst werden können. In der Regel wird den Beschäftigten des Instituts im Rahmen einer Vollversammlung im Anschluss an die Sitzung der Begutachtungsgruppe bereits ein erster Bewertungseindruck mitgeteilt.

¹⁰ Detaillierte Beschreibungen befinden sich im Appendix 3 des SEP (KNAW et al. 2003: 27ff).

Die GutachterInnen selbst geben nach der Begehung eine schriftliche Stellungnahme ab. Diese Stellungnahmen und die Aufzeichnungen des Sekretärs zur Begehung bilden die Grundlage für den Bewertungsbericht, der wiederum durch den Sekretär/ die Sekretärin vom KNAW verfasst wird. Die GutachterInnen haben dann die Möglichkeit, den Berichtsentwurf zu kommentieren. Anschließend wird der Bericht dem Institut zur Verfügung gestellt. Der wissenschaftliche Beirat des Instituts und auch das KNAW geben dann zu dem fertigen Bericht eine Stellungnahme ab.

Eine Besonderheit gegenüber dem Verfahren der WGL stellt die Möglichkeit dar, der Institutsleitung in einem *Management Letter* vertrauliche Informationen zu vermitteln. Von dieser Möglichkeit wird zwar selten Gebrauch gemacht, sie wird aber beispielsweise dann angewandt, wenn die Gutachtergruppe es für angemessen hält, eine Stellungnahme zu konkreten Personen und Personalangelegenheiten abzugeben. Im Evaluationsbericht selbst sind diese Informationen nicht enthalten.

Spezifische Formen der Zurichtung

Evaluationsverfahren nach dem SEP sind denen der WGL sehr ähnlich. Damit lassen sich aus der Perspektive des Laborkonstruktivismus ähnliche Prozesse der Zurichtung des Gegenstandes identifizieren, etwa durch die Festlegung der Evaluationseinheiten (Institute und dort spezifische Arbeitseinheiten und Forschungsprojekte), durch den sechsjährigen Evaluationsrhythmus und durch den hohen Stellenwert der Vor-Ort-Begehung. Ebenso prägen auch hier die Zusammensetzung des Begutachtungsgremiums, der spezifische Formalisierungsgrad von Mündlichkeit und Schriftlichkeit sowie die relativ große Bedeutung von Körperlichkeit als Messinstrument maßgeblich den Charakter dieser Verfahren.

Über diese Gemeinsamkeiten hinaus sind aber auch Unterschiede feststellbar, die für den Prozess der Urteilsbildung relevant sein können. Von Bedeutung scheint etwa der Selbstevaluationsbericht. Während bei Verfahren der WGL einmalig ein Bericht verfasst wird und es dann – bis zur nächsten Evaluation nach sieben Jahren – dem Institut überlassen bleibt, wie es mit dem Bericht umgeht, findet mit dem vorgeschriebenen Selbstevaluationsverfahren nach dem SEP gleichsam ein

kontinuierlicher „Zurichtungsprozess“ statt¹¹. Die Entscheidungen, welche Arbeit für die Evaluation als relevant und möglicherweise dem gewünschten Profil entsprechend eingeschätzt werden, sind somit nicht nur alle sieben Jahre zu treffen, sondern vor dem Hintergrund des jährlichen Selbstmonitoring und der auf dieser Basis nach drei Jahren stattfindenden Zwischenevaluation mehr oder weniger kontinuierlich. Denn mit dieser Vorgabe ist die implizite Annahme verbunden, dass zur Beurteilung einer wissenschaftlichen Einrichtung nicht nur die Leistung zu einem Zeitpunkt X ausschlaggebend ist, sondern auch die Entwicklungsdynamik.¹² Deutlich wird dies nicht zuletzt auch daran, dass der Leistungsentwicklung eines Instituts ein eigenes Kriterium gewidmet ist (Dynamik und Realisierbarkeit) – bewertet werden soll etwa, inwiefern ein Institut in der Lage ist, Forschungsgruppen dem aktuellen Bedarf entsprechend zu eröffnen oder zu schließen (vgl. KNAW et al. 2003: 11). Stärker noch als in den Verfahrensregeln der WGL wird somit in den Verfahrensregeln des SEP ein Zustand geschaffen, bei dem die externe Evaluation nur noch als Teil einer „Kontrolle der Kontrolle“ fungiert (Power 1997). Damit ist anzunehmen, dass auch der Zurichtungseffekt dieser Regeln weit über das eigentliche Verfahren hinausgeht.

Zwei weitere markante Unterschiede zum Verfahren der WGL betreffen die Zusammensetzung der Begutachtungskommission. Zum einen sind in den Verfahrensregeln des SEP ausdrücklich Einflussmöglichkeiten der Institute auf die Auswahl der GutachterInnen vorgesehen. Zum anderen sind an den Kommissionen nach dem SEP keine wissenschaftspolitischen VertreterInnen beteiligt. Zwar wird jedes Gremium ähnlich wie bei der WGL von einem Sekretär oder einer Sekretärin der Trägerorganisation organisatorisch betreut, die Artikulation von dezidiert politischen Interessen ist aber im Evaluationsverfahren nicht vorgesehen.

Die Institute gestalten darüber hinaus das konkrete Verfahren aktiver mit als im Falle der WGL, etwa indem der zeitliche Ablauf der Begehung und die Definition

¹¹ Wenn auch nicht im offiziellen Verfahren vorgeschrieben, ist auch bei den Instituten der WGL zu beobachten, dass zusätzlich zu den Evaluationen in Eigenregie Begehungen durch den wissenschaftlichen Beirat durchgeführt werden. Einige Institute nutzen dieses Beiratsaudit explizit zur Vorbereitung auf die WGL-Evaluation.

¹² Dieser Aspekt ist grundsätzlich zwar auch in den Verfahren der WGL angelegt, wird nach dem SEP aber deutlicher betont und formalisiert. Auch hier ist das Institut aufgefordert darzustellen, welche Konsequenzen sie aus der letzten Evaluation (vor 7 Jahren) gezogen hat. Darüber hinaus werden eine Reihe von Daten in der Selbstdarstellung als Zeitreihe erfragt.

der zu treffenden Gruppe nicht zentral durch die Trägerorganisation vorgegeben sind, sondern im Gespräch mit den Instituten entwickelt werden. Damit haben die verantwortlichen InstitutsvertreterInnen auch größeren Einfluss darauf, was in welcher Form von den GutachterInnen wahrgenommen und bewertet werden kann und was nicht.

Die Formen der Zurichtung auf den Prozess der Evaluation werden somit im SEP-Verfahren mehr von der zu evaluierenden Einrichtung mitbestimmt als bei der WGL. Gleichzeitig ist nach dem SEP die Aufgabe der Zurichtung quasi auf Dauer gestellt. Mit der Vorgabe der jährlichen Selbstberichterstattung sowie der Mid-term-Selbstevaluation wird die Evaluation gleichsam zum Monitoring-Instrument. Die GutachterInnen in den Verfahren des SEP bewerten unter einem hohen Zeitdruck. Dieser wird dadurch ausgelöst, dass der versammelten Belegschaft einer evaluierten Einrichtung bereits am Ende des zweiten Bewertungstages ein erstes Bewertungsergebnis präsentiert werden muss. Damit wird sehr frühzeitig eine Öffentlichkeit zu den Evaluationsergebnissen geschaffen, die spätere Abweichungen davon legitimationsbedürftig machen. Der Prozess der Konsensfindung im Gutachtergremium muss somit bereits während des Vor-Ort-Besuchs weitgehend zum Abschluss gebracht werden.

Mündlichkeit, Körperlichkeit und Schriftlichkeit

Wie in den Verfahren der WGL stellt der Vor-Ort-Besuch auch in den Verfahren nach dem SEP den Kern des Bewertungsprozesses dar, wodurch sowohl Mündlichkeit als auch Körperlichkeit in diesen Verfahren einen vergleichbaren Stellenwert bekommen. Ähnlich verhält es sich mit der Rolle von Schriftlichkeit, über die sowohl die Entwicklung des zu bewertenden Instituts als auch der Verlauf des Verfahrens und die Bewertungen der einzelnen GutachterInnen dokumentiert werden. Ob ihre Bedeutung als Kontrollfunktion in diesen Verfahren angesichts der jährlich zu veröffentlichenden Kennziffern ein etwas stärkeres Gewicht hat, ist empirisch zu prüfen.

3.3 Evaluationsverfahren des „Research Assessment Exercises“ (RAE) in Großbritannien

Die Research Assessment Exercises finden 2008 zum sechsten Mal statt. Die erste Bewertung von Forschungseinrichtungen dieser Art in Großbritannien war 1986, weitere folgten 1989, 1992, 1996 und 2001. Mit jedem Durchgang wurde das Ver-

fahren verändert. Sichtbar zugenommen hat im Laufe der Jahre die Transparenz der Verfahrensschritte und Vorgaben, die heute umfassend online dokumentiert sind (www.rae.ac.uk).

Getragen und organisiert wird die RAE durch den Higher Education Funding Council for England (HEFCE)¹³ und die Aufgabe ist, eine Verteilungsgrundlage für den Anteil institutioneller Forschungsförderung zu schaffen. Eine derartig systematische Evaluierung über alle Forschungseinheiten des Landes entspricht in sofern der Tradition der Forschungsförderung Großbritanniens, als dort bereits früher als in anderen europäischen Ländern die Relevanzorientierung von Forschung öffentlich debattiert und politisch gefordert wurde. Mit der Ära Thatcher wurde das Forschungsbudget insgesamt verknappt und konsequent an Ergebnisse der Leistungsmessung gebunden (Braun 1997; Leszczensky et al. 2004b). Der seitdem zu beobachtende Trend, die öffentliche Mittelvergabe in der Wissenschaft mit Evaluationsergebnissen zu verknüpfen, scheint heute im Land wenig umstritten zu sein. Immer wieder hinterfragt und aktuell neu diskutiert wird allerdings auch in Großbritannien das *Wie* der Evaluierung (Gilbert/Lipsett 2007; McCormick et al. 2007).

Die RAE sieht vor, dass jede Universität disziplinäre Forschungseinheiten definiert und diese jeweils eine Bewerbungsunterlage entwickeln. Sie enthält neben einer festgelegten Anzahl an Daten zur Institution und einer Beschreibung der Forschungsstrategie und Organisationsstruktur die Benennung von so genannten „forschungsaktiven WissenschaftlerInnen“, für die als Grundlage der Bewertung jeweils vier Publikationen zu benennen sind. Abgabefrist für die Bewerbungsunterlagen der RAE 2008 war Ende 2007; 2008 tagten dann insgesamt 67 disziplinär organisierte Bewertungsgremien, sogenannte Units of Assessment (UoA), sowie 15 übergeordnete Gremien (Main panels), welche die Konsistenz der Bewertung innerhalb der Fachcluster überprüfen sollten.

An einem Bewertungsgremium sind etwa 10-15 Personen beteiligt, ganz überwiegend WissenschaftlerInnen aus Großbritannien, die in dem fachlichen Bereich der

¹³ HEFCE wurde 1992 als Nachfolgeorganisation des University Grants Council gegründet und ist für die Verteilung institutioneller Forschungsförderung und damit für eine der zwei Säulen im Fördersystem zuständig. Der andere Teil wird über die Research Councils projektbezogen vergeben (HEFCE 2007a). In Bezug auf die RAE übernimmt der HEFCE auch Managementaufgaben für die Förderorganisationen von Schottland, Wales und Nordirland.

UoA selbst wissenschaftlich aktiv sind. Die Auswahl der Persönlichkeiten zur Besetzung dieser Gremien beginnt circa drei Jahre vor dem eigentlichen Bewertungsverfahren im Rahmen eines öffentlichen Vorschlagsverfahren: Fachorganisationen und Einrichtungen, die zum Kreis der Nutzer von Wissenschaft zu zählen sind, haben das Recht, Personen vorzuschlagen, die dann in die engere Wahl kommen.¹⁴ Rund 5.000 Vorschläge sind für das aktuelle Verfahren eingegangen, von denen circa 1.200 durch das HEFCE ausgewählt wurden.

Ein Evaluationsverfahren im Rahmen der RAE wird also letztlich von solchen WissenschaftlerInnen bestritten, die von der Evaluationsagentur, dem HEFCE, für geeignet gehalten werden. Als gewisse Mindestvoraussetzung können wir annehmen, dass diese WissenschaftlerInnen über das je eigene Institut hinaus bekannt sein müssen. Als empirische Frage bleibt jedoch, nach welchen Kriterien die jeweiligen Einrichtungen die potenziellen Mitglieder vorschlagen und welchen Einfluss die spezifische Zusammensetzung einer Bewertungsgruppe auf die spezifische Praxis eines Evaluationsverfahrens hat. Hypothetisch können wir beispielsweise annehmen, dass bei einer Rekrutierung von FachgutachterInnen auf Basis der von den Nutzerorganisationen eingereichten Vorschlägen die Anwendungsorientierung der Forschung als Qualitätskriterium in der Bewertung ein stärkeres Gewicht erhält.

Zusätzlich zu den WissenschaftlerInnen ist an jedem Gremium auch eine MitarbeiterIn des HEFCE beteiligt. Er oder sie nimmt an jeder Sitzung teil, ist zuständig für die Rahmenorganisation und für das Protokoll.

Die Bausteine der Bewerbungsunterlagen sind in den Fachgremien jeweils unterschiedlich gewichtet, den größten Ausschlag für das Resultat gibt aber immer die Bewertung der eingereichten Publikationen. Die Gremien tagen in der Regel fünf Mal im Laufe des Jahres. Zum Ende des Prozesses sind sie gefordert, für jede zu bewertende Einheit eine Note zwischen 1* (national interessanter Beitrag) und 4* (weltweit bahnbrechender Beitrag) zu vergeben. Diese Notenskala wird in eine

¹⁴ Eine Liste der nominierenden Institute ist veröffentlicht unter <http://www.rae.ac.uk/pubs/2004/03/bodies.asp>.

Formel zur Berechnung des Budgets für die nächsten sieben Jahre¹⁵ übertragen und die Resultate aller Forschungseinheiten werden online publiziert.

Spezifische Formen der Zurichtung

Welche Formen der „Laboratorisierung“ können wir in Folge der Verfahrensregeln der RAE rekonstruieren? Betrachten wir wiederum zunächst jene „Zurichtungen“, die den komplexen und in seiner Ganzheit nicht beobachtbaren Gegenstand „wissenschaftliche Arbeit“ handhabbar machen sollen. Bedeutsam auf diesem Weg ist auch hier zunächst die Definition der Bewertungseinheit. Für die RAE haben wir es dabei mit einer komplexen und mehrstufigen Selektion zu tun. Seitens des HEFCE werden insgesamt 67 so genannte Units of Assessment (UoA) definiert. Die UoAs sind nach Fachgebieten unterteilt und umfassen entweder eine ganze Disziplin (zum Beispiel Geschichte) oder einzelne Teilgebiete (zum Beispiel gibt es in der Ökonomie getrennte UoAs zu „Business and Management Studies“ und „Accounting and Finance“). Im zweiten Selektionsschritt entscheiden die Institute, welche Arbeitsgebiete in welche UoAs „passen“. Eine Bewerbung für die RAE muss sich im Unterschied zu den WGL-Verfahren in Deutschland und dem SEP in den Niederlanden nicht an den Grenzen eines Instituts oder einer Fakultät orientieren, sondern die Verantwortlichen einer Einrichtung entscheiden selbst und für jede Bewerbung neu, welche Arbeiten sie einreichen und welcher UoA sie diese zuordnen.

Für die weitere Zurichtung der wissenschaftlichen Arbeit eines Arbeitsbereiches für die Evaluation wird im RAE zudem ein deutlicher Schwerpunkt definiert: Der wesentliche Gegenstand, anhand dessen wissenschaftliche Qualität zu bewerten ist, sind Publikationen, allerdings wie beschrieben nicht die Publikationen von allen MitarbeiterInnen sondern von einer exklusiven Auswahl „forschungsintensiver Mitarbeiter“, was einen hohen Interpretationsspielraum darüber lässt, wer das ist. Stellen wir uns also einen Vergleich zwischen der wissenschaftlichen Arbeit „in der Natur“ und dem für das Evaluationsverfahren zugerichteten Gegenstand

¹⁵ Aktuell wird über die Zukunft der RAE debattiert und vorgeschlagen ist ein neues Verfahren, das für naturwissenschaftliche Fächer eine Bewertung allein auf der Grundlage von metrischen Daten vorsieht, während für die Geisteswissenschaften eine „Light touch-Peer Review“ erhalten bleiben soll (HEFCE 2007b). Ein Teil der Pläne sieht auch vor, dieses neue Verfahren bereits während der nächsten Finanzierungsperiode einzusetzen, sodass sich die Laufzeit der Budgetierung, die auf der Grundlage der RAE 2008 festgelegt wurde, verkürzen würde.

vor, dann können wir eine Vielzahl wissenschaftlicher Arbeiten annehmen, die im RAE nicht zu beobachten sind: Arbeiten, die mündlich geblieben sind; Arbeiten, die aufgeschrieben aber (noch) nicht veröffentlicht wurden; Arbeiten, die veröffentlicht wurden, aber von solchen AutorInnen verfasst sind, die dem RAE nicht gemeldet wurden oder die nicht zu den vier ausgewählten Publikationen der gemeldeten WissenschaftlerInnen zählen. Die explizierten Verfahrensregeln machen keine eindeutigen Vorgaben dazu, wie dieser Auswahlprozess vorzunehmen ist. Es wird lediglich an die Institute appelliert, keine Diskriminierungen vorzunehmen: „we are encouraging HEIs to submit the work of all their excellent researchers, including those whose volume of research output has been limited for reasons covered by equal opportunities guidelines.” (HEFCE 2005: Abs. 36). Wir finden allerdings in den Unterlagen keinerlei Definition für „excellent researchers“, womit die Interpretation dieser Aufforderung den Instituten überlassen bleibt. Die Resultate der RAE im Jahre 2001 geben Aufschluss darüber, wie unterschiedlich mit der Selektionsanforderung umgegangen wird – manche Institute beziehen nahezu alle WissenschaftlerInnen in die Bewerbung ein, andere weniger als 20 Prozent. Von InterviewpartnerInnen hören wir oft Beschreibungen, die sich als „Taktik unter Unsicherheit“ zusammenfassen lassen, so etwa:

"As you'll know, no doubt, to an extent in this process we're trying to second-guess what the RAE team may eventually do in terms of the way the funding will flow from the result. So, throughout the process we're trying to wave up on the one hand, producing returns which were quite inclusive of staff, because one of the drivers of the funding we know will be volume, the number of people who are included, and on the other hand, trying to ensure that each of the 15 submissions were as tight and as strong as possible. Because one of the other drivers, obviously, is high quality. So we're trying to sort of trade off, trying to achieve that balance between volume and quality. So the rules, in those terms, weren't explicit in the way in which the RAE team presented them. That was clearly very deliberate. The university needed to decide on its stance." (GL Institut B-GB)

Was dann im Rahmen einer UoA tatsächlich begutachtet wird, obliegt der Entscheidung des GutachterInnengremiums. Stellt man in Rechnung, dass eine UoA durchschnittlich 35 Bewertungseinheiten umfasst, die wiederum durchschnittlich 20 forschungsaktive WissenschaftlerInnen anmelden und von denen jeweils vier Publikationen einreichen, dann ist ein Gremium für die Analyse von rund 2800 Publikationen zuständig.¹⁶ Je nach Disziplin handelt es sich dabei mehrheitlich um

¹⁶ Diese Zahlen wurden als arithmetische Durchschnittswerte anhand der veröffentlichten Resultate der 2001 RAE ermittelt. Tatsächlich variieren sowohl die Anzahl der angemeldeten for-

Aufsätze, aber auch um Monografien. Die einzelnen Gremien reagieren auf diese Quantität, indem sie bereits vor der Bewertungsphase in einer öffentlichen Darstellung zur gremienspezifischen Arbeitsweise angeben, wie viel der eingereichten Publikationen sie zu lesen versprechen. In der Praxis der vergangenen Verfahren, so wurde uns berichtet, reichte die Bandbreite der tatsächlich begutachteten Publikationen von 20 bis 100 Prozent. In den Einzelverfahren besteht bei der Umsetzung der Verfahrensregeln der RAE also ein weiter Gestaltungsspielraum.

(Ausgewählte) Publikationen machen zwar den Schwerpunkt dessen aus, was in einem Evaluationsverfahren der RAE begutachtet wird, sie bestimmen den Gegenstand aber nicht allein. Eine Bewerbungsunterlage umfasst überdies zahlreiche Kennziffern der Bewertungseinheit, darunter Zahlen der MitarbeiterInnen, NachwuchswissenschaftlerInnen und technische Infrastruktur. Das Institut ist außerdem gefordert, einen freien Text über die Strategie und den organisatorischen Aufbau der Einheit zu verfassen. Diese Darstellung wurde auch in früheren Verfahren bereits gefordert, dessen Bewertung soll aber 2008 erstmals nachvollziehbar Teil der Berechnungsformel werden. Die einzelnen Bewertungsgremien geben in ihrer Methodenbeschreibung an, mit welchem Gewicht sie diesen Teil in der Bewertung versehen.

Die Ganzheit wissenschaftlicher Arbeit wird also im RAE auf spezifische Weise für das Evaluationsverfahren zugerichtet und auf ausgewählte Forschungseinheiten, Publikationen, statistische Messgrößen und eine Prosa über Strategie und Umfeld der wissenschaftlichen Einrichtung komprimiert. Für jeden dieser Zurichtungsschritte finden wir in den schriftlichen Unterlagen zum Verfahren Anhaltspunkte, nach welchen Kriterien die jeweiligen Selektionen zu erfolgen haben. Explizit wird für die Bewertung der Publikationen eine Rangordnung von national bedeutend bis international bahnbrechend beschrieben. In der praktischen Umsetzung der Selektionsleistung können wir davon ausgehen, dass individuell unterschiedliche Deutungen dieser Kriterien zum Tragen kommen. Eine empirische Frage wäre etwa, ob die Selektionen von MitarbeiterInnen und Publikationen für die unterschiedlichen UoAs nach nachvollziehbar unterschiedlichen Abwägungen vorgenommen werden, ob beispielsweise die Gewichtung von Kriterien schon bei

schungsaktiven WissenschaftlerInnen als auch die Bewertungseinheiten pro Gremium erheblich.

den Auswahlentscheidungen in den Instituten zwischen den Natur- und den Geisteswissenschaften unterschiedlich vorgenommen wird.

Mündlichkeit, Schriftlichkeit und Körperlichkeit

Als konkretes Evaluationsverfahren innerhalb der RAE können wir uns am besten die Bewertungsleistung einer UoA vorstellen. Hier findet der eigentliche Urteilsbildungsprozess statt, hier tauschen sich die Peers aus und sind gefordert, am Ende des Prozesses ein gemeinsames Ergebnis zu präsentieren. Die mündliche Interaktion ist ein wesentliches Werkzeug, um dies zu erreichen. Somit basiert auch das RAE in einem hohen Maße auf Mündlichkeit, allerdings bleibt sie hier auf die Kommunikation zwischen den GutachterInnen beschränkt.

Der Diskurs zwischen den Gutachtern und Gutachterinnen kann sowohl innerhalb eines formalisierten Rahmens stattfinden als auch in informellen Gesprächen am Rande oder außerhalb der offiziellen Gremiensitzungen. Formal folgt jedes Bewertungsgremium in etwa demselben zeitlichen Ablauf: Erste Sitzungen finden bereits rund zwei Jahre vor der ersten Bewertungsaufgabe statt. Auf der Tagesordnung dieser Treffen stehen die eigenen Bewertungsmethoden, so auch beispielsweise die Gewichtung der Bewerbungsbestandteile für die Gesamtbewertung. Für das eigentliche Bewertungsverfahren waren im RAE 2008 jeweils etwa fünf Treffen für jedes Bewertungsgremium vorgesehen. Während die ersten beiden Treffen vor allem für Auswahl und Verteilung der zu lesenden Artikel genutzt wurden, dienten die anschließenden Sitzungen der Bewertung der Artikel sowie der weiteren von den Institutionen eingereichten Dokumente. Anders als beispielsweise bei der Begutachtung von WGL-Instituten lernen sich die GutachterInnen beim RAE auch sehr viel besser kennen, da sie über einen viel längeren Zeitraum miteinander zu tun haben. Damit bieten sich Ihnen auch sehr viel mehr Gelegenheiten für informelle mündliche Kontakte – etwa Vorgespräche zur nächsten Sitzung am Telefon oder ein informeller Austausch im Rahmen von anderen Gelegenheiten, bei denen sich die GutachterInnen zwischen den Sitzungen im Laufe des Jahres treffen.

Dagegen ist eine direkte Interaktion zwischen GutachterInnen und Begutachteten im RAE in der Regel nicht vorgesehen. Die Möglichkeit zum unmittelbaren Nachfragen und damit zum Entwickeln eines tieferen Verständnisses über den Kontext der zu bewertenden Arbeit steht somit als offizielles Instrument im RAE nicht zur Verfügung. Gleichwohl ist auch hier eine informelle Mündlichkeit möglich. Weil

die Gremienmitglieder beinahe ausschließlich im Land selbst rekrutiert werden, sind viele gleichzeitig für eine Bewerbungsunterlage ihrer Einrichtung verantwortlich. Zur Vermeidung von Interessenskonflikten sehen die Verfahrensregeln lediglich vor, dass Gremiumsmitglieder den Raum zu verlassen haben, wenn über die eigene Einrichtung verhandelt wird. Auch wenn diese Praxis nun über mehrere RAE-Runden eingeübt werden konnte und formal eingehalten wird, bleibt die informelle mündliche Kommunikation zwischen Gremiumsmitgliedern und Vertretern der zu bewertenden Einrichtung als empirische Frage von Bedeutung.

Neben der Mündlichkeit wird der Schriftlichkeit nach den Regeln der RAE eine große Bedeutung beigemessen, wobei ein hohes Maß an Standardisierung und Formalisierung vorgesehen ist. Zum einen hat Schriftlichkeit angesichts der im Zentrum stehenden Publikationen einen hohen Rang im RAE. Was in den anderen Verfahren auch vor Ort durch den Körper in Erfahrung gebracht werden kann, muss im RAE einzig über das Lesen von Schriftstücken erfolgen. Schriftlichkeit ist insofern im RAE ein zentrales Instrument zur Erzeugung von Signalen über die Qualität einer wissenschaftlichen Leistung.

Daneben hat Schriftlichkeit wie in den anderen Verfahren auch eine Dokumentations- und Kontrollfunktion. Zuständig dafür ist das HEFCE. Bereits in der frühen Vorbereitungsphase werden Kriterien, Prozesse und Selbstverständnisse formuliert und umfassend online dokumentiert – wir können hier als Adressaten also immer sowohl die interne als auch die breite externe Öffentlichkeit annehmen. Jede Sitzung der Bewertungsgremien wird überdies protokolliert. Diese Protokolle sind und bleiben vertraulich und sind nur den Mitgliedern der Bewertungsgremien zugänglich. Im Vergleich zu anderen Verfahren ist die schriftliche Dokumentation des Bewertungsverfahrens im RAE in einem hohen Maße standardisiert: Jeder diskutierte Beitrag wird abschließend mit einer Note von eins bis vier versehen – alle Bewertungen, alle durch informelle und formelle Mündlichkeit und Schriftlichkeit gewonnenen Eindrücke werden also in eine dieser Noten verdichtet. Auch an das Institut wird offiziell nur diese Zahl kommuniziert; anders als beispielsweise in der WGL wird kein qualifizierender Bewertungsbericht verfasst.

Körperlichkeit im Sinne von sinnlichem Fühlen und Ermessen des Gegenstandes scheint demgegenüber innerhalb der Verfahrensregeln der RAE eine geringere Rolle zu spielen als etwa bei Verfahren mit Vor-Ort-Begehung. Dennoch lassen

sich auch in der konkreten Situation von Gremiensitzungen im Rahmen des RAE eine Vielzahl sinnlicher Eindrücke vorstellen, die den Gang von Diskussionen im Panel und damit die Urteilsbildung beeinflussen können. Wenn VertreterInnen einzelner Bewerbungen beispielsweise gleichzeitig Mitglieder des Bewertungsgremiums sind, können wir durch deren persönliche Anwesenheit ähnliche Einflüsse annehmen wie bei der persönlichen Begegnung mit MitarbeiterInnen im Rahmen einer Vor-Ort-Begehung. Auch die Gestaltung des Tagungsraumes oder die Verpflegung des Gremiums können vermittelt über den Körper einen Einfluss auf die konkreten Bewertungsprozesse haben. Indem aber die Bewertung außerhalb der zu bewertenden Institute stattfindet, haben so symbolträchtige Dinge wie Gebäude, Räumlichkeiten oder andere sinnlich wahrnehmbare Gegebenheiten der Institute im RAE zumindest keinen direkten Einfluss auf die Bewertung. Inwieweit solche Dinge von einzelnen GutachterInnen bereits zu früheren Zeiten wahrgenommen wurden und sich in deren Gedächtnis eingeschrieben haben, so dass sie gegebenenfalls in die Bewertung einfließen, bleibt der empirischen Prüfung vorbehalten.

4 Fazit

Mit diesem Beitrag wurden die Verfahrensregeln von drei Evaluationsbeispielen einer dezidierten Analyse unterzogen: das Beispiel der Wilhelm-Leibniz-Gemeinschaft in Deutschland, des “Standard Evaluation Protocol for Public Research“ in den Niederlanden und des „Research Assessment Exercises“ (RAE) in Großbritannien. Verfahrensregeln wurden dabei zunächst eingeführt als Erwartungen, die aus dem Evaluationssystem an die einzelnen Verfahren gerichtet werden und die den Raum eingrenzen, innerhalb dessen ein Verfahren sich autonom entwickeln kann. Im Zentrum stand die Frage, welche strukturierende Wirkung auf den Bewertungsprozess durch spezifische Verfahrensregeln zu erwarten sind.

Konkretisiert wurden Verfahrensregeln unter Zuhilfenahme des Laborkonstruktivismus. Ähnlich wie Knorr-Cetina das wissenschaftliche Labor als domestizierten und damit artifiziellen „Handlungskontexte begreift, das in eingegrenzter Weise auf bestimmte Erzeugungsvorgänge spezialisiert ist und seinen Untersuchungsgegenstand zu diesem Zweck in spezifischer Weise und unter Zuhilfenahme spezifischer Techniken „zurichtet“, haben wir Evaluationen als Erzeugungsprozesse von Werturteilen definiert, die auf vorhergehenden Selektionen aufbauen. Verfahrens-

regeln von Evaluationen haben wir entsprechend als Techniken oder Werkzeuge betrachtet, die dazu dienen, den Bewertungsprozess zu vereinfachen und somit eine Art Laboratorisierungseffekt auslösen. Es wurde gefragt, welche Zurichtungen von spezifischen Verfahrensregeln im Hinblick auf den zu evaluierenden Gegenstand zu erwarten sind und welcher Stellenwert dabei Mündlichkeit, Schriftlichkeit und Körperlichkeit als Instrumenten der Urteilsbildung zukommt sowie welche weiteren Selektionsleistungen damit in den jeweiligen Verfahrensregeln eingelagert sind.

Sowohl hinsichtlich der Zurichtung als auch im Hinblick auf die Rolle von Mündlichkeit, Schriftlichkeit und Körperlichkeit lassen sich die untersuchten drei Verfahren zu zwei Grundtypen verdichten: Verfahren, bei denen die Bewertung auf der Basis eines Besuches der zu evaluierenden Einrichtung erfolgt – wie in den Verfahren der WGL in Deutschland und nach dem SEP in den Niederlanden – und Verfahren, in denen die Bewertung ausschließlich auf der Grundlage von schriftlichen Dokumenten (Publikationen) außerhalb der Institution erfolgt. Sowohl die Selektionsentscheidungen zur Definition des Gegenstandes als auch die Rahmenbedingungen für die Kommunikation innerhalb eines Bewertungsprozess sind hier weitgehend verschieden.

Das auf dem Vor-Ort-Besuch aufbauende Verfahren erfordert, den komplexen Bewertungsgegenstand so herzurichten, dass er im Rahmen einer für eineinhalb bis zwei Tage vorgesehenen „Begehung“ sinnlich erfasst werden kann. Das heißt, es müssen Selektionsentscheidungen der Art getroffen werden, wie man dem Begutachtungsgremium in der vorgesehenen Zeit einen guten Gesamteindruck von der Einrichtung und ihrer Arbeit vermittelt, etwa dass Leistungen, von denen man annimmt, dass sie als wichtig angesehen werden, nicht zu kurz kommen und weniger wichtige nicht zu viel Raum einnehmen, dass aber dennoch die gesamte Breite der Arbeit hinreichend erkannt und gewürdigt werden kann. Des weiteren haben sich die vorab vorzunehmende Entscheidungen darauf zu beziehen, dass in diesen Verfahren weniger die Inhalte der in der Einrichtung entstandenen zentralen Produkte wissenschaftlichen Arbeitens – Publikationen – Gegenstand der Betrachtung sind als allgemeine Leistungsindikatoren und Organisationsmerkmale wie der Gesamtumfang an Veröffentlichungen, klassifiziert nach den Orten ihres Erscheinens (z.B. Ranking der Zeitschriften, Ansehen der Verlage bei Monographien und Sammelbänden etc.) und Umfang der akquirierten Forschungsgelder sowie Forschungsprogramme und Organisation und Management der Forschung.

Während sich jedoch die Zurichtungseffekte bei der WGL-Evaluation auf die alle sieben Jahre stattfindende Begehung beschränken, ist davon auszugehen, dass die Verfahrensregeln des SEP in den Niederlanden aufgrund der ebenfalls vorgesehenen Selbstevaluation nach drei Jahren zusammen mit dem jährlichen Selbstmonitoring quasi einen dauerhaften Zurichtungseffekt haben, so dass die externe Evaluation mehr die Funktion einer „Kontrolle der Kontrolle“ (Power 1997) zukommt.

Im Unterschied dazu sind die RAE in erster Linie auf die Bewertung von Forschungsergebnissen fokussiert – Gegenstand sind Publikationen – wiewohl die institutionelle Umgebung im Verfahren von 2008 erstmals in die Formel zur Budgetberechnung einbezogen werden sollte. Allerdings werden nicht alle Veröffentlichungen bewertet, sondern die Einrichtungen sind gefordert, für ihre so genannten forschungsintensiven MitarbeiterInnen jeweils die vier besten auszuwählen. Die Zurichtung erfolgt hier also über eine doppelte Selektion: die Auswahl der besten Mitarbeiter und die Auswahl deren bester Publikationen. Sie führt damit einerseits zu einer extremen Begrenzung dessen, was bei der Bewertung überhaupt in Augenschein genommen werden kann, sieht man einmal davon ab, dass die Bewertungsgremien die Möglichkeit haben, ein weiteres Mal zu selektieren. Andererseits erlauben die Verfahrensregeln den zu bewertenden Institutionen bei der Definition dessen, was auszuwählen ist, einen hohen Interpretationsspielraum. Wer z.B. ‚forschungsintensiv‘ ist, entscheiden sie.

Auch hinsichtlich des Einsatzes der Urteilsbildungsinstrumente Mündlichkeit, Körperlichkeit und Schriftlichkeit ist vor allem zwischen Verfahren *mit* und *ohne* Vor-Ort-Besuch zu unterscheiden. Dies gilt vor allem für den Stellenwert von Körperlichkeit auf der einen und Schriftlichkeit auf der anderen Seite. Während Körperlichkeit in Verfahren, für die ein Vor-Ort-Besuch vorgesehen ist, einen hohen Stellenwert beim Gewinnen von Eindrücken zur Bewertung der Einrichtung hat, gilt das für das stärker formalisierte RAE-Verfahren sehr viel weniger. Hier spielt stattdessen Schriftlichkeit eine zentrale Rolle. Was in den anderen Verfahren vor Ort über Kommunikationen und andere sinnliche Eindrücke in Erfahrung gebracht wird, geschieht hier über das Lesen von Publikationen.

Was Knorr-Cetina im Hinblick auf die wissenschaftliche Erkenntnisproduktion zeigen konnte, lässt sich also auch für die Evaluation von wissenschaftlichen Leistungen feststellen: Evaluationen sind stets Konstruktionsleistungen, bei denen der

zu bewertende Gegenstand aus seiner „Natur“ herausgelöst und für das technische Instrumentarium des Evaluierens zugerichtet wird. Dies geschieht je nach Verfahrensregeln auf unterschiedliche Weise, wodurch auch der zu bewertende Gegenstand in unterschiedlichem Licht erscheint. Für die Ergebnisse von Evaluationen folgt daraus, dass sie stets in Relation zu ihrem Erzeugungsprozess zu bewerten sind. Und dieser, so konnte gezeigt werden, wird zu einem erheblichen Maße durch die Verfahrensregeln determiniert.

Literatur

- Barz, Andreas (1998): Evaluation im deutschen Hochschulsystem: Ziele, Instrumente, Erfahrungen, Trends. Impulsreferat in der Arbeitsgruppe Bildung / Soziale Dienste; 1. Jahrestagung der Deutschen Gesellschaft für Evaluation, Köln, 24. September. im Internet: <http://www.degeval.de/koeln1998/barz.htm>.
- Braun, Dietmar (1997): Die politische Steuerung der Wissenschaft. Ein Beitrag zum kooperativen Staat. Frankfurt; New York: Campus.
- (2004): Wie nützlich darf Wissenschaft sein? Zur Systemintegration von Wissenschaft, Ökonomie und Politik. In: Stefan Lange und Uwe Schimank (Hg.): Governance und gesellschaftliche Integration. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 67-87.
- Gilbert, Natasha/Lipsett, Anthea (2007): The Science of Funding, in: The Guardian vom 30. Oktober 2007.
- Guston, David H. (2000): Between Politics and Science. Assuring the Integrity and Productivity of Research. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: Cambridge University Press.
- Harré, Rom (1991): Vorwort. In: Knorr-Cetina, Karin: Die Fabrikation von Erkenntnis. Zur Anthropologie der Naturwissenschaft. Frankfurt am Main: Suhrkamp.
- HEFCE, Higher Education Funding Council (2005): RAE 2008. Guidance on submissions. Im Internet: <http://www.rae.ac.uk/pubs/2005/03/rae0305.doc> (6/2008).
- (2007a): Funding Higher Education in England. How HEFCE allocates its funds. Im Internet: www.hefce.ac.uk, Stand Mai 2008.
- (2007b): Research Excellence Framework. Consultation on the assessment and funding of higher education research post 2008. Im Internet: http://www.hefce.ac.uk/pubs/hefce/2007/07_34/.
- Hirschauer, Stefan (2006): Wie geht Bewerten? Zu einer anderen Evaluationsforschung. In: Uwe Flick (Hg.): Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen. Hamburg: Rowohlt, S. 405-423.
- Hornbostel, Stefan (2008): Neue Evaluationsregime? Von der Inquisition zur Evaluation. In: Hildegard Matthies und Dagmar Simon (Hg.): Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. Leviathan Sonderheft 24/2007. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 59-82.
- KNAW/ VSNU/NWO (2003): Standard Evaluation Protocol 2003-2009 for Public Research Organisations. Amsterdam; Utrecht; Den Haag.
- Knorr-Cetina, Karin (1988): Das naturwissenschaftliche Labor als Ort der "Verdichtung" von Gesellschaft. In: Zeitschrift für Soziologie 17: 85-101.
- (1991): Die Fabrikation von Erkenntnis. Zur Anthropologie der Naturwissenschaft. Frankfurt am Main: Suhrkamp.
- (2005a): From Pipes to scopes. The flow architecture of financial markets. In: Andrew Barry (Hg.): The Technological Economy. London: Routledge, S. 122-141.

- (2005b): How are Global Markets Global? The Architecture of a Flow World. In: Karin Knorr-Cetina und Alex Preda (Hg.): *The Sociology of Financial Markets*. Oxford; New York: Oxford University Press, S. 38-61.
- Leszczensky, Michael/ Jaeger, Michael/Orr, Dominic (2004a): Evaluation der leistungsbezogenen Mittelvergabe auf der Ebene Land-Hochschulen in Berlin. Gutachten im Auftrag der Berliner Senatsverwaltung für Wissenschaft, Forschung und Kultur. Discussionpaper. HIS-Kurzinformation. A 4/2004. Hannover: Hochschul-Informationen-System GmbH.
- Leszczensky, Michael/ Orr, Dominic/ Schwarzenberger, Astrid/Weitz, Birgitta (2004b): Staatliche Hochschulsteuerung durch Budgetierung und Qualitätssicherung: Ausgewählte OECD-Länder im Vergleich. Hannover: Hochschul-Informationen-System.
- Luhmann, Niklas (2001 [1969]): *Legitimation durch Verfahren*. 6. Auflage. Frankfurt am Main: Suhrkamp.
- McCormick, Ian/ Shearer, Jen/ Weir, Tony/ Davidson, Denise/ Sykes, Rob/ Studman, Cliff/ Mischewski, Brenden/Boston, Jonathan (2007): *Managing Research Assessment: RAE2008 and beyond*. Issues in research Management and Administration. Occasional Paper No 4 July 2007. East Anglia: University of East Anglia.
- Neidhardt, Friedhelm (1988): *Selbststeuerung in der Forschungsförderung. das Gutachterwesen der DFG*. Opladen: Westdeutscher Verlag.
- Orr, Dominic/ Jaeger, Michael/Schwarzenberger, Astrid (2007): Performance-based funding as an instrument of competition in German higher education. In: *Journal of Higher Education Policy and Management* 1/2007: 3-23.
- Power, Michael (1994): From the Science of Accounts to the Financial Accountability of Science. In: *Science in Context* 7(3): 355-387.
- (1997): *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- (2008): Research Evaluation in the Audit Society. In: Hildegard Matthies und Dagmar Simon (Hg.): *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 15-24.
- Ritter, Gerhard A. (1992): *Großforschung und Staat in Deutschland. Ein historischer Überblick*. München: Beck.
- Strathern, Marilyn (2000): Afterword: accountability. and ethnography. In: Marilyn Strathern (Hg.): *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge, S.
- Weingart, Peter (1998): Wissenschaft und Forschung. In: Bernhard Schäfers und Wolfgang Zapf (Hg.): *Handwörterbuch zur Gesellschaft Deutschlands*. Opladen: Leske und Budrich, S. 720-731.