

Von den Grenzen der Evaluation

Dr. Sybille Reichert
Plenarsitzung des SWTR
Bern, 23. April 2013

1. Kontext und Treiber der Evaluationen
2. Formen der Forschungsevaluation
3. Vor- und Nachteile qualitativer und quantitativer Leistungsbewertungen
4. Lösung 1: Informed Peer Review?
5. Lösung 2: Ex ante Steuerung?
6. Lösung 3: verdichteter Dialog mit Staat und Öffentlichkeit - gegen das herrschende Wissenschaftsverständnis ansteuern

Neue Erwartungen an die Rolle der Universität

Vorher:

Kulturelle Institution, krit. Distanz

**Reproduktion der Bildungselite
Bildung eher als Ausbildung**

**Ausbildungsstätte für wenige
Professionen (Med., Ing. Jur.,
Theol., Lehrer und Staatsdiener)**

Grundlagenorientierte Wiss.

**keine Organisation sondern
“loosely coupled professional
bureaucracy“**

Präsident repräsentiert

**Senat bildet Konsens über
Disziplinen hinweg**

Jetzt:

Problemlösungsinst., osmotisch

**Bildung/ Ausbildung für 30-40%
der Bevölkerung**

**Akademisierung der Berufe –
Berufsorientierung der Akademie**

Wertschöpfungsbeitrag der Wiss.

**Kohärenter Akteur im Wettbewerb
um Ressourcen: starke interne
Koordination und Profilbildung,
Schwerpunktsetzung**

**Präsident steuert, bereitet Univ.
auf Wettbewerbe vor**

Neue Erwartungen an die Rolle des Staates

Vorher:

**Garant der Unabhängigkeit
und kritischen Distanz
(Freiheit F&L)**

**Erwartungen an Ausbildung
höherer Staatsdiener
(Juristen, Lehrer)**

**Grundfinanzierung entspricht
dem Vertrauen in Wissensch.
als öffentlichem Gut**

Jetzt:

**Rückzug aus der regulativen
Detailsteuerung, Staat nicht
geeignet als Garant d. globalen
Wettbewerbsfähigkeit**

**Gewährt größere institutionelle
Autonomie zur Anpassung an
größeren Aufgabenkatalog -
aber nur bei Einsatz neuer
Instrumente der Fernsteuerung
für Einzelziele
(Zielvereinbarungen, LOM,
Wettbewerbe, Akkreditierung)**

Neue Steuerungserwartungen u. -instrumente

- Neuer Steuerungsdiskurs I: New Public Management
 - Inst. Autonomie, Globalbudgets, staatl. Fernsteuerung durch ex post Evaluation, Forderung der Überprüfbarkeit
 - Förderung der Leistungsorientierung durch Output-basierte Mittelvergabe (LOM, Wettbewerbe)
 - Evaluation auf allen Ebenen, mit Konsequenzen für Mittelzuteilung
 - Bsp.: England seit 80er, Niederlande seit 90er
- Neuer Steuerungsdiskurs II: Network Governance
 - Austausch, gemeinsame Interessen zwischen Universitäten u. externen Akteuren (Cluster Policies)
 - Bsp.: PRES und Institut Carnot (F), Centres of Innovation (N), Exzellenzcluster, BMBF Spitzencluster (D)

Treiber der Evaluation

- Wachsende institutionelle Autonomie
 - ⇒ erhöhte Erwartung an Steuerung
 - ⇒ Verbreitung neuer Steuerungsinstrumente, leistungsorientierte Mittelvergabe, Wettbewerbe
 - ⇒ Leistungsbewertung
- Massifizierung ⇒ Suche nach leistungsstarken Inseln
- Erhöhung des Drittmittelanteils (Wettbewerb, Peer R.)
- Diversifizierung der Finanzierungsquellen und –instrumente mit entsprechendem Antrags-/Berichtsaufwand
- Mangelnde Abstimmung des Datenbedarfs
- Internationalisierung des Peer Review
- Internationalisierung und wachsende Mobilität der Studierenden und Forschenden

Arten der Forschungsbewertung: Wissenschaftssystemimmanente Beurteilungen

1. Antrags- und Berichtspflichten sowie Begutachtung im Rahmen von Drittmittelförderung von Forschungsproj.
2. Begutachtung für Veröffentlichungen in wissenschaftl. Zeitschriften
3. Externe und hochschulinterne Begutachtung bei Berufungen
4. Beantwortung von Umfragen im Rahmen von Hochschulstudien

Arten der Forschungsbewertung: Überprüfung durch externe Anspruchsgruppen

5. Rechenschaftslegung zuhanden der Träger oder Geldgeber
6. Begutachtung von Forschungseinheiten und Forschungsgruppen im Rahmen von Fachevaluationen
7. Sammlung von Forschungsoutputdaten für Mittelvergabe
8. Sammlung von Forschungsoutputdaten zuhanden von externen Akkreditierungs- oder Qualitätsagenturen
9. Umfragen zuhanden von Rankinginstitutionen.
10. Evaluation von Forschungsprogrammen
11. Evaluation forschungspolitischer Maßnahmen

Peer Review: Vor- u. Nachteile

Vorteile	Nachteile
Stärkt selbstregulierenden Kräfte des Wissenschaftssystems: Beurteilung auf Augenhöhe (expert community) – bestmögliche Methode der Fremdbeurteilung	Mangelnde Reproduzierbarkeit und Verlässlichkeit der Urteile
Effektivität der Mittelvergabe: Größerer Anteil der begrenzten Mittel für leistungsstarke Forschende und vielversprechende Projekte Gute Verlässlichkeit bei Einschätzung der Leistungsschwachen	Mainstreaming-Effekte und Verstärkung der disziplinären Ausrichtung: Benachteiligung neuer Fachgebiete, interdisziplinärer Projekte, Vermeidung risikoreicher Forschung Schlechte Verlässlichkeit bei Einschätzung bahnbrechender Forsch.
	Benachteiligung nicht etablierter Wissenschaftler/innen (Nachwuchs, Frauen, Minderheiten)
	Beeinflussung durch Interessenskonflikte

Peer Review folgt intersubjektiven Konventionen – Lamont (2009): “pragmatic conception of truth“

- Stellt nicht die Legitimität des Peer Review Verfahrens sondern nur die überhöhten Erwartungen an seine Belastbarkeit und universale Gerechtigkeit in Frage
 - Mehrwert für wissenschaftliche Qualität kaum untersucht (nur Produktivitätssteigerung, Sichtbarkeit, Mitteleinwerbung)
 - Betonung des Peer Review erhöht Einfluss der „gatekeeper/ arbiters of excellence“ der Wissenschaft
 - Nutzen von Peer-Review basierten Evaluationen umstritten
 - besser als formative E. (Einschätzung von Perspektiven im Dialog mit Wissenschaftlern) als als reine Leistungsbeurteilung
 - Grenznutzen nimmt mit zunehmender Häufigkeit ab
 - Gutachterverfügbarkeit u. –qualität nimmt ab
- ⇒ Gefahr der eskalierenden Evaluationsfrequenz

Nutzen der Evaluation (Fachevaluationen Niedersachsen)

Einschätzung zur Fachevaluation	Anteil der Zustimmung unter Wissenschaftlern
Gespräche mit Gutachtern effektiv	44%
Gutachten konnte nicht mit Gutachtern diskutiert werden	55%
Gespräche unter wissensch. Gesichtspunkten interessant	28%
Gutachter waren angesehene Vertreter des Fachs	72%
Gutachterlichen Empfehlungen waren hinreichend präzise und handlungsleitend formuliert	50% (90%)
Umsetzung der Empfehlung wichtig	56%
Keine ausreichende finanzielle oder strukturelle Unterstützung bei der Umsetzung	62%
Aufwand entspricht nicht dem Nutzen	48%
Ergebnisse = Empfehlungscharakter, nicht für Mittelzuweisung geeignet	64%

Indikatoren: Vor- u. Nachteile

Vorteile	Nachteile
Weniger Aufwand als Peer Review	z.T. erheblicher Aufwand innerhalb der Hochschulen, weil Daten nicht verfügbar; Beschränkung auf verfügbare Daten verzerrt Abbild der Leistungen (inkl. Arten des Forschungsausputs)
Vergleichbarkeit der Institutionen	Keine Vergleichbarkeit zwischen versch. Forschungskontexten (Art und Frequenz der Publikationen); unterschiedliche Aussagekraft der Indikatoren (Drittmittel, ISI Publik.)
Objektivität der Daten	Daten bedürfen der Kontextualisierung um verzerrende Effekte zu verhindern
Schneller Überblick über viele Einheiten	Bevorteilung drittmittel- und internationaler, publikationsintensiver Disziplinen (schleichender Abbau der Geisteswissenschaften)
Ersetzt ungerechte historisch gewachsene Mittelzuteilung	Überbetonung der Drittmittel führt zu einer Überstrapazierung der Gutachter und höhlt Grundfinanzierung aus (Infrastruktur, Forschungsumgebung, Nachwuchs)
	Unerwünschte Verhaltensanpassungen (Vermeidung ergebnisoffener Projekte, Fragmentierung der Publik.praxis)

These 1

Der Zeitaufwand der Wissenschaftler in der Schweiz, der für die Bewertung wissenschaftlicher Leistungen und Berichte verwendet wird, sollte empirisch ermittelt werden, um die Verhältnismäßigkeit dieses Aufwands (im Vergleich zum Zeitaufwand für die tatsächliche wissenschaftlichen Leistung in Forschung und Lehre) einzuschätzen.

These 2

Die Methodenkritik und begrenzte Verlässlichkeit der Leistungsbewertungen sollten vermehrt in Wissenschaft und Öffentlichkeit kommuniziert werden.

These 3

Die Verwendung von Leistungsindikatoren sollte möglichst auf Hintergrundinformation bei Evaluationen oder Berichten beschränkt werden, in denen diese Information auch kontextualisiert werden kann. Im Falle von Fachbereichen, in denen quantitative Bewertungen kaum Aussagekraft beanspruchen können, sollten sie gar nicht vorgesehen werden.

These 4

Die über Wettbewerb als Drittmittel oder als Teil des Grundauftrags der Universitäten zugeteilten Forschungsmittel müssen ein Gleichgewicht zwischen dynamischer Projektentwicklung und langfristiger Orientierung der Forschung erlauben.

Der Erfolg bei der Einwerbung von Drittmitteln darf nicht als Haupterfolgsmerkmal eines Forschenden oder einer Forschungseinheit gelten sondern muss im Zusammenhang mit anderen Erfolgsausweisen und vor dem Hintergrund der Disziplin und der institutionellen Rahmenbedingungen eingeordnet werden.

Informed Peer Review als Lösung?

- Quantitative Wissenschaftsindikatoren müssen „in die systematische Reflexion der Wissenschaft eingebettet sein“ (Weingart) (Kontextabhängigkeiten aufdecken)
 - Indikatoren als Hilfsmittel für Peer Review (Fragen, Bias-Korrektur)
- ⇒ Linderung möglicher Verzerrungen, aber keine Lösung der methodischen Probleme

Außerdem:

- Zunahme der Eval. ⇒ verstärkter Rekurs auf Indikatoren
- Eval.gegenstände immer häufiger jenseits der Expertise der Peers (Fächergruppen, Forschungsumgebungen als Eval.gegenstand)

Lösung:

Anpassung der Frequenz, Form und Zielsetzung der Evaluationen

- Evaluation als offener Dialog über mögliche Verbesserungen und Potentiale, als „Impuls zur Pfadkorrektur“
- Entroutinisierung, Orientierung an wirklichen Fragen, Anlass- und Strukturbezogen (Krull)
- Entkoppelung von Gratifikations- und Sanktionsmaßnahmen

These 5

Die Häufigkeit von Evaluationen an Schweizer Wissenschaftsinstitutionen sollte dringend gemindert werden, um ihren Nutzen zu erhalten und den hiermit verbundenen Aufwand zu rechtfertigen.

Evaluationen sollten nur dann durchgeführt werden, wenn sie einem genuinen Interesse der Institution und ihrer Wissenschaftler entspringen und durch deren eigene Entwicklungsfragen strukturiert sind. Die Evaluation der Leistungen einzelner Wissenschaftler sollte für diese einen Mehrwert hinsichtlich der Reflektion der eigenen wissenschaftlichen Ausrichtung bieten und so zu deren Motivation beitragen können.

These 6

Um einen offenen Dialog zu ermöglichen, der für wirkliche Qualitätsentwicklung entscheidend ist, dürfen Evaluationen nicht mit unmittelbaren Mittelverteilungskonsequenzen verbunden sein, da diese ihr gerade den Prüfungscharakter verleihen, der eine offene Selbstreflexion über eigene Schwächen und Zukunftsperspektiven unmöglich macht.

Leistungskontrollen und andere rein kontrollierende Eingriffe in das Wissenschaftssystem sollten soweit wie möglich vermieden werden, um die Grundlage des Wissenschaftssystems als einer vor allem auf intrinsischer Motivation aufbauenden kulturellen Praxis nicht zu zerstören. Leistungsbewertungen sollten nur dann zum Einsatz kommen, wenn sie die Selbstreflexionsprozesse der Wissenschaft und ihre inhärente Qualitätsorientierung unterstützen.

Leistungsanreize in der Wissenschaft

- Hohes Maß an intrinsischer Motivation
- Motivationshemmende vs. –verstärkende Wirkung von Leistungsanreizen: Produktivitätsgewinne sind keine Qualitätsgewinne
 - Kontrollierende Anreize unterminieren Motivation
 - Auf Gestaltungsfreiheit / Selbstverantwortung abzielende Anreize verstärken Motivation
- Entwicklung wissenschaftskompatibler Steuerungsformen
- Primär auf ex ante Entscheidungen konzentrieren: Auswahl und Ausbildung der Wissenschaftler
- Diversität der Ziele der Wissenschaft und der Leistungsstärke muss berücksichtigt werden

These 7

Universitäten und wissenschaftsfinanzierende Institutionen sollten Methoden und Ausmass der ex-post Evaluation und Leistungsbewertung als Qualitätssicherungsmaßnahme grundlegend überdenken und vermehrt auf die effektiveren Methoden der Qualitätsentwicklung über verbesserte ex ante-Auswahl ihrer Wissenschaftler und Wissenschaftsmanager setzen.

These 8

Leistungsförderung sollte vor allem über die Verstärkung der bereits gegebenen Motivation erfolgen, d.h. über Gewähren von Gestaltungsfreiheiten und Verstärkung von vielversprechenden Entwicklungen und Initiativen. Der wachsende Kontrollwahn von Staat und Gesellschaft gegenüber der Wissenschaft sollte ersetzt werden durch eine durch gegenseitiges Vertrauen und dichten Dialog gekennzeichnete Ermöglichungskultur.

Wie können wir das herrschende Verständnis der Wissenschaft verändern?

- Wissenschaft muss nützlich sein
- Wissenschaft ist überall, auch außerhalb der wiss. Institutionen
- Wissenschaft ist nicht nur im öffentlichen Interesse
- Wissenschaft muss durch externe Anspruchsgruppen steuerbar sein
- Wissenschaftliche Produktion muss reputiert und messbar sein, um wichtig zu sein.
- Wissenschaft muss sich vermarkten.
- Wissenschaft muss Ereignischarakter haben, um wahrgenommen zu werden.
- Wissenschaft darf nie scheitern sondern muss immer positive Ergebnisse vorweisen.

Staat und Wissenschaft brauchen:

- einen verdichteten Dialog über Inhalte, Ziele und Produkte der Wissenschaft, die unseren heutigen oder morgigen Alltag mitbestimmen
- die Betonung
 - der Eigenheit wissenschaftlicher Institutionen und ihrer optimalen Rahmenbedingungen
 - des Werts und der Innovationskraft zweckfreier Räume
 - des Werts der Umwege und des Scheiterns
 - der Effizienz des Vertrauens als Grundlage wiss. Schaffens

Die Thesen

These 1

Der Zeitaufwand der Wissenschaftler in der Schweiz, der für die Bewertung wissenschaftlicher Leistungen und Berichte verwendet wird, sollte empirisch ermittelt werden, um die Verhältnismäßigkeit dieses Aufwands (im Vergleich zum Zeitaufwand für die tatsächliche wissenschaftlichen Leistung in Forschung und Lehre) einzuschätzen.

These 2

Die Methodenkritik und begrenzte Verlässlichkeit der Leistungsbewertungen sollten vermehrt in Wissenschaft und Öffentlichkeit kommuniziert werden.

These 3

Die Verwendung von Leistungsindikatoren sollte möglichst auf Hintergrundinformation bei Evaluationen oder Berichten beschränkt werden, in denen diese Information auch kontextualisiert werden kann. Im Falle von Fachbereichen, in denen quantitative Bewertungen kaum Aussagekraft beanspruchen können, sollten sie gar nicht vorgesehen werden.

These 4

Die über Wettbewerb als Drittmittel oder als Teil des Grundauftrags der Universitäten zugeteilten Forschungsmittel müssen ein Gleichgewicht zwischen dynamischer Projektentwicklung und langfristiger Orientierung der Forschung erlauben.

Der Erfolg bei der Einwerbung von Drittmitteln darf nicht als Haupterfolgsmerkmal eines Forschenden oder einer Forschungseinheit gelten sondern muss im Zusammenhang mit anderen Erfolgsausweisen und vor dem Hintergrund der Disziplin und der institutionellen Rahmenbedingungen eingeordnet werden.

These 5

Die Häufigkeit von Evaluationen an Schweizer Wissenschaftsinstitutionen sollte dringend gemindert werden, um ihren Nutzen zu erhalten und den hiermit verbundenen Aufwand zu rechtfertigen.

Evaluationen sollten nur dann durchgeführt werden, wenn sie einem genuinen Interesse der Institution und ihrer Wissenschaftler entspringen und durch deren eigene Entwicklungsfragen strukturiert sind. Die Evaluation der Leistungen einzelner Wissenschaftler sollte für diese einen Mehrwert hinsichtlich der Reflektion der eigenen wissenschaftlichen Ausrichtung bieten und so zu deren Motivation beitragen können.

These 6

Um einen offenen Dialog zu ermöglichen, der für wirkliche Qualitätsentwicklung entscheidend ist, dürfen Evaluationen nicht mit unmittelbaren Mittelverteilungskonsequenzen verbunden sein, da diese ihr gerade den Prüfungscharakter verleihen, der eine offene Selbstreflexion über eigene Schwächen und Zukunftsperspektiven unmöglich macht.

Leistungskontrollen und andere rein kontrollierende Eingriffe in das Wissenschaftssystem sollten soweit wie möglich vermieden werden, um die Grundlage des Wissenschaftssystems als einer vor allem auf intrinsischer Motivation aufbauenden kulturellen Praxis nicht zu zerstören. Leistungsbewertungen sollten nur dann zum Einsatz kommen, wenn sie die Selbstreflexionsprozesse der Wissenschaft und ihre inhärente Qualitätsorientierung unterstützen.

These 7

Universitäten und wissenschaftsfinanzierende Institutionen sollten Methoden und Ausmass der ex-post Evaluation und Leistungsbewertung als Qualitätssicherungsmaßnahme grundlegend überdenken und vermehrt auf die effektiveren Methoden der Qualitätsentwicklung über verbesserte ex ante-Auswahl ihrer Wissenschaftler und Wissenschaftsmanager setzen.

These 8

Leistungsförderung sollte vor allem über die Verstärkung der bereits gegebenen Motivation erfolgen, d.h. über Gewähren von Gestaltungsfreiheiten und Verstärkung von vielversprechenden Entwicklungen und Initiativen. Der wachsende Kontrollwahn von Staat und Gesellschaft gegenüber der Wissenschaft sollte ersetzt werden durch eine durch gegenseitiges Vertrauen und dichten Dialog gekennzeichnete Ermöglichungskultur.