

Editorial – Die Evaluation der Soziologie und die Soziologie der Evaluation

Rainer Diaz-Bone (Universität Luzern) und Jörg Rössel (Universität Zürich)

Der Erfolg der Sozialwissenschaften im Allgemeinen und der Soziologie im Besonderen zeigt sich auch an der Allgegenwart ihrer Methoden und Verfahren. So findet sich der Begriff der «Evaluations-Gesellschaft» (Dahler-Larssen 2012), der zum Ausdruck bringt, dass und in welchem Ausmass die methodisch-empirische Bewertung sozialer Programme, Dienstleistungen, Institutionen in allen gesellschaftlichen Bereichen und in wissenschaftlichen Disziplinen heutzutage ein sowohl selbstverständlicher und doch auch zunehmend kritisch beurteilter Vorgang geworden ist. Seit Jahren mehren sich nun auch die wissenschaftspolitischen Diskurse, die die Evaluation von Disziplinen insgesamt einfordern oder kritisieren und problematisieren. Die Schweizerische Rektorenkonferenz (CRUS) hat mit ihrem Programm «*Mesurer les performances de la recherche en sciences humaines et sociales*» den Sozial- und Geisteswissenschaften in der Schweiz die Hand gereicht, um über den Einsatz von Evaluationsinstrumenten zu diskutieren und deren nähere Ausgestaltung differenzierter zu reflektieren (Zürcher in diesem Heft).¹ Unabhängig von diesem konkreten Anlass sind aber Evaluationen und darauf basierende Rankings von Forschern, Instituten und Universitäten schon länger Gegenstand der wissenschaftlichen Diskussion.

Die Soziologie soll und muss sich an dieser Diskussion in mindestens drei Hinsichten beteiligen:

Erstens werden die Soziologie, ihre Akteure und Forschungseinrichtungen ohnehin regelmässig Evaluationen unterzogen, ob die SoziologInnen

damit einverstanden sind oder nicht (vgl. für ein besonders elaboriertes Beispiel einer Evaluation Hinz in diesem Heft). Dazu zählen nicht nur die von Münch (in diesem Heft) diskutierten Evaluationen von Forschungshypothesen, sondern auch die Evaluationen von Aufsätzen, Gesuchen sowie von Bewerberinnen und Bewerbern für Positionen im soziologischen Wissenschaftsbetrieb. Eine Beschäftigung mit diesen Evaluationen erfordert sicher im ersten Schritt eine Bilanz der zahlreichen Probleme und Fallstricke, die mit Evaluationen und Rankings verbunden sind (vgl. den Beitrag von Budowski und Nollert in diesem Heft, zum Teil auch die Beiträge von Lessenich sowie von Münch in diesem Heft). Mit einer rein defensiven Haltung wird die Soziologie sich allerdings sowohl wissenschaftspolitisch als auch wissenschaftlich isolieren. Zudem wäre damit auch eine komplette Willkür in den Standards der ohnehin durchgeführten Evaluationen verbunden. Es gilt also konstruktiv über angemessene Formen der Evaluation nachzudenken: Wie von Zürcher (in diesem Heft) vorgeschlagen, geht es um die Bestimmung von zu bewertenden Leistungsdimensionen («Was-Frage» der Evaluation) und die Operationalisierung dieser Dimensionen durch angemessene Indikatoren («Wie-Frage» der Evaluation). In diese Richtung geht der Beitrag von Hinz in diesem Heft, der über die ausgesprochen differenzierte Evaluation der Soziologie in Deutschland berichtet, die vom Wissenschaftsrat initiiert wurde. Im Rahmen dieser Evaluation hat die Expertengruppe neben der Erhebung von Kennziffern für zahlreiche Evaluationsdimensionen sogar Beispielpublikationen der evaluierten Einheiten angefordert und gelesen, um eine qualitativ hochstehende Bewertung der Publikationsaktivitäten der betroffenen Forscher vorzunehmen. Es zeigt sich allerdings, dass dieser

1 Siehe auch die Beiträge im Dossier «Qualitäts- und Leistungsmessung in den Geisteswissenschaften» Bulletin 2/2011 der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW), S. 34–56.

enorme Aufwand nur zu einem ausserordentlich geringen Mehrwert führt. Am Ende kann die Gesamtbewertung der Forschungsleistung der evaluierten Einheiten allein auf der Grundlage der Publikationen im *Web of Science* sehr gut erfasst werden. Dies wird auch durch die Ergebnisse der explorativen Studie von Rössel und Pap (in diesem Heft) unterstützt. Diese können zeigen, dass die Erfassung von Publikationsaktivitäten, relativ unabhängig von der verwendeten Datenquelle, zu recht ähnlichen Ergebnissen führt. Aber diese Beiträge können natürlich nur explorative erste Anstösse zur Diskussion darstellen, die durch sorgfältige und systematische Analysen und Diskussionen von Evaluationsinstrumenten abgelöst werden müssen. Eine konstruktive Diskussion dieser Art wird es der Soziologie ermöglichen, offensiv mit eigenen Angeboten auf die gesellschaftliche und politische Aufforderung zur (Selbst-)Evaluation der Disziplin zu reagieren. Darüber hinaus kann eine solche Diskussion auch zu einem Reflexionsprozess über das Selbstverständnis des Faches beitragen. So wird in der Diskussion über Evaluationen häufig von der Soziologie als einem multiparadigmatischen Fach gesprochen (vgl. Budowski und Nollert sowie Zürcher in diesem Heft). Oder es wird herausgestellt, dass in den Sozialwissenschaften eine Pluralität von Qualitätskonventionen und Qualitätsverständnissen vorliegt und in der Praxis des Peer Reviewings eine Heterogenität von Kriterien zur Anwendung kommt (Lamont 2009). Dies ist ein Selbstverständnis der Disziplin, das von einer anders ansetzenden Position zu Evaluationen durchaus in Frage gestellt werden kann: Möglicherweise befindet sich die Soziologie überhaupt nicht in einer multiparadigmatischen Konstellation, sondern in einer vorparadigmatischen Phase, in der noch nicht einmal ein Konsens über grundlegende wissenschaftliche Standards existiert, die ein Urteil über die Fruchtbarkeit von Forschungsprogrammen zuliessen (siehe Münch in diesem Heft). Insofern sind Diskussionen über Evaluationen nicht nur eine von aussen herangetragene Pflicht,

sondern können auch die interne Diskussion in der Soziologie befruchten und das wissenschaftliche Selbstverständnis der Disziplin befördern oder zumindest zur Selbstaufklärung beitragen. Dann kann nicht nur für Teilbereiche des Faches (wie soziologische Theorie und methodologische Positionen) die Frage beantwortet werden, was die disziplinäre Struktur und auch die Verfassung des Faches in Sachen «Qualität» ist (wie auch immer diese zu bestimmen wäre). Dies dürfte gerade in einer Zeit, in der die Soziologie zunehmend durch andere Disziplinen herausgefordert wird, welche erstens bestimmte spezielle Soziologien für sich reklamieren (Politische Soziologie, Öffentlichkeitssoziologie, Bildungssoziologie usw.) oder zweitens sogar die Soziologie als Disziplin mehr oder weniger als obsolet ansehen (wie dies in den behavioral sciences geschieht), für die Soziologie höchst produktiv sein.

Zweitens sind Soziologinnen und Soziologen mitsamt ihrer Expertise und ihren Methodenkompetenzen vielfach in praktischen Evaluationen engagiert und exponieren somit auch die Rolle der Soziologie und ihres Instrumentariums in den gesellschaftlichen Diskursen über Sinn und Zweck von Evaluationen, über ihren möglichen Nutzen und ihren möglichen Schaden. «Evaluation» ist ein soziologisches Tätigkeits- und Berufsfeld. Die «Soziologie» ist also nicht nur bereits Objekt von Evaluationen, sie ist auch bereits eine eingesetzte «Ressource» – wenn auch nicht gesamthaft als Disziplin, aber doch repräsentiert durch ihre organisationstheoretische und methodische Expertise. Hierbei werden ihre Repräsentanten auch in die normativ-kritischen Problematisierungen verwickelt, sei es, dass ihre Evaluationstätigkeit in die Kritik gerät, sei es, dass sie die Evaluationspraxis anderer selbst kritisieren (Guba und Lincoln 1989).

Drittens kann die Soziologie aber Evaluationen selbst auch zum Gegenstand soziologischer Analyse machen (siehe den Beitrag von Diaz-Bone in diesem Heft, zum Teil auch die Beiträge von Münch und

Lessenich in diesem Heft). Dabei dürfte es primär um vier Fragen gehen:

- a) Es geht grundsätzlich zunächst um die ursächliche Erklärung des gerade angeführten Phänomens selbst: Warum führen Akteure und Organisationen bestimmte Formen von institutionalisierten Bewertungen durch? Allgemeine Verweise auf die Tatsache, dass Menschen schon als anthropologische Konstante Phänomene kognitiv bewerten müssen, diese Bewertungen auch gerne quantifizieren (siehe Lessenich in diesem Heft) oder auch Hinweise auf vermeintlich globale Prozesse wie die Durchsetzung des Neoliberalismus (Lessenich und Münch in diesem Heft) können sicher als Ausgangspunkt für diese Diskussionen und Forschungen dienen. Aber im nächsten Schritt müssen präzisere und differenziertere Erklärungen entwickelt werden (z. B. im Anschluss an den Beitrag von Diaz-Bone), die uns die spezifischen Formen der Institutionalisierung von Evaluationen oder auch die Strategien der Akteure erklären: Warum wurde z. B. in Deutschland eine ungeheuer aufwändige Evaluation verschiedener Fächer durch den Wissenschaftsrat durchgeführt, die dann letztlich aber komplett folgenlos blieb (Hinz in diesem Heft)? Genauso kann man sich auch die Frage stellen, warum gerade die Soziologie aus dem CHE Ranking in Deutschland im Sinne einer Exit-Option komplett ausgestiegen ist (Lessenich in diesem Heft), während die Politikwissenschaft im Sinne einer Voice-Option über eine Verbesserung des Rankings mit dem CHE verhandelt hat.² Um diese Fragen zu beantworten, benötigt man differenzierte theoretische Werkzeuge zur Erklärung des Umgangs mit Evaluationen und ihrer Institutionalisierung.
- b) Weiter kann die Soziologie die methodologischen Problematiken zum Untersuchungs-

2 Das CHE ist das Centrum für Hochschulentwicklung (www.che.de).

- gegenstand machen, die sich aus den normativen Entscheidungen ergeben, was denn als Qualitäten und Ziele von Evaluationen gelten und daher wie zu bewerten sein soll (siehe insbesondere die Beiträge von Bonvin, Lebaron und auch Diaz-Bone in diesem Heft). Hier stehen Fragen an, wie der Zusammenhang von Wertesystemen und (Qualitäts)Indikatoren in Evaluationen praktisch hergestellt wird und wie die Evaluationspolitiken auch methodologisch fundiert sind bzw. sich performativ im Bereich der Methoden artikulieren.
- c) Daran schliesst ein weiterer Aspekt unmittelbar an: Die Soziologie kann ihr vielfältiges Methodeninstrumentarium und ihre Fähigkeiten im Bereich der Methodenforschung einsetzen, um auf diesem Weg Evaluationsinstrumente selbst zu evaluieren, sie zu verbessern und methodisch-kritisch zu reflektieren. Dabei ist keineswegs nur an die sogenannte quantitative empirische Sozialforschung zu denken, die im Bereich der Evaluationsforschung sehr aktiv ist (Rossi et al. 1988; Bortz und Döring 2006), sondern auch an bestimmte Bereiche der sogenannten qualitativen Sozialforschung und ihrer Beiträge zur Evaluationsforschung (Patton 2015). Hier könnte man als Referenz beispielsweise an die systematische Anwendung qualitativer Methoden in der Evaluation von demokratischen Regimen denken, wie sie im Rahmen von sogenannten "democratic audits" stattfinden (Beetham und Weir 2000).
- d) Schliesslich kann und soll sich die Soziologie auch den Anreizwirkungen und gesellschaftlichen Folgen von Evaluationen widmen. Donald T. Campbell hatte bereits in den 1960er-Jahren die Evaluationsforschung in diesem Sinne als sozialwissenschaftliches Gestaltungsinstrument für Institutionen und institutionelle Kontexte des Handelns und Entscheidens aufgefasst (Campbell und Ross 1968; Campbell 1969). Und vielfach wird bereits angenommen, dass die Messung von

bestimmten Leistungen im Rahmen von Evaluationen selbst zu Verhaltensänderungen bei den evaluierten Akteuren führt. Es bliebe aber noch empirisch zu zeigen, dass dies generell der Fall ist und wie diese Verhaltensänderungen tatsächlich aussehen. Betrachtet man die Argumentation von Hinz (in diesem Heft) scheint die aufwändige Evaluation der Soziologie in Deutschland durch den Wissenschaftsrat keine erheblichen Auswirkungen auf die Ausstattung von Instituten oder die Verhaltensweisen von Soziologen gehabt zu haben. Daher bleibt zu erforschen, wovon solche Verhaltensänderungen abhängen und welche Richtung sie tatsächlich nehmen. Damit kann die Soziologie dann wiederum einen Beitrag zur Gestaltung von Evaluationen leisten, da unter Kenntnis von möglichen evaluationsbasierten Fehlanreizen Evaluationen besser gestaltet werden können.

Zusammenfassend betrachtet ist das Thema Evaluationen für die Soziologie eine spannende und – bedingt auch durch wissenschaftspolitische Einflüsse – immer drängendere Herausforderung. Auf der einen Seite kann sie mit ihren theoretischen und empirischen Werkzeugen an der Verbesserung von Evaluationen mitwirken und damit die Gestaltung der Evaluationen nicht einfach externen Akteuren überlassen. Auf der anderen Seite kann die Soziologie aber wiederum Evaluationen auch zu ihrem Gegenstand machen und damit die wissenschaftliche Diskussion befruchten. Insgesamt eröffnet sich ein Horizont für die Soziologie als Disziplin, der in der Ausweitung ihrer Autonomie besteht, wenn sie sich anhand selbst entwickelter sowie methodisch-reflektierter normativer Grundlagen die Evaluation als Instrument für ihre Selbstbe-

wertung zunutze machen kann, anstatt defensiv auf externe Einflussnahmen und Steuerungsversuche reagieren zu müssen.

Wir hoffen, dass dieses Themenheft eine konstruktive Diskussion über Evaluationen in der Schweizerischen Soziologie anstößt.

Literatur

- Beetham, David und Stuart Weir (2000). Democratic audit in comparative perspective. In: Lauth, Hans-Joachim/Pickel, Gert/Welzel, Christian (Hrsg.), *Demokratiemessung*. Wiesbaden: VS Verlag, S. 73–88.
- Bortz, Jürgen und Nicola Döring (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Aufl. Heidelberg: Springer.
- Campbell, Donald T. und Laurence Ross (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. *Law & Society Review* 3(1), S. 33–54.
- Campbell, Donald T. (1969). Reforms as experiment. *American Psychologist* 24(4), S. 409–429.
- Dahler-Larsen, Peter (2012). *The evaluation society*. Stanford: Stanford University Press.
- Guba, Egon und Yvonna Lincoln (1989). *Fourth generation evaluation*. Newbury Park: Sage.
- Lamont, Michèle (2009). *How professors think. Inside the curious world of academic judgment*. Cambridge: Harvard University Press.
- Patton, Michael (2015). *Qualitative research and evaluation methods*. 4. Aufl. Thousand Oaks: Sage.
- Rossi, Paul, H., Howard E. Freeman und Gerhard Hofmann (1988). *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.

Die Vermessung der Forschung: Genese, Stand und Handlungsbedarf

Markus Zürcher (Generalsekretär SAGW)

In den neunziger Jahren des letzten Jahrhunderts hat der damalige Schweizerische Wissenschaftsrat (SWR) erstmals die Forschungsleistungen der schweizerischen Universitäten in den verschiedenen Fachbereichen nach den bekannten Indikatoren (Bibliometrische Daten, Zahl der Forschungsprojekte und Patente) systematisch vermessen (SWR 1998, 1999a und 1999b). Befördert von der im Zuge des New Public Managements gewährten Autonomie etablierten Universitäten, Forschungsförderungsorganisationen sowie die zuständigen Behörden eine umfassende Rechenschaftslegung über die Leistungen in Forschung und Lehre. Heute ist die institutionalisierte Evaluation von Fakultäten, Departementen, Instituten und Ordinariaten ein mehrheitlich als bürokratisches Ärgernis wahrgenommener Bestandteil des universitären Lebens. Eher ausgeblendet wird, dass im Kerngeschäft der Forschung die Evaluation zu einer Haupttätigkeit geworden ist: Forschungsanträge abfassen, die in der Regel nach einem Leistungsausweis und einem Leistungsversprechen für die Zukunft verlangen, zu Hauf Anträge und Arbeiten von «Peers» beurteilen, gelegentlich bei institutionalisierten Evaluationen mitwirken, nehmen einen substanziellen Anteil der zeitlichen Ressourcen von Forschenden in Anspruch. Nicht selten ist die Klage, dass die Prüfung wie das Verfassen von Anträgen kaum mehr Zeit für Forschung und Lektüre lasse.

Es sei vorweggenommen: Jene Zeit, in welcher der Disput zwischen Gelehrten über Bedeutung und Wert einer Forschungsleistung entschied, wird nicht zurückkehren. Dazu ist das Forschungssystem zu gross und zu ausdifferenziert geworden. Entsprechend erfordert sowohl die Rechenschaftslegung wie die Steuerung die Erfassung und Messung eingesetzter Ressourcen sowie damit erbrachter Leistungen. Schliesslich verfügen die

Universitäten über die administrativ-technischen Voraussetzungen, das Wissen und in langjähriger Praxis erworbene Erfahrung und Kompetenzen. Bekanntlich wurde dieser Aufstieg der Forschungsmessung von einer grundsätzlich ablehnenden Haltung, teilweise heftiger Kritik, eher selten von gehaltvollen und konstruktiven Debatten begleitet. Gerade der breite und systematische Einsatz von Evaluationsverfahren erfordert dessen kritisch-konstruktive Prüfung, denn in zunehmendem Masse hängt die Entwicklung eines Forschungsbereichs von den eingesetzten Prüfverfahren und -kriterien ab. Dabei wird rasch deutlich, dass zumindest zwei grundlegende Fragen nicht hinreichend geklärt und ausdiskutiert sind. Möglicherweise müssen beide Fragen immer wieder von Neuem verhandelt werden. In jedem Fall ist es primär an den jeweiligen Fachgemeinschaften, diese zu beantworten.

Gefragt werden muss erstens nach dem «Wie», d. h. zu klären ist, anhand welcher Kriterien und mit welchen Verfahren eine bestimmte Leistung zu beurteilen ist. Die Wie-Frage ist für die Bestimmung der Qualität und der Wirksamkeit einer Leistung relevant. Zweitens ist zu bestimmen, «Was» zu beurteilen ist, also die Frage nach der Leistungsart oder der Leistungsdimension. Die Was-Frage bestimmt das Portefeuille bzw. das Angebot und damit in der Regel auch den Ressourceneinsatz einer Fakultät, eines Departementes oder eines Ordinariates. Das von der Schweizerischen Rektorenkonferenz (CRUS) für die Jahre 2008–2016 lancierte Programm «*Mesurer les performances de la recherche en sciences humaines et sociale*» bietet die Möglichkeit, die Verfahren und Instrumente der Leistungsmessung sowie deren Einsatz entlang der beiden Fragen kritisch zu reflektieren und dabei zugleich einen Beitrag zu deren künftigen Ausgestaltung zu leisten (<http://www.swissuniversities>).

ch/de/hochschulraum/projekte-und-programme). Deshalb hat die Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW) ihre Fachgesellschaften aufgefordert, sich zu überlegen, für welche Leistungen sie anhand welcher Kriterien beurteilt werden wollen.

In den Kernbereich und in die Kernkompetenz fällt die Frage nach den Kriterien guter Forschung und damit die Frage, wie zu messen ist. Dabei zeigt sich, dass in Mehrparadimgenwissenschaften, denen die überwiegende Mehrheit der Sozial- und Geisteswissenschaften zuzuordnen sind, kein gemeinsames Verständnis guter Forschung besteht. Namentlich in der nicht standardisierten, qualitativen, von konkurrierenden Schulen und sich teilweise widersprechender methodischer Ansätze dominierten Forschung fehlte bislang und wohl bis heute in der Praxis der Begutachtung ein gemeinsam geteiltes Set von Beurteilungskriterien. Die Vermutung, dass Schismen zwischen konkurrierenden Schulen die schlechte Erfolgsquote qualitativer Sozialforschung mitverantworten, war dem Wissenschaftspolitischen Rat der Sozialwissenschaften (CPS) der SAGW Anlass, in einem mehrjährigen Prozess unter Einbezug zahlreicher Forschenden «Das Manifest qualitative Sozialforschung» zu erarbeiten. Dabei wurden neun Qualitätskriterien formuliert, entlang derer die qualitative Forschung beurteilt werden soll und einen für die Beurteilungspraxis bedeutsamer Grundsatz formuliert: «Jede Qualitätsbeurteilung einer qualitativen Studie muss die Prämissen, auf denen die Studie beruht, berücksichtigen. Jedes qualitative Forschungsprojekt muss anhand des ihm zugrunde liegenden Paradigmas beurteilt werden» (SAGW 2010, 19–20).

Nicht allein Methodenpluralismus, sondern auch unterschiedliche Wissenschaftsverständnisse führen dazu, dass auch die seit langem praktizierte, kaum hinterfragte und entsprechend breit akzeptierte Peer Review einer kritischen Prüfung nicht stand hält: Forschungsarbeiten zeigen, dass Gutachten sehr unterschiedlich ausfallen, sich die

hohe Streuung hingegen stark reduziert, wenn die Gutachter Kenntnis von den Beurteilungen anderer haben (siehe u. a. Lamont 2009). Dies kann als Hinweis dafür gedeutet werden, Panels einzusetzen, welche die Güte einer Arbeit im Diskurs aushandeln, was gleichsam als Surrogat für den gelehrten Disput gesehen werden kann. Aufschlussreich haben Hans-Dieter Daniel, Sven Hug und Michael Ochsner in verschiedenen Arbeiten die unterschiedlichen Einschätzungen zugrundeliegenden Ursachen aufgedeckt (SAGW 2011, 42–43; SAGW 2012a, 157–171): In mehreren Fragerunden erfassten sie mittels Repertory-Grid-Interviews und Delphi-Technik die subjektive Sicht der Forschenden auf Forschungsqualität fachspezifisch und fachübergreifend. Ziel war es dabei, das Qualitätsverständnis explizit zu machen, um über mehrstufige Validierungsrunden zu Qualitätskriterien zu gelangen, die einem Fachbereich angemessen sind und von den in diesem Bereich tätigen Forschenden anerkannt werden. Deutlich wurden dabei zumindest zwei fundamentale Sachverhalte:

1. Die Einschätzung von Qualitätskriterien hängt unter anderem mehr vom Forschungsverständnis und von der Position im Forschungssystem ab: Faktor- und Clusteranalysen von Qualitätskriterien lassen ein traditionelles und ein modernes Konzept von Forschung erkennbar werden. ProfessorInnen, Habilitierte und DoktorandInnen unterscheiden sich bei der Einschätzung und Gewichtung von einzelnen Qualitätskriterien deutlich. Dasselbe gilt für verschiedene Fachbereiche.
2. Es ist indes möglich, auch zu fachübergreifend akzeptierten Qualitätskriterien zu gelangen. Das Qualitätsverständnis muss jedoch in jedem Fach ausgehandelt und explizit gemacht werden. Ebenso sollte ein breites Spektrum von möglichen Kriterien berücksichtigt werden.

Wenden wir uns nun noch kurz der ebenso bedeutsamen Frage zu, was denn überhaupt zu beurteilen ist und damit der Frage nach der Leistungsart oder

Tabelle 1 Bewertungsmatrix: Forschungsrating Anglistik/Amerikanistik (Stand: Juni 2011)

Bewertungskriterium	Bewertungsaspekt	Daten (Auswahl)
I. Forschungsqualität	1. Qualität des Outputs	Qualitativ: z. B. Exemplarische Publikationen, Publikationsliste nach Typen
	2. Quantität des Outputs	Quantitativ: z. B. Anzahl wissenschaftlicher Publikationen nach Typen
II. Reputation	1. Anerkennung	Qualitativ: z. B. Wissenschaftliche Preise und Auszeichnungen
	2. Professional Activities	Qualitativ: z. B. Herausgeberschaften, Gutachtertätigkeiten, Ämter in Fachverbänden
III. Forschungsermöglichung	1. Drittmittelaktivitäten	Qualitativ: z. B. Liste wissenschaftlich begutachteter Drittmittelprojekte Quantitativ: z. B. Ausgabesummen wissenschaftlich begutachteter Drittmittel
	2. Nachwuchsförderung	Qualitativ: z. B. Liste der abgeschlossenen Promotionen, Liste strukturierter Promotionsprogramme/Graduiertenkollegs Quantitativ: z. B. Anzahl der Promotionsprogramme, Anzahl der drittmittelfinanzierten Doktoranden
	3. Infrastrukturen und Netzwerke	Qualitativ: z. B. Liste Sammlungen, Archive, digitaler Datenbanken
IV. Transfer an ausser-universitäre Adressaten	1. Personaltransfer	Qualitativ: z. B. Weiterbildungsangebote
	2. Wissensvermittlung	Qualitativ: z. B. Medien und Öffentlichkeitsarbeit

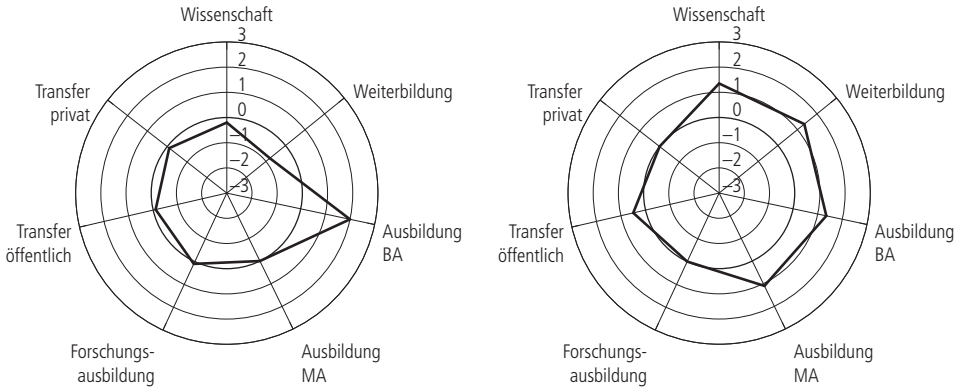
In Anlehnung an: Wissenschaftsrat (2011), Bewertungsmatrix: Forschungsrating Anglistik/Amerikanistik (Stand: Juni 2011), Köln: Wissenschaftsrat.

der Leistungsdimension. Zwei meines Erachtens überzeugende Ansätze seien hier kurz präsentiert. Vier zentrale Leistungsdimensionen hat der Deutsche Wissenschaftsrat mit der Forschung, der Lehre, der Nachwuchsförderung und dem Wissenstransfer in der untenstehenden Bewertungsmatrix identifiziert und damit zugleich den der universitären Lehre und Forschung seitens der Gesellschaft zugedachten Auftrag abgebildet (SAGW 2012a, 161):

Der Mehrdimensionalität tragen ebenso die von Diana Ingenhoff für die Kommunikations- und Medienwissenschaften ebenfalls “bottom up” erhobenen Aktivitätsprofile Rechnung (SAGW 2011, 44–45). Mit Blick auf die notwendige und erwünschte Vielfalt sind Aktivitätsprofile überdies für die Orientierung und Positionierung von wissenschaftlichen Einheiten und Instituten äusserst hilfreich.

Überzeugt davon, dass die Forschung wie zahlreiche andere Bereiche der Gesellschaft rechen-schaftspflichtig sind, und gestützt auf die hier kurz dargelegten Befunde, hat die SAGW eine klare und einfache Position zur Frage der Leistungsbeurteilung formuliert (SAGW 2012b, 32–37): Eine selbstbestimmte Leistungsbeurteilung stärkt die Geistes- und Sozialwissenschaften, weil diese zur Sichtbarkeit ihrer Arbeiten beiträgt, eine Verständigung über Standards, Kriterien und Ziele erfordert und damit wesentliche Voraussetzungen für die dringend gebotene Stärkung der Zusammenarbeit schafft. Die Aushandlung von gemeinsam geteilten Qualitäts- und Leistungskriterien trägt entscheidend zu einem gemeinsamen Selbstverständnis bei und damit zur Stärkung unserer spezifischen Wissenschaftskultur, die sich gegenüber anderen und mächtigeren Wissenschaftskulturen behaupten

Abbildung 1 Aktivitätsprofile: Zwei Beispiele



ten muss. Dies erfordert aber die Bereitschaft der Fachgesellschaften, diese Diskussion engagiert und offen zu führen. Die Evaluationspezialistin Michèle Lamont hat die hohe Bedeutung dieses Diskurses für die Geistes- und Sozialwissenschaften vortrefflich auf den Punkt gebracht:

Finally, a concern for the use of consistent criteria and the bracketing of idiosyncratic taste is more salient in the sciences than in the social sciences and the humanities, due in part to the fact that in the latter disciplines evaluators may be more aware of the role played by (inter) subjectivity in the evaluation process. While the analogy of democratic deliberation appears to describe well the work of the social sciences and humanities panels, the science panels may be best described as functioning as a court of justice, where panel members present a case to a jury. (SAGW 2011, 47–48)

Die Position von Lamont kann als Plädoyer für den gelehrten Disput in einer den veränderten Verhältnissen angepassten Form gelesen werden. Wenn nicht eine von den Fachgesellschaften geführte Debatte über Standards und Güterkriterien, what else?

Literatur

- Lamont Michèle (2009). *How Professors Think – Inside the Curious World of Academic Judgement*. Cambridge: Harvard University Press.
- SAGW (2010). *Manifest Methoden qualitativer Sozialforschung*. Bern.
- SAGW (2011). *Bulletin 2, Dossier Qualitäts- und Leistungsmessung in den Geisteswissenschaften*. Bern.
- SAGW (2012a). Ochsner Michael, Hug Sven E. und Daniel Hans-Dieter, «Wie wollen und sollen die Geisteswissenschaften Qualität und Leistung messen und steuern», in: SAGW 2012a. *Für eine neue Kultur der Geisteswissenschaften?*, S. 157–171.
- SAGW (2012b). *Positionspapier. Für eine Erneuerung der Geisteswissenschaften*. Bern.
- SWR (1998). *Beitrag zu einem Indikatorensystem für die Wissenschafts- und Technologiepolitik*. Bern.
- SWR (1999a). *Forschungslandkarte Schweiz 1997. Bibliometrische Indikatoren der schweizerischen Forschung in den Jahren 1993–1997*. Bern.
- SWR (1999b). *Forschungsstatus Schweiz 1998. Bibliometrische Wissenschaftsindikatoren zur schweizerischen Forschung im internationalen Vergleich*. Bern.

Die Publikationsleistungen der universitären Soziologieabteilungen in der Schweiz

Jörg Rössel und Ilona Pap (Universität Zürich)

Einleitung

Evaluationen von Forschungsleistungen und darauf basierende Rankings von Forschungseinrichtungen und WissenschaftlerInnen sind ein hochgradig umstrittenes Thema, insbesondere in den Geisteswissenschaften, zum Teil aber auch in den Sozialwissenschaften. Auf der einen Seite kann auf verschiedene Argumente verwiesen werden, die für derartige Evaluationen und Rankings sprechen: Erstens liegt es nahe, dass die durch öffentliche Mittel finanzierte Wissenschaft gegenüber eben jener Öffentlichkeit auch rechenschaftspflichtig ist und ihre Leistungen ausweisen muss. Zweitens sind universitäre Wissenschaftler in ihrem Alltag regelmässig im Sinne von Evaluationen beschäftigt: In der Regel evaluieren wir ja nicht nur Hypothesen (siehe den Beitrag von Münch in diesem Heft), sondern wir bewerten die Leistungen von Studierenden, von NachwuchswissenschaftlerInnen, die Qualität von Publikationen sowie Forschungsanträgen und wir rekrutieren Bewerber für Stellen auf der Grundlage ihrer Leistungen, die wir selbstverständlich vorher evaluieren. Drittens ist es in einem kompetitiven Feld wie den Wissenschaften durchaus gerechtfertigt, Ressourcen und Anerkennung auf der Basis von Leistungsvergleichen zu allozieren. Neben den offenkundig existierenden ideologischen Vorbehalten gegenüber Leistungsbewertungen in den Sozialwissenschaften können gegen Evaluationen aber auch plausible Argumente ins Feld geführt werden. Erstens sollte nicht unterschlagen werden, dass Evaluationen Zeit und andere Ressourcen verbrauchen und daher das Leistungspotenzial der Wissenschaft sogar schmälern können (Frey 2006). Zweitens gibt es durchaus auch kritische Fragen zur Reliabilität von Leistungsbewertungen, die allerdings durch

sorgfältige Methoden überwiegend ausgeräumt werden können (siehe den Beitrag von Hinz in diesem Band). Drittens wird häufig darauf verwiesen, dass generell Leistungen mit unterschiedlichen Dimensionen erfasst werden müssen und dass dies insbesondere für multiparadigmatische Wissenschaften wie die Soziologie gilt. Viertens können Evaluationen auch Fehlanreize setzen, so dass Wissenschaftler nur noch die evaluierten Leistungskriterien maximieren und dabei andere wichtige Aufgaben vernachlässigen.

Die skizzierte kritische Diskussion mahnt zu einer gewissen Vorsicht bei der Durchführung von Evaluationen. Wir wollen hier dennoch im Anschluss an den Beitrag von Bernauer und Gildardi (2010), der die Publikationsleistungen in der schweizerischen Politikwissenschaft analysiert, in diesem kurzen Aufsatz den Publikationsoutput der verschiedenen soziologischen Abteilungen an den Universitäten in der Schweiz untersuchen. Dafür gibt es eine Reihe von Gründen. Erstens scheint es uns wichtig, mit den politischen Wissenschaften gleichzuziehen und eine Diskussion über Evaluationen und Rankings auch in der schweizerischen Soziologie anzustossen. Das Beispiel der von Jürgen Gerhards (2002) in Deutschland angestossenen Diskussion über Rankings in der Soziologie zeigt deutlich, dass es derartige explorative Analysen braucht, um eine weitere Reflexion in Gang zu bringen. Zweitens fokussiert die Analyse auf Publikationsleistungen, da diese die Quantität und Qualität von wissenschaftlichen Leistungen – im Gegensatz zum Beispiel zu inputbasierten Messungen, wie Drittmitteln – am besten erfassen können. Dies wird auch in dem Beitrag von Hinz sehr deutlich (in diesem Heft), der zeigt, dass sich eine Vielzahl von Dimensionen der Leis-

tungsbewertung statistisch hervorragend in den Publikationsziffern bündeln lassen. Drittens ist auch in einer multiparadigmatischen Wissenschaft unbestritten, dass Publikationen den wichtigsten Leistungsindikator darstellen. Um allerdings eine zu einseitige Perspektive zu vermeiden, werden unterschiedliche Messungen von Publikationen berücksichtigt, die einerseits auf dem ISI Web of Knowledge (das auch den Social Science Citation Index und andere Datenbanken einschliesst) basieren, das vor allem Publikationen in internationalen, begutachteten Fachzeitschriften erfasst, und andererseits wurde Google Scholar in die Analyse miteinbezogen, das Publikationen in ganz unterschiedlichen Arten von Medien einschliesst. Damit können mögliche Einschränkungen, die durch eine Verwendung von nur einer Datenbank entstehen, vermieden werden (Frey und Rost 2010). So werden in Google Scholar auch Publikationen in den Landessprachen, Bücher und Sammelbandbeiträge berücksichtigt, die im ISI Web of Knowledge unterrepräsentiert sind. Insofern sollte auf dieser Grundlage eine gute Einschätzung der Publikationsleistungen der universitären Soziologie in der Schweiz möglich sein. Schliesslich kann es sich hier nur um eine explorative Analyse handeln, da erstens nur Professuren berücksichtigt werden und zweitens die Auswertung der Datenbanken von zahlreichen Entscheidungen abhängig ist. Insofern können die Einzelergebnisse auch nicht ohne weiteres reproduziert werden. Die Muster des Publikationsoutputs sind gegenüber diesen Designentscheidungen sehr robust.¹ Insbesondere muss an dieser Stelle hervorgehoben werden, dass der Umfang von Publikationen und Zitationen stark vom Alter eines Wissenschaftlers abhängig ist. Dies haben wir in unserer Studie nicht korrigiert. Standorte mit einer überdurchschnittlichen Anzahl von jüngeren WissenschaftlerInnen könnten durch diese Entscheidung benachteiligt sein. Eine

1 Eine ausführliche Dokumentation der Vorgehensweise, die eine Replikation der Ergebnisse ermöglicht, kann bei den Autoren angefordert werden.

abschliessende Analyse der Publikationsaktivität der Schweizer Soziologie müsste im Rahmen eines ausführlichen Projektes behandelt werden und nicht in einem kurzen Bulletinbeitrag. Dieser kann nur als Anregung für weitere Diskussionen dienen.

Daten und Methoden

Die explorative Studie folgt in wesentlichen Zügen den Vorgaben der Untersuchung von Bernauer und Gilardi (2010). Für die Analyse berücksichtigt wurden alle Inhaber von Professuren (einschliesslich Assistenz- und Forschungsprofessuren) an neun universitären Hochschulen in der Schweiz.² Insgesamt konnten 62 Personen in die Untersuchung aufgenommen werden. Sowohl die Analyse des ISI Web of Science, wie auch die Analyse von Google Scholar wurde an bestimmten Stichdaten durchgeführt, so dass sich keine Verzerrungen aus unterschiedlichen Erhebungszeitpunkten ergeben konnten.³ Das ISI Web of Science enthält Informationen über Publikationen in über 2000 Zeitschriften in 50 sozialwissenschaftlichen Disziplinen. Wir haben Publikationen in allen Fachrichtungen berücksichtigt, allerdings wurden nur Aufsätze in die Analyse aufgenommen (keine Rezensionen, Editorials usw.). Dagegen werden in Google Scholar Publikationen aller Art, von Monografien, über Sammelbände, Handbücher bis hin zu Artikeln in Publikumszeitschriften berücksichtigt. Dies ist eine wichtige Ergänzung, da gerade in der Soziologie das Publizieren in Fachzeitschriften, anders als in der Politikwissenschaft, nicht durchgängig akzeptiert ist. In diesen beiden Datenbanken wurden jeweils drei Indikatoren erhoben: erstens die Gesamtzahl von Publikationen einer Person, zweitens die Anzahl von Zitationen (einschliesslich

2 Die Liste der Personen wurde im Vorstand der SGS und mit Vertretern der jeweiligen Hochschulen validiert und kann auf Anfrage von den Autoren zur Verfügung gestellt werden. Die Personen waren jeweils zum 31.12.2014 Inhaber der Professuren.

3 Für die Analyse von Google Scholar wurde die Software «Publish-or-Perish» verwendet. www.harzing.com.

Tabelle 1 Zusammenhang zwischen den Indikatoren (Google Scholar und ISI)

	Werte aus Google Scholar		h-Index
	Anzahl Publikationen	Anzahl Zitationen	
Werte aus ISI	0.52 (4.725)	0.68 (7.27)	0.69 (7.312)
R ²	0.27	0.47	0.47

Angegeben sind die standardisierten Regressionskoeffizienten und die t-Werte in Klammern, N=62.

Selbstzitationen) dieser Publikationen und drittens der sogenannte *b*-Index. Der *b*-Index gibt die Anzahl (*b*) von Publikationen an, die eine bestimmte Mindestzahl von Zitationen (*b*) aufweisen. Hat ein Forscher also einen *b*-Index von 3, so weist er drei Publikationen, mit mindestens drei Zitationen auf. Die Konstruktion des *b*-Index «belohnt» eine kontinuierliche Publikationsleistung, die kontinuierlich Zitationen erzeugt. Dagegen «bestraft» er Forscher, die eher wenige Publikationen mit sehr vielen Zitationen aufweisen.

Angeichts der Tatsache, dass ISI Web of Science und Google Scholar ganz unterschiedliche Arten von Publikationen aufnehmen, kann man sich fragen, ob auf der Grundlage der beiden Datenbanken überhaupt eine konsistente Publikationsleistung der verschiedenen Abteilungen für Soziologie gemessen wird. Um dies zu prüfen, wurden die Korrelationen zwischen den drei Indikatoren berechnet, jeweils auf der Grundlage von Google Scholar und ISI Web of Science. Ähnlich wie dies Bernauer und Gilardi (2010) auch für die Politikwissenschaft in der Schweiz feststellen konnten, weisen die Ergebnisse einen sehr starken Zusammenhang für die Messungen auf der Grundlage von ISI Web of Science und Google Scholar nach. Die standardisierten Regressionskoeffizienten liegen zwischen 0.5 und 0.7. Dies würde dafür sprechen, dass die beiden Messungen eine konsistente Erfassung der zugrundeliegenden Grösse, der Publikationsleistung, ermöglichen. Zudem spricht es gegen das Vorliegen einer Verzerrung bei der Messung.

Ergebnisse

Publikationen

Die Ergebnisse der Analyse sollen nun für die drei Indikatoren vorgestellt werden. Dabei werden die Resultate immer auf der Ebene der Abteilungen, nicht auf der Ebene von Personen, präsentiert, um erstens eine gewisse Anonymität zu wahren und zweitens tatsächlich die Leistung von Forschungseinheiten zu vergleichen (Hinz in diesem Band). Damit wird auch der kollektive Charakter von Forschungsleistungen ansatzweise erfasst, der von Münch (siehe den Beitrag in diesem Heft) betont wird. Für die grafische Darstellung greifen wir im Anschluss an Bernauer und Gilardi (2010) auf Boxplots zurück. In den Boxplots wird der Median für jede Abteilung durch eine fett gedruckte Linie dargestellt. Die Boxen um diese fett gedruckte Linie stellen die 25 und 75% Perzentile dar. Die Whiskers markieren das erste und das neunte Dezil. In jeder Grafik wird auch der Durchschnitt für alle soziologischen Abteilungen in der Schweiz durch eine gestrichelte Linie dargestellt. Zudem ist zu beachten, dass die y-Achse eine logarithmische Skalierung aufweist. Diese Skala wurde gewählt, da sonst die Grafiken zu weit auseinandergezogen wären und damit unübersichtlich wären. Zu den Angaben wurde 1 addiert, worauf sie mit dem natürlichen Logarithmus transformiert wurden.

Die Resultate in den beiden Grafiken für die Publikationen zeigen deutlich, dass die Soziologie an der ETH, der Universität Bern und an der Universität Zürich in der Schweiz den quantitativ umfangreichsten Publikationsoutput aufweisen.

Abbildung 1 Publikationen im ISI (rangiert nach Median)

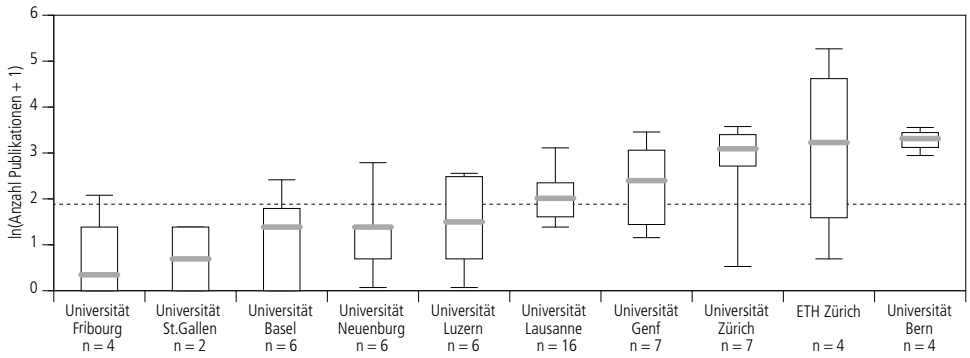
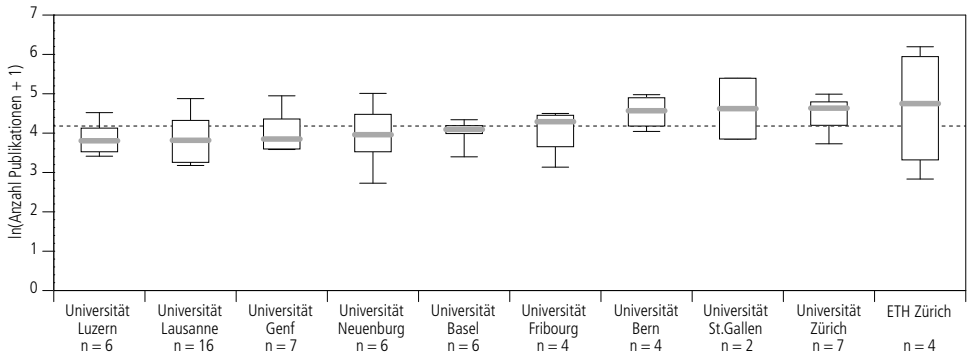


Abbildung 2 Publikationen in Google Scholar (rangiert nach Median)



Während die Universität Zürich jeweils an dritter Stelle steht, wechseln die Positionen der ETH und Bern je nach Messung. Die Soziologie in St. Gallen kann allerdings bei der Messung über die Datenbank Google Scholar auch einen relativ hohen Wert erreichen. In der Romandie erreicht die Universität Genf besonders hohe Werte, wenn man die ISI Web of Science Datenbank heranzieht. Misst man die Publikationen in Google Scholar, weist Fribourg den höchsten Wert in der Romandie auf.

Zitationen

Betrachtet man im nächsten Schritt die Anzahl von Zitationen, die die Publikationen aus den verschie-

denen Soziologieabteilungen auf sich versammeln können, so zeigt sich ein vergleichbares Bild wie beim Publikationsoutput. Die Soziologie an der ETH, der Universität Bern und der Universität Zürich weist die grösste Anzahl von Zitationen auf. Im Gegensatz zur Publikationsmessung basierend auf Google Scholar weist die Universität St. Gallen bei den Zitationen keinen besonders hohen Wert auf. Anders als bei den Publikationen im ISI Web of Science ist bei den Zitationen in der Romandie die Universität Lausanne führend.

Abbildung 3 Zitationen im ISI (rangiert nach Median)

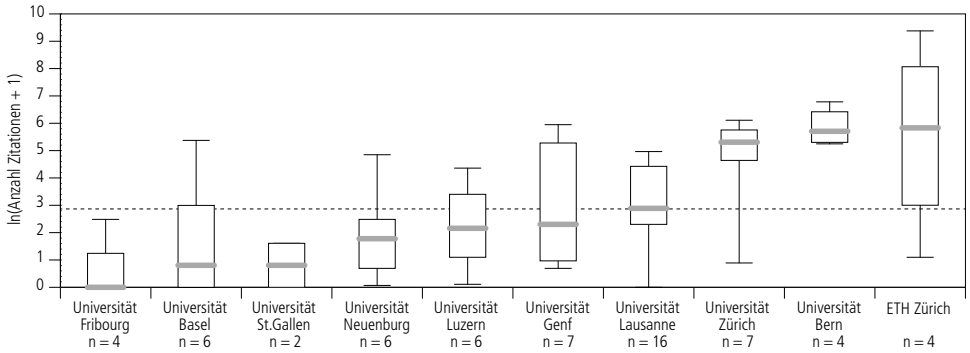
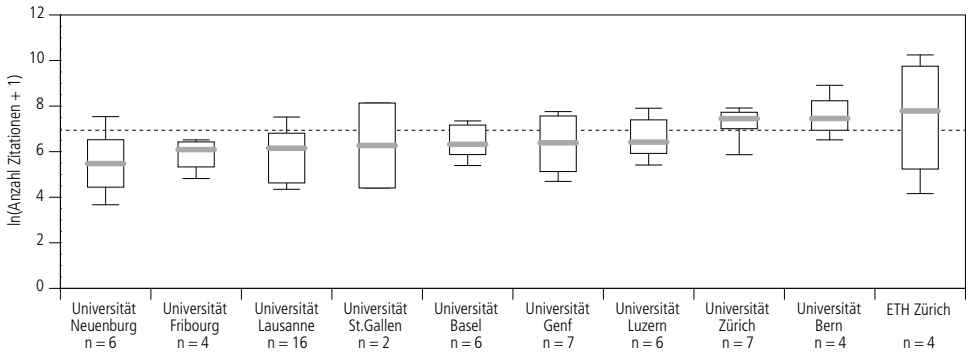


Abbildung 4 Zitationen in Google Scholar (rangiert nach Median)



h-Index

Die Analyse des *h*-Index ergibt im Wesentlichen keine neuen Resultate. Weiterhin liegen die ETH, die Universität Bern und die Universität Zürich mit ihren Publikationsleistungen an der Spitze, allerdings ist die Universität Lausanne in der Romandie führend in ihren Publikationsleistungen.

Zusammenfassung

Die vorliegende Analyse basiert auf der Prämisse, dass es durchaus gute Gründe für das Durchführen von Evaluationen auch in den Sozialwissenschaften gibt. Allerdings wird mit dieser kleinen Publikation keine abschliessende Bewertung der Publikations-

leistungen der verschiedenen universitären soziologischen Abteilungen in der Schweiz angestrebt, sondern lediglich eine explorative «Probebohrung», die Anlass für weitere Diskussionen und für zukünftige genauere Studien geben soll.

Betrachtet man die Resultate der durchgeführten Analyse, so lassen sich vor allem drei Hauptergebnisse feststellen. a) Die Ergebnisse der Auswertungen des ISI Web of Science und von Google Scholar führen in hohem Masse zu übereinstimmenden Ergebnissen. Dies spricht gegen die Vorstellung, dass die Berücksichtigung von unterschiedlichen Dimensionen bei der Evaluation von Forschungsleistungen die Resultate stark

Abbildung 5 h-Index im ISI (rangiert nach Median)

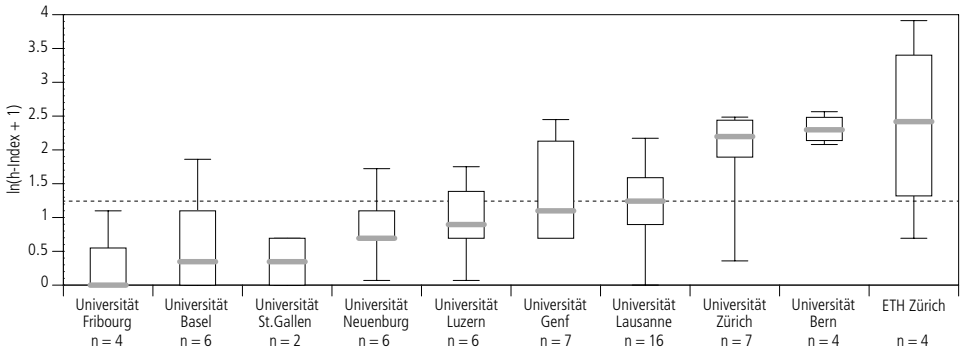
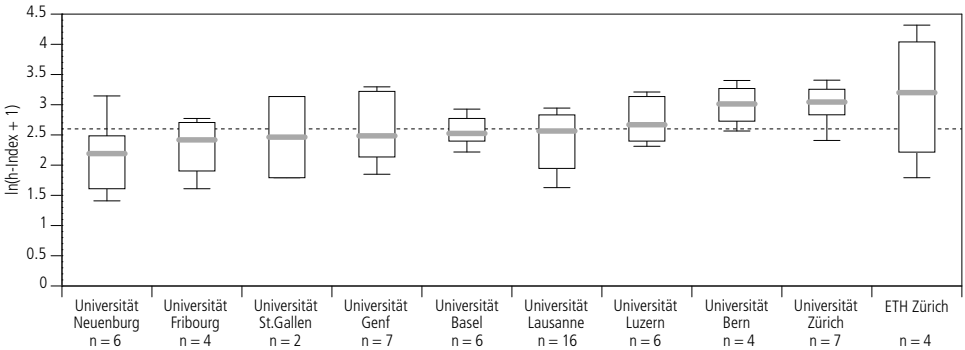


Abbildung 6 h-Index in Google Scholar (rangiert nach Median)



beeinflusst. Obwohl die beiden Datenquellen ganz unterschiedliche Arten von Publikationen erfassen, sind die Ergebnisse hochgradig übereinstimmend. Einheiten, die viele Publikation im ISI Web of Science aufweisen, publizieren also auch in anderen Kontexten überdurchschnittlich viel. b) Die drei in ihrem Publikationsoutput führenden soziologischen Einheiten in der Schweiz finden sich an der ETH, der Universität Bern und an der Universität Zürich. Dabei sieht die Reihenfolge für fast alle Indikatoren und Messungen folgendermassen aus: Die ETH erweist sich als publikationsstärkste Einheit und danach folgen die Universitäten Bern und Zürich dicht aufeinander. In der Romandie

sind die Universität Lausanne und die Universität Genf die publikationsstärksten Einheiten. c) Dieses Ergebnis variiert in Abhängigkeit von der Messung, ob man also ISI oder Google Scholar verwendet, ob man sich Publikationszahlen, Zitationen oder den *h*-Index anschaut, nur geringfügig. Insofern scheint es sich hier um eine robuste Messung zu handeln. Damit wird deutlich, dass die vorliegende Analyse vielen skeptischen Einschätzungen von Evaluationen und Rankings widerspricht. Damit dürfte sie auch eine gute Grundlage für weitere Diskussionen zu diesem Thema in der Schweizer Soziologie darstellen.

Literatur

- Bernauer, Thomas und Fabrizio Gilardi (2010). Publication Output of Swiss Political Science Departments. *Swiss Political Science Review* 16: 279–303.
- Frey, Bruno S. (2006). Evaluitis – Eine Neue Krankheit? Institute für Empirical Research in Economics. University of Zurich. *Working Paper Series* No. 293.
- Frey, Bruno S. und Katja Rost (2010). Do Rankings Reflect Research Quality? *Journal of Applied Economics* 13: 1–38.
- Gerhards, Jürgen (2002). Reputation in der deutschen Soziologie – zwei getrennte Welten. *Soziologie* 2/2002: 19–33.

Weitgehend überraschungsfrei, folgenlos und (so) nicht wiederholbar: das Forschungsrating der Soziologie in Deutschland

Thomas Hinz (Universität Konstanz)

Allgemeiner Hintergrund

Mitte der letzten Dekade startete der Wissenschaftsrat (WR) in Deutschland, eine Einrichtung getragen von Vertretern von Bund, Ländern und Wissenschaft zur kontinuierlichen Politikberatung in Sachen Gestaltung und Reformierung des deutschen Wissenschaftssystems, eine Reihe von fachbezogenen Evaluationen. Das Ziel bestand in einer belastbaren Bewertung von Forschungsleistungen der in Deutschland aktiven wissenschaftlichen Einrichtungen. Ganze Fächer und ihre Vertretung an Universitäten und ausseruniversitären Forschungsinstituten sollten auf den Prüfstand – und dies in einem transparenten und von den beteiligten WissenschaftlerInnen möglichst weit akzeptierten Verfahren. Der WR betonte damals vor allem den Informationsbedarf von Wissenschaft, Politik und Öffentlichkeit. Das angestrebte Forschungsrating sollte den verantwortlichen Akteuren möglichst hilfreiche und transparente Informationen zur Qualität der geleisteten Forschung liefern. Ausdrücklich war kein Forschungsranking beabsichtigt. Und anders als in Grossbritannien, wo zunächst von 1986 bis 2008 die *Research Assessment Exercise* (RAE) und anschliessend das *Research Excellence Framework* (2014) durchgeführt wurden, war mit dem Forschungsrating keine Allokation von Forschungsmitteln verbunden. Allerdings waren substantielle Schlussfolgerungen aus dem Rating nicht ausgeschlossen. Zunächst sollten wissenschaftsadäquate Methoden und Verfahrensweisen erprobt werden. Die damit angestrebte Bewertung von Forschungsleistungen war in Deutschland wie andernorts mit einer Reihe von grundsätzlich zu klärenden Fragen verbunden: Wie können die wissenschaftlichen Leistungen möglichst objektiv,

reliabel und valide gemessen werden – und dies in einem möglichst akzeptierten, transparenten und vom Aufwand her vertretbaren Verfahren? Welche Rolle spielen dabei Kriterien, die im eigentlichen Sinn nicht wissenschaftsimmanent sind – wie der Transfer wissenschaftlicher Erkenntnisse in die Öffentlichkeit? Welche Eigenheiten der unterschiedlichen Wissenschaftsdisziplinen müssen dabei berücksichtigt werden?

Den Hintergrund für diese Fragen bildeten Debatten um die Messbarkeit von wissenschaftlichen Leistungen. In der Wissenschaft ist die Bewertung von Qualität und Leistungen auf die Urteile von Peers angewiesen (Bornmann und Daniel 2003). WissenschaftlerInnen forschen dann erfolgreich, wenn sie unter ihresgleichen anerkannt werden. Peers sind für die Bewertung von eingereichten Publikationen und Forschungsanträgen sowie für die Bestenauslese im Rahmen von Preiskomitees und Berufungsverfahren verantwortlich. Die Urteilskraft der Peers wird insbesondere deshalb akzeptiert, weil der Wissenszuwachs meist kumulativ organisiert ist, sich Wissenschaft immer mehr spezialisiert, oder mit anderen Worten, Spezialisten mehr und mehr für die Rezeption und Anerkennung durch andere Spezialisten forschen und veröffentlichen. Letztlich ist das Verfahren des Peer Reviews trotz einiger Kritik, etwa an der möglichen Bevorzugung von etablierter und risikoloser Forschung bzw. Blindheit gegenüber Innovationen (Münch 2009), alternativlos. Die Urteile von Peers sind für die Bewertung von Forschungsleistungen also zentral. Gleichwohl erhalten Fragen der wirtschaftlichen Verwertung von wissenschaftlichen Erkenntnissen, etwa durch Patente oder Fragen der gesellschaftlichen Relevanz seit einigen Jah-

ren vermehrt Aufmerksamkeit. Die verständliche Vermittlung von Wissenschaft in der (und in die) Öffentlichkeit wird zunehmend gefordert und gefördert. Diese Anforderungen nach allgemeiner Verwertbarkeit und Verständlichkeit besitzen im Vergleich zu den Urteilen der Peers jedoch eine klar nachrangige Bedeutung. Das Forschungsrating des WR sollte allerdings bewusst mehrdimensional ausgerichtet sein, weil der WR eben auch die Interessen der Öffentlichkeit und Politik vertritt. Neben der Forschungsleistung sollten Faktoren wie Förderung jüngerer WissenschaftlerInnen und Transfers in andere gesellschaftliche Bereiche beurteilt werden. Von 2006 bis 2011 wurden die Fächer Soziologie, Chemie, Elektro- und Informationstechnik sowie Anglistik/Amerikanistik evaluiert. Weiterhin wurden Vorüberlegungen zur Anwendbarkeit des Forschungsratings in den Geisteswissenschaften angestellt.

Forschungsrating in der Soziologie

Die Soziologie wurde bereits im Jahr 2006 als sozialwissenschaftliches Fach für eine Pilotstudie ausgewählt (vgl. zum Vorgehen und zu Ergebnissen Neidhardt 2006, 2008, 2009). Das eingesetzte Evaluationsverfahren bestand in einem sog. *informed Peer Review*. Dies bedeutete, dass in der Bewertungsgruppe unter Vorsitz von Friedhelm Neidhardt nicht nur Kennzahlen heranzuziehen waren, sondern diese Kennzahlen, etwa die Anzahl der Publikationen und Drittmittelwerbungen, durch ein Expertengremium (eben die «Bewertungsgruppe») reflektiert und diskutiert werden sollten. Die Mitglieder dieser Bewertungsgruppe sollten die von den begutachteten Einrichtungen eingesandten Schriften lesen und ihre Leseindrücke bei der anschliessenden Diskussion der Forschungsqualität einbringen. Insgesamt deckte das Forschungsrating in der Soziologie folgende Dimensionen ab: Forschungsqualität, Impact/Effektivität, Effizienz, Nachwuchsförderung, Transfer, Wissensvermittlung. Als wissenschaftsimmanent sind vor allem die ersten beiden Kriterien der For-

schungsqualität und des erzielten Impacts zu sehen. Zur Bewertung der Forschungsqualität wurden vor allem die Veröffentlichungen, zur Bewertung des Impacts wurde zusätzlich die Einwerbung qualifizierter Drittmittel herangezogen.

Eine zentrale operative Frage des Forschungsratings lautete, auf welcher organisatorischen Ebene die Evaluation angesetzt werden sollte. Man unterschied letztlich zwischen Forschungseinrichtungen, also Universitäten oder ausseruniversitären Forschungsinstituten, und Forschungseinheiten, meist Professuren oder Arbeitsgruppen innerhalb dieser Einrichtungen. Die konkrete Definition der zu bewertenden Forschungseinheiten überliess man letztlich den Forschungseinrichtungen, was zu einer gewissen Uneinheitlichkeit führte. An einer Universität wurden alle Professuren als eine gemeinsame Forschungseinheit begriffen, an den meisten Universitäten orientierte sich die Definition der Forschungseinheiten an den Professuren. In der Konsequenz konnten die Bewertungen nicht über alle Einheiten sinnvoll verglichen werden.

Die Pilotstudie in der deutschen Soziologie stand vor weiteren Herausforderungen. Im Fach Soziologie ist die Veröffentlichungskultur sehr heterogen. Vor allem jüngere und international orientierte WissenschaftlerInnen versuchen, ihre Forschungsergebnisse in Zeitschriften mit Peer Review Verfahren zu publizieren. Daneben halten sich jedoch auch Buchpublikationen oder Beiträge für thematisch orientierte Sammelbände. Gängige Datenbanken für Publikationen, etwa das *Web of Science* oder *Scopus*, erfassen jedoch nur Zeitschriftenpublikationen, und hiervon auch nur eine Auswahl. Alle führenden Zeitschriften der Soziologie in Deutschland werden allerdings im *Social Sciences Citation Index* (SSCI) erfasst; damit waren die Veröffentlichungen in diesen Zeitschriften Teil der Datengrundlage des Forschungsratings. Um die Arbeit der Bewertungsgruppe vorzubereiten, wurde eine Literaturdatenbank erstellt – mit einem wirklich problematischen Resultat: Die Einträge in die Publikationsdatenbanken der Sozialwissenschaften

deckten nur einen Teil aller Veröffentlichungen ab und enthielten Beiträge, die dem Kriterium eines ernsthaften Peer Reviews nicht genügten. So war als empirische Grundlage für das Forschungsrating eine ergänzende Erfassung derjenigen Publikationen nötig, die über die nachgewiesenen Zeitschriftenbeiträge in den Datenbanken hinauswiesen. Dazu gab die Bewertungsgruppe allen begutachteten Einheiten die Möglichkeit, die vom WR vorbereitete Liste mit den in Datenbanken recherchierten Publikationen eigenhändig zu ergänzen. Von Zitationsanalysen wurde abgesehen. Für Bücher und Beiträge in Sammelbänden wären solche Zitationsanalysen möglich, insofern sie in Zeitschriften zitiert werden. Dies ist allerdings mit grossem Aufwand verbunden – und weiterhin natürlich unvollständig. Das Verfahren des informierten Peer Reviews sollte sich nicht auf die Beurteilung von Datenbankeinträgen beschränken: Jede Forschungseinheit sollte (grössenabhängig) eine gewisse Anzahl von Veröffentlichungen an die Bewertungsgruppe einsenden, die dann von zugewiesenen FachgutachterInnen gelesen wurden.

Insgesamt waren 16 FachgutachterInnen zwischen Februar 2006 bis März 2008 mit der Evaluation beschäftigt. Es waren 254 Forschungseinheiten in 57 Forschungseinrichtungen zu begutachten. Für die wissenschaftsimmanenten Dimensionen wurde eine fünfstufige Ordinalskala eingesetzt (mit den Kategorien «exzellent», «sehr gut», «gut», «befriedigend» und «nicht befriedigend»). Jedes Mitglied der Bewertungsgruppe hatte bis zu 80 Einzelbewertungen von Forschungseinheiten abzugeben. Die Gruppe traf sich zu elf Sitzungen an insgesamt 15 Sitzungstagen. Insgesamt wurde pro Mitglied der Bewertungsgruppe eine Arbeitslast von zwei bis drei Arbeitsmonaten angegeben (Neidhardt 2008).

Wie lassen sich die Ergebnisse dieses aufwendigen Prozesses zusammenfassen? Die Bewertungsgruppe beurteilte vier Prozent der Einheiten im Hinblick auf die Forschungsqualität als «exzellent», 18 Prozent als «sehr gut», 38 Prozent als «gut»,

24 Prozent als «befriedigend» und neun Prozent als «nicht befriedigend». Die Bewertungsgruppe konstatierte dabei, dass die LeistungsträgerInnen über die Einrichtungen breit verteilt sind und eine überwiegend kleinteilige Organisation der Forschung an einzelnen Professuren vorherrscht. Dies ist der Tatsache geschuldet, dass die Universitäten in der Regel mit wenigen Professuren ein breites Themenspektrum in der Lehre abdecken wollen. Insgesamt wurde eine schwach ausgeprägte internationale Ausrichtung festgestellt, nur 16 Prozent der herangezogenen Publikationen wurden ausserhalb Deutschlands veröffentlicht. Hinsichtlich der Publikationskultur wurde auch in anderer Hinsicht die vermutete Heterogenität bestätigt: Etwa die Hälfte der Publikationen besteht aus Beiträgen für Sammelbände bzw. der Herausgabe solcher Sammelwerke – zumeist ohne eigene Qualitätssicherung. Nur ein Drittel der Publikationen erschien in Fachzeitschriften, davon wiederum nur ein Drittel in referierten Zeitschriften. Zudem konstatierte die Bewertungsgruppe, dass es anders als in den Naturwissenschaften hohe Forschungsqualität auch an relativ kleinen Einheiten geben könne.¹

Lohnt der Aufwand?

In die Bewertung der Forschungsqualität sind – wie gesagt – die Leseindrücke der Bewertungsgruppe eingeflossen. Wie in einer Ad-Hoc Gruppe beim Soziologiekongress der DGS 2012 in Bochum gezeigt, lässt sich die Bewertung des Ratings der Forschungsqualität von Einheiten und Einrichtungen zu weiten Teilen allein aus der grössengewichteten Anzahl der Aufsatzpublikationen rekonstruieren – und das obgleich die Publikationsdatenbank ja in mancherlei Hinsicht als suboptimal kritisiert wurde.² Lediglich bei einer sehr kleinen Anzahl

- 1 Alle Unterlagen und Berichte sind dokumentiert unter: <http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/forschungsrating/aktuelles.html>.
- 2 Zusammen mit Andreas Diekmann (ETH Zürich), der Mitglied der Evaluationsgruppe war, seinem Mitarbeiter Mathias Naef (ETH Zürich), der die Zitationen der damals angeführten Literatur recher-

der Forschungseinheiten haben die Leseindrücke der Gutachtergruppe zu einer spürbaren Abweichung von der standardisierten Beurteilung nach Perzentilen der gewichteten Publikationsanzahlen in Zeitschriften geführt. Anders gesagt: Die enorm aufwendige Lesearbeit liefert in der Summe kaum zusätzliche Informationen. Wissenschaftsgesteuerte Drittmittelprojekte (oder auch Drittmittelsummen) haben das Ergebnis des Ratings für die Forschungseinheiten kaum oder gar nicht beeinflusst. Diese vielleicht überraschenden Resultate lassen sich indes leicht erklären. Veröffentlichungen in Zeitschriften liegt zumindest bei referierten Zeitschriften bereits ein Qualitätsurteil zu Grunde. Die Mitglieder der Bewertungsgruppe orientierten sich bei der Vergabe ihrer Urteile überdies an dem Perzentilwert der blossen Anzahl der Zeitschriftenaufsätze. Drittmittelprojekte erhöhen zwar naturgemäss die zur Verfügung stehenden Ressourcen, was aber nicht unbedingt zu einer Steigerung der grössengewichteten wissenschaftlichen Produktivität führt. Natürlich gibt es eine Mindestgrösse einer Forschungseinheit, um angesichts eines in Deutschland beachtlichen Beitrags der universitären Einheiten zur Lehre und Verwaltung überhaupt sinnvoll forschen zu können, aber Drittmittelerwerbungen tragen als solche noch nicht zu mehr Forschungsqualität pro MitarbeiterIn bei. Dies erscheint angesichts der Kontrolle auf den Output (eben die Publikationen in Zeitschriften) solcher Drittmittelerfolge auch logisch.

chierte, und Katrin Auspurg (Goethe Universität Frankfurt), hat der Autor versucht, die Bewertungen der Evaluationsgruppe für die Forschungseinheiten aus den zusammengetragenen Daten zu rekonstruieren. Die Ergebnisse wurden bei einer Ad-Hoc Gruppe des Soziologiekongresses vorgestellt und diskutiert. Eine Rekonstruktion auf der Ebene der Forschungseinrichtungen stammt von Riordan, Ganser und Wolbring (2011).

Kritische Würdigung und Empfehlungen zur Weiterentwicklung

Das Forschungsrating der Soziologie wurde im Fach selbst eher skeptisch aufgenommen (Baier 2011; Münch 2009). Zwar wurde im Vergleich zu stärker umstrittenen Erhebungen (etwa das sog. CHE Ranking) die sorgfältige und sehr transparente Durchführung positiv hervorgehoben. Gleichzeitig wurde Kritik an der faktischen Eindimensionalität geübt.³ Die Kritiker betonten, dass das Forschungsrating letztlich eine einseitige und auf die Publikation in Fachzeitschriften beschränkte Konstruktion von soziologischer Exzellenz darstelle und eine Stratifikation in ein zuvor durch gleichberechtigte DiskursteilnehmerInnen gekennzeichnetes Feld einführe (Münch und Baier 2009). Eine einmal etablierte Stratifikation tendiere überdies zur Selbstreproduktion, weil sich die Akteure zukünftig an dieser Konstruktion orientierten.

Der WR selbst gelangte zu einer viel positiveren Einschätzung. Er resümiert, dass mit den Pilotstudien in allen vier Fächern ein Meilenstein im Sinne der Entwicklung eines Modells guter Praxis der vergleichenden Forschungsbewertung erreicht wurde (Wissenschaftsrat 2013). Hervorgehoben wird die Verbindung aus hoher methodischer Qualität, fachspezifischer Anpassungsfähigkeit, potenziell breitem Nutzen und der im Zuge der Pilotstudien bewiesenen Lernfähigkeit. Trotz des damit verbundenen Aufwands wird eine Fortführung und breitere Anwendung des Forschungsratings empfohlen.

Man kann eine abschliessende Würdigung des Forschungsratings für das Fach Soziologie noch einmal anders formulieren. Aufgrund des unbestritten enormen Aufwands der Bewertungsgruppe und damit verbundener vielfältiger Kontingenzen erscheint das Forschungsrating letztlich nicht in derselben Form wiederholbar. Damit sind wichtige

3 Was allerdings nur bestätigt, dass die Forschungsqualität letztlich die anderen Dimensionen dominiert. Man könnte sagen: Genau dieses Resultat ist eigentlich gewollt.

Anforderungen an die Reliabilität des Verfahrens nicht erfüllt. Weiterhin muss bedacht werden, dass der Bedarf am Forschungsrating in der Soziologie zumindest viel geringer war, als der WR vermutete. Es ist nicht bekannt, dass nach Vorstellung der Ergebnisse eine Diskussion über schlecht oder nur befriedigend evaluierte Einheiten und Einrichtungen stattgefunden hätte, geschweige denn, dass Universitäten oder Ministerien Konsequenzen gezogen hätten. Immerhin 61 Einheiten hatten in fünf Jahren keinen einzigen Artikel in einer referierten Zeitschrift veröffentlicht, fast ebenso viele nicht ein einziges wissenschaftsgesteuertes Drittmittelprojekt eingeworben.

Ein Grund für das letztlich geringe Interesse könnte die Orientierung auf Forschungseinheiten (in den meisten Fällen waren das einzelne Professuren) sein. Das Forschungsrating in der Soziologie zeigte insbesondere das auf, was man auch ohne Forschungsrating wusste. Die Forschungsleistung dieser meist kleinen Einheiten lässt sich nämlich ziemlich genau durch bibliometrische Verfahren bestimmen. Neben der blossen grössengewichteten Anzahl von Veröffentlichungen könnten auf dieser Ebene auch die Zitationen ermittelt werden. Davon wurde in der Pilotstudie abgesehen – aus damals nachvollziehbaren Gründen. Allerdings wäre die Forschungsleistung durch eine gemeinsame Bewertung von Publikationsoutput und dessen Einfluss adäquater zu beschreiben. Die eigenen und knapp berichteten Auswertungen mit den Daten des Forschungsratings Soziologie unterstützen diese Sichtweise.

Sinnvoller und informativer erschiene eine Orientierung eines Forschungsratings an den Forschungseinrichtungen. Auch die Vorgaben und Ziele der Evaluation wären dann zu ergänzen. Auf der Ebene der Einrichtungen könnte auf Grundlage zunächst standardisierter Daten eine informierte Diskussion unter Peers zur Forschungsausrichtung, einer erkennbaren Profilierung, zu genutzten bzw. ungenutzten Synergien und zur Einbindung in internationale Forschungszusam-

menhänge erfolgen. All dies macht nur auf der Ebene von Forschungseinrichtungen Sinn. Die Peers wären in diesem Fall effizienter eingesetzt, als wenn sie die Veröffentlichungen von Kollegen (im Fall von Peer Review gestützten Zeitschriften: *nochmals*, im Fall mancher Sammelwerke vielleicht *erstmal*) lesen und bewerten. Auch die sonstigen Leistungsfaktoren – etwa die Einwerbung von wissenschaftsgesteuerten Drittmitteln – wurden in der Regel bereits vor dem Forschungsrating durch Peers evaluiert. Aber was bislang fehlt, ist eine Gesamtschau dazu, wie die einzelnen Forschungseinheiten zusammenpassen und wie zielführend und kreativ sie zusammenarbeiten. Dazu könnte der Sachverstand eines informierten Peer Review bei einer neuen Runde des Forschungsratings, zu der der WR bereit zu sein scheint, gewinnbringend genutzt werden.

Literatur

- Baier, Christian (2011). Wissenschaft regieren. Eine diskursanalytische Studie zum Forschungsrating des Wissenschaftsrates. *Soziale Welt* 62: 371–390.
- Bornmann, Lutz und Hans-Dieter Daniel (2003). Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. In: Schwarz, Stephanie/Teichler, Ulrich (Hrsg.): *Universitäten auf dem Prüfstand*. Frankfurt: Campus, 207–225.
- Neidhardt, Friedhelm (2006). Forschungsrating der deutschen Soziologie durch den Wissenschaftsrat. *Soziologie* (35): 303–308.
- Neidhardt, Friedhelm (2008). Das Forschungsrating des Wissenschaftsrates. Einige Erfahrungen und Befunde. *Soziologie* (37): 421–432.
- Neidhardt, Friedhelm (2009). Stärken und Schwächen der Soziologie in Deutschland. *Soziologie* (38): 40–48.
- Münch, Richard und Christian Baier (2009). Die Konstruktion der soziologischen Realität

- durch Forschungsrating. *Berliner Journal für Soziologie* (19): 295–319.
- Münch, Richard (2009). Die Konstruktion soziologischer Exzellenz durch Forschungsrating. *Soziale Welt* (60): 63–90.
- Riordan, Patrick, Christian Ganser und Tobias Wolbring (2011). Zur Messung von Forschungsqualität. Eine kritische Analyse des Forschungsratings des Wissenschaftsrats. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (63): 147–172.
- Steuerungsgruppe (2008). Forschungsleistungen deutscher Universitäten und ausseruniversitärer Einrichtungen in der Soziologie [Steuerungsgruppe der Pilotstudie Forschungsrating im Auftrag des Wissenschaftsrates]. Abgerufen: http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/Pilotstudie_Forschungsrating_Soziologie/pilot_ergeb_sozio.pdf.
- Wissenschaftsrat (2013). Empfehlungen zur Zukunft des Forschungsratings. Drs 3409-13. Abgerufen: <http://www.wissenschaftsrat.de/download/archiv/3409-13.pdf>.

Evaluation 2.0: von der wissenschaftlichen zur managerialen Qualitätssicherung?

Richard Münch (Universität Bamberg)

Unisono hören wir, dass die Wissenschaft in Zeiten knapper öffentlicher Finanzen nicht mehr mit ausreichender finanzieller Unterstützung rechnen könne, ohne regelmässig Rechenschaft über ihr Tun abzulegen. Für diesen Zweck sei es notwendig, sich kontinuierlich einer Evaluation zu stellen, die der Öffentlichkeit vermittelt, was in Forschung und Lehre geleistet wird. Und man ist dann schnell dabei, zu argumentieren, es sei besser, die Evaluation selbst in die Hand zu nehmen und die geeigneten wissenschaftsadäquaten Kriterien zu verwenden, als es irgendwelchen wissenschaftsfernen Instanzen zu überlassen, die beanspruchen, mit Rankings für die geforderte «Transparenz» der Wissenschaft für die Öffentlichkeit zu sorgen. Ich möchte hier zeigen, dass diese sehr gängige Argumentation verkennt, wie problematisch die Errichtung eines derartigen Evaluationsregimes für das Gedeihen der Wissenschaft selbst ist und wie wenig sie erkennt, dass Finanzknappheit und die Forderung nach mehr Transparenz nicht unumstössliche Tatsachen sind, sondern ideologisch geprägte soziale Konstruktionen darstellen, die nicht sakrosankt, sondern höchst bestreitbar sind. Im Wesentlichen geht es darum, dass das neue Evaluationsregime der primären wissenschaftlichen Qualitätssicherung eine von wissenschaftsfremden Interessen gesteuerte sekundäre Qualitätssicherung überstülpt, sodass die primäre Qualitätssicherung nicht mehr adäquat funktioniert. «Evaluation 2.0» ist deshalb, anders als es derartige Ankündigungen üblicherweise beanspruchen, kein Fortschritt, sondern ein Rückschritt.

Primäre wissenschaftliche Qualitätssicherung

Selbstverständlich ist es einem gut sozialisierten Wissenschaftler oder einer gut sozialisierten Wis-

senschaftlerin nicht gleichgültig, ob seine bzw. ihre publizierten Forschungsergebnisse der kritischen Prüfung standhalten und bei den Fachkollegen bzw. Fachkolleginnen Beachtung finden. Das Streben nach Anerkennung für die erbrachten Leistungen ist ein wesentlicher Bestandteil des wissenschaftlichen Habitus, der im akademischen Sozialisationsprozess erworben wird. Ebenso selbstverständlich ist es für gut sozialisierte Wissenschaftler oder Wissenschaftlerinnen, ihre Forschungsprodukte auf den Prüfstand der Kritik zu stellen, um zu sehen, ob man damit weiterarbeiten kann oder besser andere Wege einschlägt. Wir lernen das von der ersten Seminararbeit bis zum Habilitationsverfahren. Und damit ist es längst nicht genug. Dass alles, was wir von uns geben, der Kritik unterworfen wird und nur das überleben kann, was der Kritik standhält, ist wissenschaftlicher Alltag. «Lasst Hypothesen sterben, nicht Menschen», hat Karl Popper (1958) einmal sein Credo des Falsifikationismus auf den Punkt gebracht. Nach diesem Credo sind viele Irrtümer Wegbereiter der nie zum Ziel gelangenden Annäherung an die Wahrheit. In der Wissenschaft ist es deshalb nichts Ehrenrühriges, Irrtümer zu begehen. Auf viele Fehlschläge kommen nur wenige Fortschritte der Erkenntnis.

Die Evaluation ist demnach jeder wissenschaftlichen Forschung inhärent. Diese ist ohne Evaluation gar nicht möglich. Es ist jedoch ein wesentlicher Bestandteil der wissenschaftsimmanenten Evaluation, dass sie nur Hypothesen und nicht Menschen sterben lässt, weil es eine gemeinsam geteilte Überzeugung aller Wissenschaftler und Wissenschaftlerinnen ist, dass die Forschung hohe Risiken des Scheiterns eingehen muss, wenn Fortschritte der Erkenntnis erzielt werden sollen. Es besteht auch ein Einverständnis darüber, dass aus

Fehlschlägen viel gelernt werden kann. Fehlschläge führen nicht gleich zum Verlust der Anerkennung. Dazu gehört auch, dass Forschungsergebnisse als Kollektivgut der wissenschaftlichen Gemeinschaft verstanden werden, als ein Gut, das nur in Zusammenarbeit hervorgebracht werden kann und das allen zu gleichen Teilen zur Verfügung steht (Callon 1993). Jede Forschungsarbeit beginnt mit dem Stand der Forschung auf einem Gebiet, für dessen Nutzung nichts bezahlt werden muss. Alles, was auf diesem Gebiet von anderen WissenschaftlerInnen erarbeitet wurde, inklusive aller Fehlschläge, ist eine entscheidende Ressource für jedes aktuelle Forschungsprojekt, das schon zu Beginn auf einer imaginären Kooperation mit vielen anderen WissenschaftlerInnen beruht und im weiteren Verlauf nur im Austausch mit realen und imaginären KooperationspartnerInnen durchgeführt werden kann. Jede Publikation von Forschungsergebnissen aus dem Projekt geht in den kollektiven Wissensbestand ein und bildet ihrerseits eine Ressource für weitere Forschungsarbeiten, unabhängig davon, ob sich die Ergebnisse halten lassen oder nicht.

Natürlich leisten einzelne WissenschaftlerInnen einen mehr oder weniger grossen Beitrag zum Erkenntnisfortschritt, und zwar in ganz unterschiedlicher Form. Sie mögen viel publizieren und für jede publizierte Einheit kontinuierliche aber nicht übermässig grosse Resonanz erzeugen, z. B. gemessen in Zitaten pro Einheit, oder für wenige Einheiten viel Resonanz erfahren, für viele andere dafür wenig. Sie mögen weniger publizieren und damit viel Aufmerksamkeit erzeugen. Sie mögen wenig publizieren und dafür auch wenig Beachtung finden. Für das Kollektivgut Erkenntnisfortschritt ist es völlig gleichgültig, wie sich Publikationen und Zitationen verteilen. Auch der am wenigsten aktive und beachtete Wissenschaftler ist Teil des kollektiven Produktionsprozesses und selbst wenn er nur die Schar derer vergrössert, die den höchst erfolgreichen vielzitierten WissenschaftlerInnen die Ehre erweisen und sich freuen, zu der Gemeinschaft derer zu gehören, die als Ganze solche

Erfolge hervorgebracht hat, wer immer es war, der oder die auf den Schultern der Gemeinschaft am Ende einen speziellen Durchbruch erzielt hat. Es gibt ja die berühmte 80/20-Verteilung von Publikationen und Zitationen, das heisst 80 Prozent aller Publikationen werden von etwa 20 Prozent aller Wissenschaftler hervorgebracht und 80 Prozent aller Zitate beziehen sich auf etwa 20 Prozent aller Publikationen. Wer nicht begreift, dass die Produktion von Erkenntnis ein kollektiver Prozess ist, kommt angesichts dieser Verteilung leicht zu dem Schluss, dass man auf 80 Prozent aller Wissenschaftler verzichten könnte. Stephen Cole hat z. B. diese Hypothese aufgestellt, hat dafür aber in einer Studie zur Physik im Zeitraum von 1963 bis 1976 keine Bestätigung gefunden. Die Zahl von mindestens einmal zitierten Physikern variierte in diesem Zeitraum mit der Zahl der beschäftigten Physiker (Cole 1992, 220). Es war also keineswegs so, dass zusätzlich eingestellte Physiker überflüssig waren und an der Zahl der beachteten Physiker nichts geändert haben. Cole hat dabei ein Forschungsdesign gewählt, das dem kollektiven Charakter der Erkenntnisproduktion noch nicht einmal gerecht wird, weil er diese ausschliesslich den mindestens einmal zitierten Physikern zurechnet. Dabei übersieht er, dass auch die nicht zitierten Physiker an der kollektiven Erkenntnisproduktion teilhaben, und selbst wenn sie nur zeigen, welche Wege nicht zum Erfolg führen oder als Unterstützer der Erfolgreichen wirken, z. B. indem sie deren Beiträge an Studierende vermitteln, die sie wiederum mehr oder weniger erfolgreich in die weitere kollektive Erkenntnisproduktion einspeisen.

Für die Erklärung der 80/20-Verteilung bieten sich drei strukturelle Ursachen an: Erstens ist der Spielraum für Neues, das Aufmerksamkeit auf sich zieht, begrenzt und Fortschritte können im Verhältnis zum Bestand an Wissen nur in begrenztem Umfang erzielt werden. Schon von daher liegt es nahe, dass sich etwa 80 Prozent der Forschung im Bereich des vorhandenen Wissensbestandes bewegt, im Sinne von Thomas Kuhn (1967) Nor-

malwissenschaft betreibt. Zweitens arbeitet wohl die grösste Zahl der Wissenschaftlerinnen und Wissenschaftler gar nicht an der Forschungsfront, sondern ist mit dem Transfer des bestehenden Wissens in die Praxis, mit der Unterrichtung von Studierenden oder mit der Vermittlung des Wissens an die Öffentlichkeit beschäftigt. Es ist nicht zu erwarten, dass daraus Publikationen hervorgehen, die im Web of Science registriert und zitiert werden. Es ist deshalb paradox, dass Evaluatoren von Instituten der angewandten Forschung den dort tätigen WissenschaftlerInnen im Zuge der Einrichtung regelmässiger Evaluationen vorgehalten haben, zu wenig in den Fachzeitschriften publiziert zu haben und dort zu wenig zitiert werden. Das ist schlichtweg nicht ihre Aufgabe, sodass die Evaluatoren offensichtlich ihre Aufgabe nicht richtig verstanden haben. Drittens dürfte die Aufmerksamkeitsökonomie eine wesentliche Rolle bei der Erklärung der 80/20-Verteilung bei den Zitationen spielen. 10 000 WissenschaftlerInnen, zum Beispiel, haben schlichtweg nicht mehr Zeit, um mehr als 2 000 unter ihnen überhaupt Aufmerksamkeit zukommen zu lassen, ihre Publikationen zu lesen und zu zitieren. Grössere, über nur wenige Zitate hinausreichende Aufmerksamkeit kann aus zeitökonomischen Gründen wiederum nur einem Bruchteil von den 2 000 geschenkt werden, nach der 80/20-Verteilung nur 400 und davon kann wiederum nur 80 höchste Aufmerksamkeit zukommen. Es ist deshalb schon nach der Aufmerksamkeitsökonomie vollkommen deplatziert, die Zitationsquoten der Erfolgreichsten zum Massstab für alle zu machen. Die besonders Erfolgreichen ragen ja gerade deshalb heraus, weil die vielen anderen grundsätzlich nur sehr begrenzte oder in den meisten Fällen überhaupt keine Aufmerksamkeit auf sich ziehen können. Die besonders Erfolgreichen gibt es überhaupt nur im Kontrast zu den vielen weniger Erfolgreichen.

Für die wissenschaftsimmanente Qualitätssicherung ist die ungleiche Verteilung von Erfolgen kein Problem. Sie ist ausschliesslich auf Hypothe-

sen und nicht auf Personen gerichtet. Aus der Perspektive der Förderung des Erkenntnisfortschritts ist es völlig gleichgültig, an welchem Ort, zu welcher Zeit von welchen Personen Erkenntnisse produziert werden, die der Kritik standhalten und Anstösse für weitere Erkenntnisproduktion liefern, solange wissenschaftliches Wissen ein Kollektivgut ist. Es kann in Cambridge, Massachusetts, in London, Honolulu, Sao Paulo, Shanghai, Lille, Greifswald oder Bologna sein. Da Diversität eine unabdingbare Quelle für Erkenntnisfortschritt ist, bietet eine breite Streuung von Orten und Personen bessere Bedingungen als die Konzentration auf wenige Orte und Personen. Je grösser die Zahl unterschiedlicher Quellen, umso grösser die Chancen, dass Neues entsteht. Dass Diversität nicht in Beliebigkeit umschlägt, dafür sorgt die Kritik als genuin wissenschaftliches Instrument der Qualitätssicherung. Es gibt kein anderes Feld, auf dem die Kritik als Instrument der Qualitätssicherung eine so grosse Rolle spielt. Im Sozialisationsprozess wird das Kritikeinstecken und das Kritiküben so ausgiebig gepflegt, dass diese Prozedur zum festen Bestandteil des wissenschaftlichen Habitus wird. Der gut sozialisierte Wissenschaftler hat gelernt, mögliche Kritik schon selbst zu antizipieren und verfährt selbst nach der Methode, Hypothesen aufzustellen und zu prüfen, Argument und Gegenargument einander entgegenzustellen. Bevor er mit neuen Erkenntnissen an die Öffentlichkeit tritt, hat er schon den inneren Diskurs geführt. In vielen Fällen kann nur geringe Produktivität darauf zurückgeführt werden, dass WissenschaftlerInnen ein so kritisches Überich ausgebildet haben, dass sie jede Erkenntnis schon verwerfen, bevor sie an die wissenschaftliche Öffentlichkeit gelangen, während andere über ein stärkeres Ich verfügen und das kritische Überich nur begrenzt zur Wirkung kommen lassen. Auch das kann vom übergeordneten Gesichtspunkt des Erkenntnisfortschritts aus betrachtet eine sinnvolle Arbeitsteilung sein.

Ein grosser Spielraum für Diversität ist die Quelle, aus der sich Erkenntnisfortschritt speist,

während institutionalisierte Kritik für die Ordnung des Wissens sorgt. Sie klopf alle Beiträge zum wissenschaftlichen Erkenntnisfortschritt auf ihre logische Stringenz, theoretische Fundierung, methodische Sorgfalt und empirische Haltbarkeit ab. Das geschieht im permanenten Prozess des Peer Review. Promotionen, Habilitationen, Beförderungen, Berufungen, Forschungsanträge, eingereichte Aufsätze bei Fachzeitschriften durchlaufen alle diesen Prozess der kritischen Prüfung. Wir können deshalb annehmen, dass die Wissenschaft eher Gefahr läuft, zarte Pflanzen, aus denen neue Erkenntnisse erwachsen können, durch exzessive Kritik zu zertrampeln, als dass der Kreativität zu viel Spielraum gegeben würde. Auf diese Weise wird der Erkenntnisfortschritt eher gehemmt als befördert. Genau aus diesem Grund hat Paul Feyerabend (1970) sein Plädoyer für methodischen Pluralismus dem popperschen Falsifikationismus entgegengestellt. Und Imre Lakatos (1970) hat aus dieser Kontroverse den Schluss gezogen, Forschungsprogrammen Zeit zum Aufblühen zu lassen, bevor sie von der Kritik zermalmt werden. Zu diesem Zweck sollen sie einen stabilen Kern aufbauen, um den ein Schutzgürtel von veränderbaren Annahmen gezogen wird.

Sekundäre wissenschaftsexterne Qualitätssicherung

Auf jeden Fall können wir sagen, dass der Spielraum für Diversität, Abweichung von herrschenden Standards und Kreativität auf der einen Seite und die Qualitätssicherung durch Peer Review auf der anderen Seite in einem ausgewogenen Verhältnis zueinander stehen müssen, damit gesicherter Erkenntnisfortschritt möglich ist. Genau dieses Verhältnis wird nun im Zuge der Qualitätssicherungsoffensive von New Public Management in der Wissenschaft nach der Seite der Qualitätsprüfung verschoben, die ein grösseres Gewicht bekommt als jemals zuvor. Dabei werden vier entscheidende Transformationen vollzogen: (1) Die manageriale Qualitätssicherung richtet sich nicht mehr auf die

Erkenntnisproduktion als kollektiver Prozess der Herstellung und Aneignung eines Kollektivgutes, sondern auf die Erkenntnisproduktion von einzelnen WissenschaftlerInnen und einzelnen Institutionen. (2) Gegenstand der kritischen Prüfung sind nicht Hypothesen, sondern Menschen bzw. Institute. (3) Die neue manageriale Qualitätssicherung wird dem primären wissenschaftsimmanenten Prozess der Qualitätssicherung als sekundärer Prozess übergestülpt. (4) Bezugspunkt der managerialen Qualitätssicherung ist nicht der wissenschaftliche Erkenntnisfortschritt, sondern die Positionierung von Individuen und Institutionen im Wettbewerb um Ressourcen und Prestige.

Die Protagonisten der Verbreitung von managerialer Qualitätssicherung in der Wissenschaft argumentieren gern, dass sie nur für etwas eintreten, was immer schon Teil der wissenschaftlichen Praxis ist und was nun lediglich systematisiert und standardisiert und damit objektiviert und schliesslich zum Vorteil aller transparent gemacht wird (Küpper 2009). So wird versucht, der sekundären managerialen Qualitätssicherung die Weihe der primären wissenschaftlichen Qualitätssicherung zu geben, um einen legitimen Status in der akademischen Welt zu erhalten. Dieser Versuch verkennt den grossen Unterschied zwischen primärer und sekundärer Qualitätssicherung. Die primäre Qualitätssicherung regelt den genuin wissenschaftlichen Wettbewerb um Erkenntnisfortschritt, die sekundäre Qualitätssicherung ist ein manageriales Instrument zur Steigerung der Wettbewerbsfähigkeit von Organisationen, die im ökonomischen Wettbewerb um Ressourcen und Prestige stehen. Dieser ökonomische Wettbewerb zielt auf Monopolrenten, mit deren Hilfe sich als Unternehmen verstandene Universitäten dauerhaft Wettbewerbsvorteile in der zirkulären Akkumulation von Ressourcen und Prestige sichern.

Dieser ökonomische Wettbewerb hat nach aussen und nach innen fatale Folgen für den Erkenntnisfortschritt. Nach aussen führt die wachsende Ungleichheit in der Verfügung über Ressourcen

und Prestige dazu, dass sich die beachtete Wissensproduktion auf wenige Zentren konzentriert und die breite Peripherie als Quelle des Erkenntnisfortschritts ausscheidet. Es bildet sich eine oligopolartige Struktur der wissenschaftlichen Forschung mit einem massiven Verlust an Diversität heraus. Nach innen können sich die universitären Unternehmen – ganz anders als die Wissenschaft als kollektiver Prozess – keinen Flop leisten, weil das einen Ressourcen- und Prestigeverlust mit sich bringt. Deshalb benötigen sie ein durchgreifendes strategisches und operatives Management. Sie müssen ihr Personal auf Linie bringen und in Clustern organisieren, um bei der Einwerbung von Drittmitteln in grossem Stil zu reüssieren. Sie müssen ihre Angestellten zur Verbundforschung zwingen, um dabei zum Erfolg zu gelangen. Und sie müssen jeden einzelnen Wissenschaftler und jede einzelne Wissenschaftlerin vom Anfang bis zum Ende der Karriere einer detaillierten Kontrolle unterwerfen.

Um das leisten zu können, benötigen universitäre Unternehmen Zahlen. Nur so können sie nach aussen Erfolge dokumentieren und nach innen an Stellschrauben drehen, um Ressourcen und Prestige zu akkumulieren. Diese Zahlen werden ihnen von Rankings bereitgestellt. Die von diesen Rankings verwendeten Indizes können sie für das interne Monitoring ihrer «Geschäftsprozesse» verwenden. Indikatoren für die Drittmittelinwerbung, Publikationen und Zitationen beherrschen deshalb zunehmend die Praxis des internen Qualitätsmanagements. Was dabei geschieht, ist die zunehmende Verengung der wissenschaftlichen Praxis auf das, was die eingesetzten Kennzahlen messen. Die Folge ist eine massive Schrumpfung des Spielraums für Kreativität. Die heranwachsenden WissenschaftlerInnen werden auf Konformität mit den herrschenden Standards getrimmt, weil nur auf diese Weise das Peer Review-Verfahren bei der Einreichung von Forschungsanträgen bei Förderinstitutionen und von Aufsätzen bei Fachzeitschriften überstanden wird. Ein Spielraum für Eigensinn bleibt nur noch einer ganz kleinen

ausserwählten Schar von Wissenschaftlern an den elitären Standorten, während die breite Masse an der kurzen Leine der permanenten Kontrolle geführt wird. Die äussere Stratifikation in Elite und Masse spiegelt sich in der Abspaltung eines elitären Habitus ganz weniger Stars von einem konformistischen Habitus in der breiten Mittelschicht, während es sich die akademische Unterschicht der periphersten Institutionen wiederum leisten kann, ihre eigenen nonkonformistischen Wege zu gehen. Besonders kritische, heterodoxe Positionen sind dann am ehesten an solchen kaum sichtbaren Orten zu finden.

Die Folgen dieser Überwucherung der genuin wissenschaftlichen Praxis im Spannungsfeld von Diversität und kritischer Prüfung durch das strategische und operative Management universitärer Unternehmen für die Wissenschaft sind verringerte Spielräume für Diversität und Kreativität und damit eine massive Schrumpfung des Potenzials, neue Wege der Erkenntnis gehen zu können. Vor dieser Gefahr wird von gewichtigen wissenschaftlichen Vereinigungen mit allem Nachdruck gewarnt. Namhafte Vereinigungen kritisieren den unreflektierten Gebrauch quantitativer Publikations- und Zitationsindizes wie dem Journal Impact Factor zur Messung des Einflusses von Fachzeitschriften und dem Hirsch-Index (Hirsch 2005) zur Messung des Einflusses eines Wissenschaftlers bei der Entscheidung über Beförderungen, Förderanträge und Berufungen (vgl. Bornmann und Daniel 2008). Die International Mathematical Union hat dazu einen vielbeachteten kritischen Bericht verfasst (Adler et al. 2009). Auf der Tagung der American Society for Cell Biology wurde im Dezember 2012 dazu die San Francisco Declaration on Research Assessment verabschiedet, der sich eine Vielzahl von Institutionen und WissenschaftlerInnen angeschlossen haben: 155 individuelle und 82 institutionelle Erstunterzeichner, inzwischen weitere 12055 individuelle und 547 institutionelle Unterzeichner (DORA 2012).

Expansion trotz massiver Kritik: Die Überlagerung der primären durch die sekundäre Qualitätssicherung als Gegenstand soziologischer Erklärung

Trotz massiver Kritik ist die Praxis der managerialen Qualitätssicherung mittels Kennzahlen nicht im Rücklauf, sondern ganz im Gegenteil auf dem Weg der weiteren Expansion. Diese Paradoxie ruft nach einer soziologischen Erklärung, die ich hier aus Platzgründen nur andeuten kann. Die Funktionalität dieser Evaluationspraxis für die Wissenschaft kommt ganz offensichtlich nicht in Frage, jedoch ihre Funktionalität für die als Unternehmen begriffene Universität und ihre Investoren, zu denen Regierungen, die Drittmittelgeber, Sponsoren und Studierende zu zählen sind. Es handelt sich also um eine Art ökonomischer Invasion in das akademische Feld, in dem der heteronome Pol des Wettbewerbs zwischen Universitätsunternehmen um die zirkuläre Akkumulation von Ressourcen und Prestige den wissenschaftlichen Wettbewerb um Erkenntnisfortschritt und Anerkennung am autonomen Pol überlagert (Münch 2014). Diese ökonomische Invasion speist sich wiederum aus der globalen Verbreitung der neoliberalen Lehre, die öffentliche Güter sowie ihre Herstellung und Nutzung denselben Produktions- und Konsumtionsbedingungen wie private Güter unterwerfen möchte und damit ein Programm der Gleichschaltung aller Funktionsbereiche der Gesellschaft mit der Ökonomie verfolgt. In dem Masse, in dem diese Lehre eine hegemoniale Stellung einnimmt, können die kontraproduktiven Effekte der ökonomischen Gleichschaltung auf die Wissenschaft nicht mehr erkannt werden, sodass sich der entsprechende Strukturwandel der Wissenschaft ohne nennenswerten Widerstand vollziehen kann. Die letzte Bastion der Wissenschaft sind die wissenschaftlichen Fachgesellschaften. Diese geraten in diesem Prozess in einen zunehmend verschärften Konflikt mit den universitären Unternehmen. Sie sind der Machtfaktor der Wissenschaft, auf den es bei der Abwehr der ökonomischen Invasion in das

Feld der Wissenschaft ankommt. Das neue Regime der sekundären Qualitätssicherung bezieht seine Legitimation letztlich aus der Verdrängung des Modells der professionellen Treuhänderschaft der Wissenschaft, das auf Vertrauen in die professionelle Selbstregulation setzt, durch das institutionenökonomische Prinzipal-Agenten-Modell, das auf dem grundsätzlichen Misstrauen aufbaut, dass der Agent in die eigene Tasche wirtschaftet und deshalb unter Kontrolle des Prinzipals gehalten werden muss. Erst die moderne Informationstechnologie liefert das Instrument, um daraus ein Panoptikum der totalen Kontrolle zu machen (Foucault 1977). Die Expansion der sekundären Qualitätssicherung trotz heftiger Kritik verdankt sich dem Machtverlust der wissenschaftlichen Fachgesellschaften im Verhältnis zu den als Unternehmen geführten Universitäten und der globalen Verbreitung einer Misstrauenskultur im Zuge der Hegemonie des Neoliberalismus im öffentlichen Diskurs und der Ökonomie im Diskurs der Sozialwissenschaften.

Literatur

- Adler, Robert, John Ewing und Peter Taylor (2009). "Citation Statistics. A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)". *Statistical Science* 24(1): 1–14.
- Bornmann, Lutz und Hans Dieter Daniel (2008). "What do citation counts measure? A review of studies on citing behavior". *Journal of Documentation* 64(1): 45–80.
- Callon, Michel (1993). "Is Science a Public Good?" *Science, Technology & Human Values* 19(4): 395–424.
- Cole, Stephen (1992). *Making Science. Between Nature and Society*. Cambridge, MA: Harvard University Press.
- DORA (2012) San Francisco Declaration on Research Assessment. <am.asb.org/dora/>. Abgerufen am 31.12.2014.

- Feyerabend, Paul K. (1970). *Against Method*. London: Verso.
- Foucault, Michel (1977). *Überwachen und Strafen. Die Geburt des Gefängnisses*. Frankfurt a. M.: Suhrkamp.
- Hirsch, Jorge E. (2005). "An Index to Quantify an Individual's Scientific Research Output". *Proceeding of the National Academy of Sciences of the United States* 102(46): 16569–16572.
- Kuhn, Thomas (1967). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Küpper, Hans-Ulrich (2009). «Evaluation von Forschung und Lehre». In: Max-Emanuel Geis (Hrsg.). *Hochschulrecht im Freistaat Bayern. Handbuch für Wissenschaft und Praxis*. Heidelberg: Springer: 201–215.
- Lakatos, Imre (1970). "Falsification and the Methodology of Scientific Research Programmes". In: Imre Lakatos & Alan Musgrave (Eds.). *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press: 91–196.
- Münch, Richard (2014). *Academic Capitalism. Universities in the Global Struggle for Excellence*. London und New York: Routledge.
- Popper, Karl R. (1958). *Die offene Gesellschaft und ihre Feinde*. Bern: Francke.

Academic Rankings: Are They Really “Here to Stay”?

Stephan Lessenich (Ludwig-Maximilians-Universität München)*

Managers, administrators, policy makers, journalists and the public at large all like the simple numerical ordering of any sort of phenomena, from the Forbes’ richest people to Billboard’s Top 100 music charts, because it is readily accessible. Thus, it only was a question of time until university rankings entered the scene of numerology, and it comes as no surprise that they were met with both a sense of relief and greed by those who primarily use them. Decisions about the fate of something are made easier, and can be more easily (albeit superficially) justified, when this “something” can be expressed in numbers and ranked from (supposedly) best to worst.

The idea of university rankings dates back to the 1880s, in the form of classifications of universities in the United States. Yet, its modern form was actually born only in 2003, with the release of the Shanghai league table (now known as the Academic Ranking of World Universities). When launched a decade ago, academic rankings were immediately embraced by journalists, governments, policy makers and university administrators, and they attracted the strong interest of faculty and students (and their parents) as well. At the same time, however, university managers and scientists often fear them – rankings are on “a thin line between love and hate” (Salmi and Saroyan 2007, 10).

Obsession with rankings grew rapidly and was soon globalized, not least because rankings also support a huge business industry. Their impact is constantly increasing and nowadays there are more than a dozen institutions producing Global University Rankings. Within less than a decade, rankings have become important instruments for various aspects of higher education: university

reform, department curricula, faculty recruitment, promotion and wages, research funding, student admissions and tuition fees, a student’s future job prospects. As a result, they are being heavily advertised and covered by the media as soon as they are published. Their publication is intensively accompanied by press releases and public gloating from universities ranked at the top of the lists. Not least, they trigger reactions – pride and exultation, flurry and irritation – at different governmental and administrative levels.

Academics – including scientists, sociologists and even sociological theorists – are humans, and as humans they like numbers too. However, academics are pretty strange human beings: they are used to debate, comment, criticize, evaluate, reject and eventually propose alternatives to whatever becomes orthodoxy. In fact, it is these characteristic traits of scientists that lay at the very heart of scientific progress. In addition, most of them certainly know how to read numbers better than managers, administrators, politicians or journalists, and they are aware of the dangers of reducing value to what can be counted numerically. Finally, they are especially trained in detecting the assumptions behind those numbers, and in identifying patterns and propensities in them.

Thus, it is not really surprising that academics received rankings with great skepticism, questioning both their estimation and their performance. The critical literature rapidly increased in the years following their emergence. University rankings have been heavily criticized for (1) many methodological issues related to the indicators used and their weightings, (2) English speaking countries dominating the rankings, (3) teaching quality being hard, if at all, to measure, and (4) arts, humanities and social sciences being relatively under-represented. As Usher and Savino (2007, 13)

* Thanks go to Kostas Stergiou who contributed important ideas to this paper.

aply state: “In fact, most indicators are probably epiphenomena of an underlying feature that is not being measured.” And what is even more important from a sociological perspective, rankings have also been questioned for their “symbolically violent character (...) as a form of social categorization and hierarchization” (Amsler 2013, 155).

University rankings, global or not, have led to several boycotts throughout the world. Thus, after the publication of the 1997 and 1998 rankings of universities in the Asian and Pacific region, 35 universities refused to participate in the 1999 survey and as a result the initiative was terminated. Patricia McGuire, the president of Trinity University (Washington DC), boycotted U. S. News & World Report rankings: “Rip it up and throw it away. That’s the advice I’m giving my fellow college and university presidents this month as the ‘reputation survey’ from U. S. News & World Report lands on our desks. I am one of 12 presidents who wrote a letter urging colleagues to take a stand for greater integrity in college rankings – starting by boycotting the magazine’s equivalent of the ‘American Idol’ voting process.” (McGuire 2007) European education politics, often slavishly focused on what is (if only allegedly) going on in US academic life, should be more attentive to such incidents and processes as well.

Probably the most recent rejection of rankings is evident from the boycott of the German Centre for Higher Education Development (Centrum für Hochschulentwicklung – CHE) rankings by German sociologists. The CHE ranking has been very effective during the last decade in contributing to the political construction of a landscape of “good” and “bad” departments at German universities. It has done so by suggesting to be able to measure the relative quality of academic teaching by way of ranking the subjective satisfaction scores of a small sample of students (frequently not more than 10% of the totality) in different disciplines. However, rather than being a reliable instrument in advising students of which university department to go to

if they want to get a good education, the CHE ranking has proved to be welcomed by politics and bureaucrats as a seemingly self-evident measure of “excellence” and “non-excellence” in academic teaching. In a higher education system which, as in the German case, is ever more influenced by the power of numbers, teaching rankings are a further instance of constructing an academic “reality” of differences in quality. An artificial reality which eventually operates like a self-fulfilling prophecy: in the end, it may result in a cemented division of winners and losers, with departments with supposedly “excellent” teaching records attracting the more ambitious and mobile students.

The consequences of such boycotting of rankings might be either favorable or harmful to the individual institution(s) – ultimately, this is an empirical question. Many maintain that boycotting is not going to solve matters because “rankings are here to stay.” Yet, as Amsler (2013) rightly claims, ranking is neither a professionally necessary nor an institutionally inevitable activity. There are good – not only scientific, but also ethico-political – reasons to turn away from the ranking business. Actually, rankings are not “here to stay” if we do not want them to. As Peter Murray-Rust (Cambridge) stated to Zoë Corbyn (2009) – regarding journal metrics, yet equally applicable to rankings: “Higher education has to take control of academic metrics if it is to control its own destiny (...) it should determine what is a metric and what isn’t.” Yet, even if academics take control of metrics, the problem of measuring scientific quality remains. Simplified counting and ranking, even if organized by academics themselves, will still have serious limitations, and thus will not be the solution if the same type of power struggles and reputation games remain – and attention is restricted to what “counts” in numerical terms.

German sociologists have chosen another path. In cooperation with the academic association of German historians (Verband der Historiker und Historikerinnen Deutschlands – VHD),

the German Sociological Association (Deutsche Gesellschaft für Soziologie – DGS) has developed the information website Studium.Org/ (www.studium.org). This website presents extensive data on every single Bachelor program (and, starting in May 2015, also on Master programs) in sociology being offered by German universities. Users can search along a major set of quantitative and qualitative criteria for their preferred Sociology department – without being externally guided by a ranking function suggesting clear-cut differences between “strong” and “weak” university sites. This instrument is meant to create transparency instead of producing hierarchy. So far, the academic associations for communication and media sciences as well as for pedagogy have joined this initiative and will make available analogous information on their respective teaching programs by the end of this year.

Only time will tell if rankings actually “are here to stay.” But even if they should, there will always be a better alternative.

References

- Amsler, Sarah (2013). “University ranking: a dialogue on turning towards alternatives”. *Ethics in Science and Environmental Politics*, 13: 155–166.
- Corbyn, Zoë (2009). “A threat to scientific communication”. <http://www.timeshighereducation.co.uk/407705.article> (accessed 21 February 2015).
- McGuire, Patricia (2007). “Rank this, U. S. News”. <http://www.latimes.com/la-oe-mcguire14may14-story.html> (accessed 21 February 2015).
- Salmi, Jamil and Alenoush Saroyan (2007). “League Tables as Policy Instruments: Uses and Misuses”. *Higher Education Management and Policy* 19(2): 1–38.
- Usher, Alex and Massimo Savino (2007). “A Global Survey of University Ranking and League Tables”. *Higher Education in Europe* 32(1): 5–15.

L'évaluation en sociologie : un enjeu au carrefour des dynamiques du champ scientifique mondial

Frédéric Lebaron (Université de Versailles Saint-Quentin-en-Yvelines)

Si l'évaluation occupe une place centrale dans l'activité des sociologues, et si ceux-ci ne cessent d'en débattre de façon formelle ou informelle (voir notre contribution au dossier de *Global Dialogue* en 2014), l'enjeu de l'évaluation n'en a pas moins pris une importance nouvelle depuis quelques années. L'une des raisons évidentes de cette évolution est la conjonction de l'internationalisation de la recherche, avec, entre autres, l'hégémonie accrue de la langue anglaise dans le monde scientifique, et la montée en puissance de l'outil bibliométrique au sein des procédures évaluatives, notamment s'agissant des recrutements et des carrières. Ces évolutions concernent l'ensemble des disciplines académiques et scientifiques.

Elles suscitent en sociologie de fortes résistances, qui prennent des formes variées. Parmi celles-ci, une grande partie semblent relativement proches de celles rencontrées dans de nombreuses autres disciplines des sciences humaines et sociales, particulièrement les disciplines littéraires : les productions particulières de nos disciplines ne sauraient faire l'objet de procédures de standardisation de type industriel, relevant d'une économie des singularités mal traitée par le nouveau modèle d'organisation de la recherche.

Chaque production sociologique nécessite une évaluation interne spécifique, et les critères de celle-ci, multiples et contextuels, ne peuvent être codifiés. Les carrières des sociologues seraient encore plus difficiles à étalonner, du fait de leur caractère encore plus « multidimensionnel », liée à des parcours biographiques plus ou moins complexes, qu'il est pourtant toujours nécessaire de prendre en compte.

Face aux positions de refus, le champ scientifique continue cependant de voir s'affirmer chaque jour la logique de l'évaluation bibliométrique à

divers niveaux, cela en dépit de critiques internes tout autant qu'externes, comme menée par Yves Gingras, utilisateur de l'outil bibliométrique et très critique à l'égard des dérives de son usage.

Faut-il tenter de sortir de l'aporie qui s'est installée, entre le modèle standardisé d'un côté et la résistance au nom de la singularité d'autre part ? Ou peut-on faire évoluer le débat en proposant quelques lignes directrices ? C'est ce que nous essaierons de faire dans ce qui suit.

La quantification des activités des sociologues n'a pas attendu le mouvement de standardisation en cours. Le comptage des productions, des heures de travail, et en particulier de cours, la mesure de l'impact des ouvrages par leur diffusion, et bien d'autres formes de mises en nombre, sont présents, à des titres divers, dans le quotidien des sociologues depuis longtemps.

La nouveauté de notre période est l'apparition sur une large échelle de métrologies envahissantes, centrées principalement sur l'activité de production scientifique (l'enseignement et l'administration restant largement à l'écart de ce mouvement de quantification). Elles sont de plus en plus souvent institutionnalisées dans des listes de revues assorties de points, des calculs d'indices ou d'autres procédures formalisées.

L'une des forces de ces pratiques est qu'elles nourrissent la tendance à la création de classements et de palmarès qui peuvent être réalisés à plusieurs échelles : celle des individus, des départements et laboratoires, des universités et des pays. Dans ce système emboîté de mise-en-équivalence, la qualité de science est en théorie facilement évaluable, comparable et cumulable. Le « classement de Shanghai » en est la manifestation la plus visible.

Dans la réalité, l'obsession de la mesure rencontre de nombreux obstacles intellectuels, et la

pratique bibliométrique est beaucoup plus diverse qu'on ne le croit : on peut, schématiquement, décider *a priori* d'attribuer des points à des revues (comme c'est souvent le cas en sciences économiques, en tout cas en France) et calculer la production en conséquence ; on peut utiliser directement des mesures d'impact citationnel à partir de bases de données couvrant un échantillon de l'ensemble des productions (comme avec le *h*-index fondé sur Google Scholar). Dans tous les cas, les biais sont multiples, et les mesures toujours difficiles à interpréter sans une bonne connaissance des domaines scientifiques couverts, et des contextes nationaux de l'activité sociologique.

En particulier, la domination de la langue anglaise est trop souvent considérée comme allant de soi, dans les sciences humaines et sociales, alors qu'une large partie de la production, dans la plupart des pays non-anglophones, continue d'être effectuée dans la langue nationale et reste souvent absente (même si la numérisation fait évoluer cette situation) des classements internationaux pour cause d'invisibilisation. On peut penser que la production dans chaque langue devrait donner lieu, par exemple, à des analyses d'impact spécifiques. Si celles-ci révéleraient évidemment la position asymétrique de la langue anglaise (citée et beaucoup moins « citant »), cela permettrait aussi d'introduire une vision beaucoup plus riche et polycentrique des controverses sociologiques à l'échelle mondiale.

Une des voies qui semblent les plus intéressantes pour réguler la pratique de l'évaluation quantitative des productions consiste en effet à cantonner celle-ci à d'éventuels « éléments d'information » à la disposition des évaluateurs, mais ne pouvant pas conditionner directement la décision évaluative : il s'agit de refuser toute forme d'algorithmique qui ferait découler un classement et une décision d'un indice, sans la médiation d'une discussion collégiale.

Une deuxième voie, parallèle, consiste à pluraliser au maximum les outils de mesure, en distinguant clairement ce qui relève de la quantité produite (nombre d'articles, calculé selon diverses

méthodes pour tenir compte de la taille, du nombre d'auteurs, etc.), de l'impact (avec les citations, mais, là encore, calculées dans diverses bases de données, représentatives de sous-espaces et de langues différentes au sein du champ scientifique, et rapportées à divers indicateurs comme l'ancienneté, les conditions de carrière, etc.).

En affirmant la multidimensionnalité de l'activité de production scientifique, qui ne relève pas, tant s'en faut, de la seule production d'articles dans les revues pour les pairs, on rend possible une analyse plus solide des dynamiques de recherche, plus ou moins individuelles et collectives et associant des composantes diverses et hétérogènes.

À côté du renouvellement de l'évaluation des productions, les deux autres grands secteurs d'activité professionnelle des sociologues, l'enseignement et l'animation de la communauté, restent mal évalués car souvent peu définis et finalement faiblement constitués en objets légitimes pour la réflexion et l'action de la communauté scientifique. Là encore, on peut penser que l'explicitation de « grilles de critères » complexes et multidimensionnels permettrait au moins de dégager un socle commun de préoccupations. On sait bien sûr que chaque évaluateur a tendance à « pondérer » différemment l'investissement dans des fonctions de direction administrative, l'animation de revues, l'évaluation d'articles, etc. Mais on peut là encore souhaiter que chaque comité d'évaluation consacre à la discussion de ces critères un ensemble de débats préalables, en poussant la formalisation explicite (mais non la quantification) aussi loin que possible.

Car l'un des enjeux de l'évaluation est bien sûr la légitimité même des décisions prises sur la base des évaluations collégiales. Dans un contexte de tensions budgétaires et de restructuration autoritaire de l'enseignement supérieur et de la recherche, des procédures collégiales ouvertes et pluralistes, fondées sur des critères clairs et s'appuyant, sans en fétichiser aucun, sur un maximum d'éléments objectifs, devraient permettre de renforcer la confiance dans le bon fonctionnement de notre discipline.

Sociologie de l'évaluation et évaluation de la sociologie : quelques réflexions

Jean-Michel Bonvin (Université de Genève)

Comment évaluer la sociologie comme discipline scientifique et les départements de sociologie au sein des universités ? Quels critères utiliser ? Qui est habilité à mener une telle évaluation ? Autant de questions difficiles qui nécessitent de porter le regard sur l'activité d'évaluation à proprement parler en vue de poser quelques jalons pour une sociologie de l'évaluation. L'évaluation est une activité complexe qui vise à dire ce qui a de la valeur et, par là même, ce qui en a moins, voire pas du tout. Elle est porteuse d'enjeux normatifs de grande ampleur et demande qu'on dispose d'une théorie et d'une épistémologie adéquates. Ce court texte énonce quelques réflexions dans cette direction.

Une première question qu'il convient de se poser est celle des critères mobilisés dans l'évaluation. Classiquement, cette question porte sur le choix des bons critères, postulant ainsi qu'il y a des critères meilleurs que d'autres et qu'ils peuvent donc être appliqués légitimement en toutes situations. Suivant cette conception, on devrait dire que ce qui fait la qualité du travail d'un ouvrier ou d'un employé de service public pourrait être déterminé sans ambiguïté. Il suffirait dès lors d'identifier ces critères et, sur cette base, évaluer les activités déployées par les personnes concernées ; on aboutirait à des classements distinguant les meilleurs, désignés comme *best practices*, et incitant les autres à s'aligner sur leurs pratiques. Nombreuses sont les études sociologiques qui montrent la fragilité empirique d'une telle conception de l'évaluation : les théories de la justification (Boltanski, Thévenot), mais aussi des auteurs comme Paugam ou Zimmermann, mettent en évidence la pluralité des ordres de grandeur à partir desquels on peut évaluer la qualité du travail (ou d'une personne ou d'un bien, etc.). Celle-ci peut relever de considérations marchandes, en lien avec la rentabilité, mais aussi d'une logique industrielle

(focalisée sur l'efficacité technique), civique (le respect du savoir-faire traditionnel), etc. Surtout, ces auteurs insistent sur l'incommensurabilité de ces divers ordres de grandeur, en d'autres termes il est impossible de décréter la supériorité absolue de l'un d'eux sur les autres, chacun a sa validité et un mécanisme d'évaluation qui privilégierait un ordre de grandeur aux dépens des autres reviendrait à imposer indûment sa supériorité. Bref, la sociologie de l'évaluation fait le constat de la pluralité des critères légitimement mobilisables et de l'illégitimité d'en privilégier un au détriment des autres. La question, dès lors, n'est plus celle de la sélection des bons critères, mais de la mise en place de mécanismes d'évaluation qui laissent place à cette pluralité de critères. La question des critères d'évaluation (donc celle de la qualité d'une activité, d'un produit, d'un service, etc.) est indécidable dans l'absolu, elle ne peut être tranchée que par des compromis situés et provisoires entre les personnes concernées.

La sociologie, en tant qu'objet de l'évaluation cette fois-ci, n'échappe pas à cette règle. Diverses conceptions de la « bonne » sociologie ou du « bon » sociologue ou encore du « bon » département de sociologie coexistent, qui sont pour la plupart ancrées dans des positions étayées et légitimes. Dès lors, privilégier l'une de ces conceptions aux dépens des autres reviendrait à nier cette pluralité et à ne pas reconnaître à sa juste valeur l'activité scientifique déployée suivant des standards alternatifs. Dans l'évaluation des activités scientifiques, des tendances fréquemment observées consistent à valoriser les activités de recherche au détriment des enseignements, à donner la priorité aux publications dans les revues dont l'*impact factor* est élevé aux dépens des publications jugées moins prestigieuses. De plus, l'accent est souvent mis sur la quantité des activités réalisées au risque de nuire à leur qualité.

Ces tendances s'inscrivent dans la volonté d'identifier un critère d'évaluation unique (ou un ensemble très cohérent de critères d'évaluation), qui soit applicable à l'ensemble des personnes travaillant dans le champ. Il s'agit ici d'établir un classement dont la validité ne soit pas contestable et dont les effets sur la pratique consistent dans l'émulation des moins bien classés, en vue d'une amélioration globale de la qualité de l'activité scientifique au sein des universités. Le postulat sous-jacent consiste à dire que le « bon » sociologue est avant tout un chercheur qui publie autant que possible dans les revues les mieux cotées. C'est là une conception du « bon » sociologue dont la validité paraît peu contestable, mais ce n'est précisément qu'une des conceptions possibles du « bon » sociologue et mettre l'accent exclusivement sur celle-ci peut entraîner des phénomènes de dévalorisation des activités qui ne s'inscrivent pas dans la même logique.

Le philosophe et économiste indien, Amartya Sen, souligne au contraire l'impossibilité de parvenir à de tels classements suivant lesquels x serait, dans l'absolu, « meilleur » sociologue ou scientifique que y car il aurait publié dans une revue à *impact factor* plus élevé. Selon lui, précisément en raison de la pluralité indépassable des ordres de grandeur et des critères d'évaluation légitimes, il conviendrait de se contenter de classements partiels, indiquant que tel ou tel sociologue ou tel ou tel département de sociologie est « meilleur » sur tel ou tel critère ou relativement à telle ou telle dimension, et donc d'abandonner la volonté d'établir une hiérarchie claire et unique. Il ne s'agit pas ici de contester la qualité des travaux publiés dans les revues les plus prestigieuses et les mieux cotées, elle est indéniable. Notre argument ne porte pas sur la pertinence de ces critères d'évaluation, mais sur leur tendance à l'hégémonie qui aboutit à ne pas laisser une place suffisamment valorisée à d'autres visions de ce qu'est la « bonne » sociologie. Si l'on suit Amartya Sen, la solution ne consiste donc pas à remplacer ces critères par d'autres, qui comporteraient le même risque de non-reconnaissance de la pluralité, mais

à trouver des modalités d'évaluation alternatives et complémentaires, qui permettent de valoriser diverses conceptions de la « bonne » sociologie et du « bon » sociologue. Il conviendrait alors de se contenter de classements partiels sans chercher à désigner le meilleur, ce qui, trop souvent, au lieu de susciter les émulations, est source de rancœurs ou de frustrations.

Le concept d'objectivité positionnelle développé par Amartya Sen permet de préciser notre propos, il explicite pourquoi une telle reconnaissance de la pluralité ne coïncide pas avec le relativisme. Suivant ce concept, toute personne occupe une position spécifique dans l'ordre social en vertu de sa trajectoire biographique, de son statut social, etc. Depuis cette position, elle porte un regard sur la réalité sociale qui lui est propre, mais qui a néanmoins une certaine objectivité, que Sen qualifie de positionnelle, dans le sens où elle n'est pas absolue mais liée à la position occupée par cette personne. Ainsi, la violence des jeunes ne sera pas définie de la même façon si elle est envisagée par un chef de service administratif, un travailleur social ou l'un des jeunes directement concernés. Pourtant, chacun de ces points de vue est doté d'une objectivité positionnelle propre et chacun des acteurs concernés est donc légitime à s'exprimer et à apporter des informations dans le processus de qualification ou d'évaluation de la situation sous examen. Il en va de même pour la sociologie et l'activité scientifique : tous les acteurs impliqués sont marqués par une objectivité positionnelle spécifique qui les conduira à privilégier une définition différente de ce qu'est la « bonne » sociologie. Si les critères d'évaluation définissent la qualité de l'activité scientifique de manière trop étroite, le risque existe que certaines objectivités positionnelles ne soient pas reconnues et donc que le travail de certains collègues ne soit pas apprécié à sa juste valeur, alors même qu'il est objectivement de bonne qualité.

La reconnaissance du pluralisme des modes d'évaluation légitimement mobilisables implique donc une certaine incomplétude des critères d'éva-

luation, qui ne doit pas être confondue avec le relativisme du “anything goes”. En effet, il existe, dans le domaine scientifique comme ailleurs, un consensus sur la validité de certains critères d'évaluation : dans le champ académique l'enseignement et la recherche sont importants, le service à la cité est également une composante à prendre en compte dans l'évaluation d'un sociologue ou d'un département de sociologie, etc. L'incomplétude ne revient pas à remettre en cause ces critères, mais plutôt à suggérer que leur hiérarchisation ne devrait pas être imposée d'en haut et de manière stricte à l'ensemble des sociologues ou des départements de sociologie. Il conviendrait au contraire de laisser une marge d'interprétation ou de manœuvre aux acteurs concernés, membres de la collectivité académique locale notamment. Il conviendrait aussi d'abandonner la prétention consistant à établir des classements complets suivant des critères identiques pour tous. C'est en effet aussi à l'échelon local, de manière située, que la question de l'évaluation devrait être tranchée. A ce niveau pourrait se déployer une évaluation contextualisée, qui prendrait en compte l'ensemble des paramètres pertinents, notamment tout ce qui concerne les moyens à disposition. Il semble en effet discutable d'évaluer à la même aune des sociologues ou des départements de sociologie disposant de moyens budgétaires ou logistiques très différents. De plus, cette « territorialisation » de l'activité évaluative permettrait de mieux prendre en compte les objectivités positionnelles des personnes concernées et de s'efforcer de résoudre la question de l'incommensurabilité des ordres de grandeur par le débat et l'échange d'arguments. L'échelon local ne devrait cependant pas avoir toutes les clés de l'évaluation entre ses mains, ce qui pourrait aboutir à des formes de justice locale similaires à celles que Weber dénonçait. Des critères objectifs, transver-

saux à l'ensemble des sociologues et départements de sociologie, sont nécessaires, mais ces critères devraient être suffisamment ouverts pour laisser un espace à la délibération locale sur la question, difficile et sans réponse univoque, de ce qu'est la « bonne » sociologie.

En résumé, quel que soit le champ où se déploie l'activité évaluative, l'investigation sociologique montre l'existence d'une pluralité des valeurs ou ordres de grandeurs légitimes, qui peuvent être à bon droit invoqués lors de l'évaluation d'une personne, d'une activité ou d'un bien donnés. Ce constat vaut également pour la sociologie. Dès lors, toute imposition d'une norme ou d'une valeur particulière revient à écarter les autres. Par contraste, le souci de l'incomplétude et la volonté de ne pas hiérarchiser les critères d'évaluation visent à reconnaître la pluralité des conceptions possibles du « bon » travail ou de la « bonne » sociologie. Ils autorisent également le déploiement de processus évaluatifs délibératifs à l'échelon local, processus qui sont au contraire entravés par l'imposition de critères uniformes d'évaluation. A notre sens, ces enseignements de la sociologie de l'évaluation devraient être pris en compte aussi pour l'évaluation de la sociologie ou, plus largement, de toute activité scientifique.

Quelques références

- Boltanski, Luc et Laurent Thévenot (1991). *De la justification*. Paris: Gallimard.
- De Munck, Jean et Bénédicte Zimmermann (2014). “Evaluation as Practical Judgment”. *Human Studies*, published on line 11 September 2014.
- Sen, Amartya (1993). “Positional Objectivity”. *Philosophy and Public Affairs* 22(2), 126–45.

«Forschungsleistung» in der Soziologie? Fallstricke bei der Leistungsmessung in einer paradigmatisch und kulturell fragmentierten Wissenschaft

Monica Budowski und Michael Nollert (Universität Freiburg)

Forschungsleistungen lassen sich nur messen, wenn innerhalb einer Disziplin ein Konsens besteht, was «gute Wissenschaft» ausmacht. Robert Mertons *Ethos der Wissenschaften* (1973) sieht vor, dass WissenschaftlerInnen ihre Forschungsergebnisse einem Prüfungsverfahren unterwerfen (organisierten Skeptizismus, Peer-Reviewing) und bei der Prüfung von Ergebnissen von KollegInnen sachfremde Kriterien, z. B. Geschlecht, Nationalität, ausblenden (Universalismus). Das Peer-Reviewing von Forschungsleistungen sollte demzufolge reliabel, valide und fair sein (Daniel 1993). Zwar besteht eine gewisse Validität des Peer-Reviewing, gemessen an der Zitation von akzeptierten Leistungen. Die Gutachterurteile korrelieren jedoch nur schwach, und in der Beurteilungspraxis lassen sich Urteilsfehler identifizieren (Shatz 2004; Gläser und Laudel 2006; Bornmann 2007; Hornbostel und Olbrecht 2007).

Diese Befunde sprechen jedoch nicht gegen Mertons Ethos, sondern gegen die Vorstellung, dass es einen von allen Disziplinen gemeinsam geteilten Katalog von wissenschaftlichen Gütekriterien gibt ("unity of science") (Feyerabend 1988). SozialkonstruktivistInnen zufolge macht es daher wenig Sinn, Forschungsleistungen zu vergleichen, die wie jene in der Soziologie sich an einer Vielfalt von «epistemologischen Kulturen», Paradigmen oder «Schulen» orientieren (Knorr-Cetina 1999).

Auch wenn man – wie Tashakkori und Teddlie (1998) – dennoch einräumt, dass es einen Kern von Qualitätskriterien/Leistungsstandards gibt, denen sich alle epistemologische Kulturen bzw. Paradigmen gleichermaßen verpflichtet fühlen, z. B. Regelgeleitetheit, intersubjektive Nachvollziehbarkeit, Transparenz des Verfahrens, ist doch

gerade für die Soziologie charakteristisch, dass sie stark paradigmatisch fragmentiert ist (Balog und Schulein 2008). Dabei verlaufen die paradigmatischen Grenzen nicht nur quer zu den disziplinären, sondern sind in vielen Forschungsbereichen sogar ausgeprägter als die interdisziplinären. So finden beispielsweise Systemtheoretiker oder Rational Choice-Anhänger zuweilen mehr Verständnis und Resonanz ausserhalb als innerhalb der Soziologie.

Angesichts der Vielfalt von unterschiedlichen Paradigmen, verstanden als «(..) allgemein anerkannte wissenschaftliche Leistungen, die für eine gewisse Zeit einer Gemeinschaft von Fachleuten massgebend Probleme und Lösungen liefern» (Kuhn 1979, 10), als kognitives System, das aus einer theoretischen, einer empirischen, einer methodologischen und programmatischen Komponente zusammengesetzt ist (Weingartner 1998, 9–11), ist die Soziologie folglich keine «reife Normalwissenschaft» im Sinne von Thomas Kuhn (1979) (vgl. z. B. Scott 2005 in Bezug auf die Soziologie). Abgesehen von Befürwortern «grosser Theorien» (Mills 1959) bedauert kaum jemand diese Vielfalt an Paradigmen, sondern sieht den intradisziplinären Wettbewerb vielmehr als Quelle von Innovativität und Kreativität.

Die Soziologie umfasst also eine Vielzahl von Denkstilen und Denkkollektiven, d. h. Gemeinschaften von Menschen, «die im Gedankenaustausch oder in gedanklicher Wechselwirkung stehen» und sich häufig als «Träger geschichtlicher Entwicklung eines Denkgebietes, eines bestimmten Wissensbestandes und Kulturstandes, also eines besonderen Denkstils» (Fleck 2006 [1935]) betrachten. Instrumente und Indikatoren von Forschungsleistungen, die diese «Koexistenz ri-

valisierender Paradigmen» (Schurz 1998) ignorieren, messen folglich nicht in erster Linie die Qualität von Forschungsleistungen, sondern die Qualitätsstandards der momentan vorherrschenden Paradigmen und Vorlieben der involvierten GutachterInnen (Guetzkow et al. 2004; Gläser und Laudel 2006).

Hinzu kommt, dass die Soziologie nicht bloss über eine transnational vernetzte Scientific Community mit entsprechenden hauptsächlich englischsprachigen Publikationsorganen und Veranstaltungen, sondern auch über viele kleinere Communities verfügt, die sich auf regionale und nationale Themen konzentrieren und ihre Befunde hauptsächlich in der Nationalsprache publizieren. Messinstrumente und Indikatoren, die ausschliesslich auf Datenbanken englischsprachiger Fachzeitschriften beruhen, benachteiligen daher vor allem all jene WissenschaftlerInnen, die vorzugsweise in sog. nationalen C-Zeitschriften publizieren (Hicks 1999).

Herkömmliche Verfahren zur Evaluation der Forschungsleistungen tendieren folglich dazu, die Diffusion der Forschungsergebnisse auf nationalen Tagungen, die Teilnahme an Podiumsdiskussionen, Vortragsreihen oder Sammelbände, Monografien und Beiträge in den Massenmedien und weitere Service Public- und Transfer-Leistungen im universitären Kontext zu vernachlässigen. Mit anderen Worten: Es stellt sich die Frage, inwiefern bei der Leistungsevaluation dem Begründungszusammenhang gegenüber dem Entdeckungs-/Verwertungszusammenhang (Friedrichs 1990) ein zu grosses Gewicht eingeräumt wird.

Zu beachten ist auch, dass Evaluationsinstrumente dazu veranlassen, nur noch zu leisten, was gemessen wird. Figuriert z. B. die Anzahl der Publikationen als Indikator für Qualität, schwinden Anreize, umfangreiche und zeitraubende Bücher wie etwa Norbert Elias' *Prozess der Zivilisation* (1939) zu schreiben. Herkömmliche Qualitätsindikatoren stimulieren also Verhalten, das den wissenschaftlichen Fortschritt gefährdet. In der

wissenschaftssoziologischen Literatur werden vorab zwei Effekte diskutiert (Weingart 2005):

Verlust intrinsischer Arbeitsmotivation: WissenschaftlerInnen sollten möglichst unabhängig von politischen und ökonomischen Zwängen Ideen entwickeln und gegebenenfalls überarbeiten oder verwerfen können (Dewett und Denisi 2005). Herkömmliche Evaluationen fördern eher eine extrinsische Arbeitsmotivation bei der die Erfüllung von Benchmarks im Vordergrund stehen (Amabile 1998; Osterloh und Frey 2008).

Anpassung an und Manipulation der Leistungskriterien: Wird eine hohe Zahl an Studierenden oder Dissertationen als gute Leistung betrachtet, steigt die Neigung, die (Lern-)Standards zu senken. Wissenschaftliche Zusammenarbeit wird untergraben, wenn die Leistung nur dem ersten Antragsteller von Forschungsprojekten angerechnet wird. Wenn die Zahl der Publikationen a priori als valider Forschungs- oder Qualitätsindikator betrachtet wird, wird es irrational, sich mit viel Zeitaufwand in neue und möglicherweise unmodische Forschungsthemen einzuarbeiten. Dagegen ist rational die Reputation zu erhöhen, indem man selbst «eigene» Zeitschriften gründet, indem man Daten manipuliert oder nationale und globale Zitierkartelle aufbaut (Schulze und Frank 2003). Kurzum: Nicht mehr die WissenschaftlerInnen selbst oder die Peers, sondern die Messverfahren und Indikatoren zur Forschungsleistung entscheiden, welche Leistungen forciert und welche vernachlässigt werden (Frey 2006).

Problematisch an herkömmlichen Evaluationsverfahren ist schliesslich, dass sie die WissenschaftlerInnen als EinzelkämpferInnen konzipieren und folglich vornehmlich die Leistungen der «Stars» («genius scientists», Latour 1987) estimieren. Die traditionelle Vorstellung, dass wissenschaftlicher Fortschritt auf Kooperation beruht, bleibt dabei genauso auf der Strecke wie die inspirierenden Beiträge vermeintlich «mittelmässiger» Teammitglieder, den sogenannten «Wasserträgern» (Knorr-Cetina 1999; Braun 2004).

Fazit: SoziologInnen tun gut daran, sich bei der Messung ihrer Forschungsleistung daran zu erinnern, dass ihre Disziplin paradigmatisch fragmentiert ist und es folglich Sinn macht, die Koexistenz unterschiedlicher Qualitätsstandards anzuerkennen. Die Kreativität und Innovativität der Soziologie wird unterminiert, wenn sie sich bei der Evaluation ihrer Forschungsleistungen einseitig einer naturwissenschaftlichen Epistemologie unterordnet.

Literatur

- Amabile, Teresa (1998). How to Kill Creativity. *Harvard Business Review* 76(5), 76–87.
- Balog, Andras und Johann August Schüle (Hrsg.) (2008). *Soziologie, eine multiparadimatische Wissenschaft. Erkenntnisnotwendigkeit oder Übergangsstadium*. Wiesbaden: VS Verlag.
- Bornmann, Lutz (2007). Peer Review in den Wissenschaften – eine Analyse des Begutachtungsverfahrens aus der Sicht wissenschaftssoziologischer Theorien. *Schweizerische Zeitschrift für Soziologie*, 33(2), 327–347.
- Braun, Dietmar (2004). Forschungsnetzwerke fördern intellektuelle Fremdbestäubung. *Informationen zur Europäischen und internationalen Forschung*, Nr. 8, 1.
- Daniel, Hans-Dieter (1993). *Guardians of Science. Fairness and Reliability of Peer Review*. Weinheim: Wiley-VCH.
- Dewett, Todd und Angelo Denisi (2004). Exploring Scholarly Reputation: It's More Than Just Productivity. *Scientometrics*, 60(2), 249–272.
- Elias, Norbert (1968 [1939]). *Über den Prozess der Zivilisation. Soziogenetische und psychogenetische Untersuchungen*. Frankfurt am Main: Suhrkamp.
- Feyerabend, Paul (1988). *Against Method* (revised edition). London: Verso.
- Fleck, Ludwik (2006 [1935]). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*. Frankfurt am Main: Suhrkamp.
- Frey, Bruno (2006). *Evaluitis – Eine Neue Krankheit*. Working Paper Nr. 293. Institut für Empirische Wirtschaftsforschung der Universität Zürich, Zürich.
- Friedrichs, Jürgen (1990). *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.
- Gläser, Jochen und Grit Laudel (2006). Advantages and Dangers of “Remote” Peer Reviews. *Research Evaluation*, 14(3), 186–198.
- Guetzkow, Joshua, Michele Lamont und Gregoire Mallard (2004). What Is Originality in the Humanities and the Social Sciences? *American Sociological Review*, 69(2), 190–212.
- Hicks, Diana (1998). The Difficulty of Achieving Full Coverage of International Science Literature and Bibliometric Consequences. *Scientometrics*, 44(2), 193–215.
- Hornbostel, Stefan und Meike Olbrecht (2007). *Peer Review in der DFG: Die Fachkollegiaten*. iFQ Working Paper No. 2, November 2007, Bonn.
- Knorr-Cetina, Karin (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Kuhn, Thomas S. (1979). *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt am Main: Suhrkamp.
- Latour, Bruno (1987). *Science in Action*. Cambridge Massachusetts: Harvard University Press.
- Merton, Robert (Hrsg.) (1973). *The Sociology of Science. Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- Mills, C. Wright (1959). *The Sociological Imagination*. Oxford: Oxford University Press.
- Osterloh, Margrit und Bruno Frey (2008). *Anreize im Wissenschaftssystem*. CREMA – Center for Research in Economics, Management and the Arts, Switzerland Zürich.
- Tashakkori, Abbas und Charles Teddlie (1998). *Mixed Methodology. Combining Qualitative and Quantitative Approaches*. Thousand Oaks, London, New Delhi: Sage Publications.

- Schulze, Günther und Björn Frank (2003). Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility. *Economics of Governance* 4, 143–160.
- Schurz, Gerhard und Paul Weingartner (Hrsg.) (1998). *Koexistenz rivalisierender Paradigmen*. Opladen: Westdeutscher Verlag.
- Scott, John (2005). Sociology and Its Others: Reflections on Disciplinary Specialisation and Fragmentation. *Sociological Research Online*, Vol. 12. <http://www.socresonline.org.uk/10/1/scott.html>
- Shatz, David (2004). *Peer Review: A Critical Inquiry*. Lanham: Rowman & Littlefield.
- Weingart, Peter (2005). Impact of Bibliometrics Upon the Science System: Inadvertent Consequences. *Scientometrics*, 62(1), 117–131.

Evaluationen als Dispositiv – eine konventionentheoretische Perspektive

Rainer Diaz-Bone (Universität Luzern)

Die Anfänge der Evaluation und der Evaluationsforschung liegen in der sozialwissenschaftlichen Analyse der Wirkungen von Politikprogrammen und gesellschaftlichen Reformprogrammen. Zuerst erfolgte dies in den USA seit den späten 1950er-Jahren (Weiss 1974; Rossi et al. 1988). Später kam die Evaluationsforschung dann auch in den westeuropäischen Ländern in den 1980er-Jahren allmählich auf, zu einer Zeit, als die Sozialwissenschaften begannen, in den westlichen Ländern das Regierungshandeln und die Regierungsprogramme nicht nur einer sozialkritischen Analyse zu unterziehen, sondern das methodische Wissen der Sozialwissenschaften systematisch in den Dienst für das Regierungshandeln selbst zu stellen (Hellstern und Wollmann Hrsg. 1984; Wottawa und Thierau 1989). Ursprünglich waren die Wirkungen von solchen Programmen Gegenstand der Evaluationsforschung, die zur Behebung sozialer Probleme wie Arbeitslosigkeit, Armut, Diskriminierung, Kriminalität u. a. lanciert worden waren. Die Rationalisierung des Regierungshandelns und die Verbesserung sozialer Verhältnisse (die Behebung sozialer Missstände) waren wichtige Impulse für die Evaluationsforschung. Regierungen wurden so aufgefasst als wissenschaftlich informierte Agenturen für die Erreichung des sozialen Fortschritts. Und bald schon wurden Evaluationen fester Bestandteil des Regierungshandelns und der Praxis öffentlicher Verwaltungen. Begriffe wie «Rationalisierung» und «Verbesserung» wurden eingeführt in Zeiten, in denen man die staatliche Aufgabe gerade auch in der Behebung sozialer Probleme gesehen hatte (Giel 2013; Stockmann und Mayer 2014). Diese Begriffe hatten damals eine scheinbar evidente Bedeutung, die man in der Verfolgung eines kollektiven Gemeinwohls sehen konnte. Mit dem Wandel des Staatsverständnisses haben sich auch die Bedeu-

tung dieser Begriffe sowie die Verwendungen von Evaluationen gewandelt, insbesondere durch den sogenannten Neoliberalismus, der in der Privatisierung vormals staatlicher Aufgaben und öffentlicher Güter gesehen wird. Zugleich sind – sozusagen als Gegenbewegung dazu – nutzenorientierte und partizipative Formen der Evaluation entwickelt worden, die die sozialwissenschaftliche Evaluationsforschung auch in den Dienst zivilbürgerlicher Anliegen stellt und Fragen der Macht, der Politik sowie der Ethik als zentral für die Evaluationsforschung ansehen (Guba und Lincoln 1989; Patton 2008). Insbesondere der Ansatz von Guba und Lincoln hat sich bereits Ende der 1980er-Jahre gegen eine Indienststellung der Evaluationsforschung für die alleinigen Zwecke von «Auftraggebern» gewendet, was als Kritik am «Managerialismus» formuliert worden ist, der zu Unfairness und Entmachtung anderer Gruppen von Beteiligten bzw. Betroffenen führe (Guba und Lincoln 1989; Giel 2013). Evaluationen sind heutzutage in vielen sozialen Feldern allgegenwärtig. Die öffentlichen Verwaltungen setzen in sie immer noch die Hoffnung, sie als ein Steuerungsinstrument der Wissenschaft einsetzen zu können. Universitäten mit ihren Einheiten (Fakultäten, Institute, Seminare) sowie Studienprogramme werden seit Jahren und zunehmend Evaluationen unterzogen. Aber Evaluationen geraten mit dem Ausmass ihrer Verbreitung auch in die – gerade auch massenmediale – Kritik. Es sind verschiedene Aspekte ihrer gegenwärtigen Praxis, die kritisch beurteilt werden.

(1) Einmal, dass Evaluationen oftmals zweckentfremdet werden, dies in dem Sinne, dass sie lediglich aus legitimatorischen Gründen erfolgen. Ein Beispiel sind Lehrevaluationen, wenn sie faktisch Studierendenzufriedenheitsbefragungen sind und keine Evaluationen der Lehrqualität. Sie

dienen dann der Selbstlegitimierung universitärer Verwaltungen und sie erfassen nicht den Zusammenhang von Lehrqualität und Lernerfolg. Anstatt die Lehrpraxis zu evaluieren werden faktisch Meinungsumfragen unter Studierenden durchgeführt. Letztere werden dabei weniger als nach wissenschaftlichen Standards zu Qualifizierende, sondern zuerst als an Universitäten «ingeschriebene Kunden» betrachtet, die es zufriedenzustellen gilt. Gegen studentische Zufriedenheit ist zunächst wenig einzuwenden, die Verwechslung studentischer Zufriedenheit mit der kompetenten Beurteilung von Lehrqualität ist dagegen problematisch. Solche «Evaluationen» können für die Beteiligten (Studierende und Dozierende) aufwändig sein. Auch wenn sie ihnen als wenig sinnvoll und so als illegitim erscheinen.

(2) Weiter stehen Evaluationen in der Kritik, wenn ihre Form den zu evaluierenden Produkten und Praktiken nicht angemessen erscheint. Das betrifft nicht nur das Design von Evaluationen als Formen der Sozialforschung (Bortz und Döring 2006; Kromrey 2001, 2009). Damit ist auch gemeint, dass Informationen erhoben werden, die nicht als valide für die Erfassung der zu evaluierenden Qualitäten angesehen werden. Das betrifft nicht nur die Frage, ob die evaluationsrelevanten Aspekte erfasst werden, sondern auch, wie sie in eine kognitive Form gebracht werden, wie sie also «in-formiert» werden. Laurent Thévenot hat darauf aufmerksam gemacht, dass Organisationen für die gelingende Koordination nicht nur in eine materielle Instrumentierung investieren müssen, sondern dass auch eine Investition in die kognitive Instrumentierung erfolgen muss, die den Beteiligten ermöglicht, die relevanten Produkte und Praktiken auf ihre Qualitäten beurteilbar zu machen. Thévenot (1984) hat diese Investition «Forminvestition» genannt. Quantitative Indikatoren können sinnvolle kognitive Formen sein, aber die Frage, unter welchen Bedingungen eine Quantifizierung eine für die Koordination zielführende Forminvestition ist oder ob narra-

tive Formen oder auch eine Kombination beider geeigneter sind, wird nicht immer ausreichend vorab beraten – oftmals erscheinen quantitative Indikatoren selbstlegitimierend. Damit zusammen hängt die Frage, ob die Evaluation auch für die von ihr betroffenen Personen handlungsrelevante Informationen generieren kann. Denn die für die Indikatorenbildung herangezogenen Definitionen beruhen auf Konventionen. Alain Desrosières hat dies prägnant formuliert: «Quantifier, c'est convenir puis mesurer» (Desrosières 2008, 10). Stehen diese fundierenden Konventionen den Informationsbedürfnissen und den Koordinationsanforderungen der betroffenen Personen entgegen, dann sind entsprechend die Quantifizierungen aus Sicht der Evaluierten unbrauchbar (siehe auch den Beitrag von Bonvin in diesem Bulletin). Die Konventionen, die die Qualität von zu evaluierenden Sachverhalten quantifizierbar machen sollen, werden im Ansatz der *Economie des conventions* «Qualitätskonventionen» genannt (Diaz-Bone 2015). Durchführung (Design) von Evaluationen, Art der Forminvestition, herangezogene Qualitätskonventionen sowie Indikatorenbildung stehen also als Verbund vielfach in der Kritik, die gerade auch von den Sozialwissenschaften gegenüber den Verwaltungen geäußert wird.

(3) Hinzu kommt ein dritter Aspekt, der gerade aus Sicht der *Economie des conventions* evident ist. Qualitäten können in vielfacher Weise bestimmt werden, Strategien der Quantifizierung sind dabei nicht die einzigen Formen. Der Grund ist, dass es eine Pluralität von Qualitätskonventionen gibt, auf die Akteure sich in ihrer Koordination, aber auch in Situationen der Kritik und Rechtfertigung berufen können. (Boltanski und Thévenot 2007; Lamont 2009; Diaz-Bone 2015). Bezogen auf Evaluationen bedeutet dies, dass es eine Pluralität denkbarer Evaluationsformen für eigentlich jede Institution oder jedes Programm geben kann. Die Praxis der Evaluation ist also keineswegs vorab durch den zu evaluierenden Sachverhalt sachlich festgelegt. Weiter bedeutet dies auch, dass eine Evaluation

eine Kombination verschiedener Qualitätskonventionen heranziehen kann und es nicht vorab evident ist, dass sie auf nur einer Konvention als Prinzip der Vergleichbarmachung von Qualitäten beruht. Wenn eine zu evaluierende Disziplin selbst durch eine Pluralität von Qualitätskonventionen geprägt ist, dann erscheint die Evaluation anhand nur einer Qualitätskonvention von Wissenschaft als schlichtweg unangemessen. Die Kritik, die sich dann Bahn bricht, ist diejenige an einer Evaluationspraxis, die die Pluralität (Heterogenität) von Wissenschaftsqualitäten ignoriert und damit keine validen Resultate zu Tage fördern kann, die für die Nutzerinnen und Nutzer wirklich nützlich sein könnten. Allerdings bedeutet dies nicht, dass solche «monistischen» Evaluationen, die nur ein Vergleichsprinzip heranziehen, keine Effekte hätten, denn sie generieren eigene Interpretationen und Machtwirkungen.

(4) Werden Evaluationen für die staatliche Regulierung und Mittelzuteilung an Wissenschaft durch Verwaltungen eingesetzt, dann wird die Evaluation immer dann als Mittel der Verfolgung eines Gemeinwohls sowie der Analyse von Wirkungszusammenhängen unterminiert und transformiert zu einem «Dispositiv der Macht» (Foucault 1978), wenn Verwaltungskriterien (Auftraggeberkriterien) an die Stelle von Wissenschaftskriterien (Nutzerkriterien) treten und Verwaltungen Evaluationen asymmetrisch für ihre Zwecke einsetzen. Das erfolgt nicht erst durch den sogenannten Neoliberalismus oder durch die Ökonomisierung der Wissenschaft, sondern durch die Verwaltungen selbst, wenn sie für das Verwaltungshandeln Konventionen heranziehen, die für die Verwaltungsvollzüge praktikabel sind, aber mit der Wissenschaftspraxis und deren Qualitätsbestimmungen nicht kohärent sind. Auf diese Verwaltungspraxis reagieren die evaluierten Gruppen vorrangig damit, dass sie Evaluationen als Dispositiv ausser Kraft zu setzen versuchen. Wenn dies gelingt, dann können solche Evaluationen in der Praxis tatsächlich nicht die «Qualität der Lehre

und Forschung» repräsentieren, was dann auch schnell so angeprangert wird, mit der Folge, dass solche Evaluationen schnell diskreditiert sind. Was fehlt ist die Grundlage für eine geteilte kollektive Intentionalität, die anvisiert, für alle relevanten und beteiligten Gruppen nützlich Wissen zu erzielen und es fehlt die Kooperationsbereitschaft aller involvierten Gruppen, da die Wahrnehmung ist, dass kein kollektives Gemeinwohl, sondern ein Partikularinteresse verfolgt wird. Sollen Evaluationen als «Dispositiv der Macht» fungieren können, dann kann auch eigentlich kein echtes Interesse an einer validen Repräsentation der «Qualität von Lehre und Forschung» bestehen, wenn dafür die durch dieses Dispositiv zu Bemächtigenden zugleich als Expertinnen und Experten gelten müssten. Von daher erklärt sich, dass viele dieser Evaluationen Instrumente wie Software-Programme mit vorformulierten Evaluationsfragen verwenden oder überwiegend organisatorische Kriterien von Verwaltungen heranziehen wie Anzahl betreuer Arbeiten, Anzahl durchgeführter Gutachten, Umfang der angeworbenen Drittmittel. Viele dieser Indikatoren sind zudem in verschiedener Weise interpretierbar. Die Anzahl betreuer Arbeiten kann die Überlastung und daher Betreuungsmängel *oder* die Beliebtheit und daher die Lehrqualität von Dozierenden indizieren. Intensive Gutachtertätigkeit kann den herausragenden Expertenstatus anzeigen *oder* die Neigung sich für Gutachten in dem eigenen Forschungsbereich gerne zur Verfügung zu stellen und als zuverlässig zu gelten. Bei Drittmittelwerbungen wird der Grad der Exzellenz der Forschung indiziert *oder* die Neigung, solche Forschungen zu betreiben, die viel finanzielle Mittel aber eher wenig eigene Zeitinvestition benötigen, so dass die Zeit für die Antragstellung gut investiert ist – im letzten Fall wird Wissenschaftsmanagement als Forschungsexzellenz indiziert. Das Gründungsmoment der Evaluationsforschung, fortschrittliche Reformprogramme auf ihre Wirkungen hin kausal und empirisch-methodisch zu analysieren, wird verkehrt zu einem Benchmarking, in dem die

Kriterien nicht durch die Wissenschaft bzw. ihre Disziplinen, sondern durch andere vorgegeben werden, nach denen sich die Wissenschaft dann auszurichten hat. Man kann daher umgekehrt folgern, dass eine Evaluation in dem Mass geeignet sein kann, die Qualität(en) von Lehre und Forschung mit ihren Resultaten valide zu repräsentieren, wie weit die evaluierte Wissenschaft unabhängig ist, die eigenen Qualitätskonventionen und Forminvestitionen der Evaluation zugrunde zu legen und die Evaluation letztlich selbst mit Bezug auf diese durchzuführen. Dies ist eine notwendige, wenn auch keine hinreichende Bedingung.

Die disziplinären Fachgesellschaften sind durch die *Schweizer Akademie der Geistes- und Sozialwissenschaften* (SAGW) aufgefordert worden, eigene disziplinspezifische Evaluationskriterien zu entwickeln. Diese Anregung kann eine spezifische Relevanz erhalten. Denn die *Schweizerische Gesellschaft für Soziologie* (SGS) kann die Evaluationspraxis daraufhin evaluieren und auszurichten helfen, ob und wie sie als Dispositiv für und durch die Disziplin selbst fungiert bzw. fungieren kann. Das macht eine soziologische Kritik von Evaluationen als Verwaltungspraxis unumgänglich, wie sie durch viele der Beiträge in diesem Bulletin auch angestrebt bzw. praktiziert wird. Diese ist aber als «Soziologie der Evaluation» (Salais 2010a, 2010b; Lamont 2012) bislang noch wenig avanciert und sie hätte Allianzen einzugehen mit Ansätzen der Soziologie der Quantifizierung (Espeland und Stevens 2008; Desrosières 2008; Salais 2012), beide wären selbst aber nicht durch eine SGS zu entwickeln, sondern im Fach selbst. Eine Kritik der Evaluation sollte doch aber weiterführen. Bislang fehlt beispielsweise eine wissenschaftliche und unabhängige Instanz, die Evaluationen soziologischer Lehre und Forschung berät, begleitet und durchführt. Diese müsste durch die soziologische Fachorganisation ins Werk gesetzt, legitimiert und auch beaufsichtigt werden. Stattdessen hat sich eine Evaluationsbranche entwickelt, in der Unternehmen diese Dienstleistungen anbieten.

Aber wer evaluiert diese privatwirtschaftlichen Evaluationsdienstleister, wenn sie soziologische Qualität(en) beurteilen sollen? Dagegen könnte eine unabhängige Instanz als eine *soziologische Evaluationsagentur* und zugleich als ein Dispositiv für die Positionierung der Soziologie sowie auch für die Autonomie ihrer Wissenschaftsbewertung fungieren. Dies nicht nur in der Weise, dass sie sich damit der Bemächtigung durch andere entziehen kann, sondern dass die so durchgeführten Evaluationen der Pluralität der Qualitätskonventionen des Fachs angemessen Rechnung tragen und für seine Selbstbeobachtung, Selbstbewertung und Selbstreflexion sowie vielleicht auch für seine Selbststeuerung nützlich sein können.

Literatur

- Boltanski, Luc und Laurent Thévenot (2007). *Über die Rechtfertigung*. Hamburg: Hamburger Edition.
- Bortz, Jürgen und Nicola Döring (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Aufl. Heidelberg: Springer.
- Desrosières, Alain (2008). *Pour une sociologie historique de la quantification. L'argument statistique I*. Paris: Mines ParisTech.
- Diaz-Bone, Rainer (2015). *Die «Economie des conventions»*. Grundlagen und Entwicklungen der neuen französischen Wirtschaftssoziologie. Wiesbaden: Springer VS Verlag.
- Espeland, Wendy und Mitchell Stevens (2008). A sociology of quantification. *European Journal of Sociology* 49(3), S. 401–436.
- Foucault, Michel (1978). *Dispositive der Macht*. Berlin: Merve.
- Giel, Susanne (2013). Theoriebasierte Evaluation: *Konzepte und methodische Umsetzung*. Münster: Waxmann.
- Guba, Egon und Yvonna Lincoln (1989). *Fourth generation evaluation*. Newbury Park: Sage.

- Hellstern, Gerd-Michael und Hellmut Wollmann, (Hrsg.) (1984). *Handbuch zur Evaluierungsforschung, Bd. 1*. Opladen: Westdeutscher Verlag.
- Kromrey, Helmut (2001). Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. *Sozialwissenschaft und Berufspraxis* 24(2), S. 1–23.
- Kromrey, Helmut (2009). *Empirische Sozialforschung*. 12. Aufl. Stuttgart: Lucius & Lucius.
- Lamont, Michèle (2009). *How professors think. Inside the curious world of academic judgment*. Cambridge: Harvard University Press.
- Lamont, Michèle (2012). Towards a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38, S. 201–221.
- Patton, Michael (2008). *Utilization-focused evaluation*. 4. Aufl. Thousand Oaks: Sage.
- Rossi, Paul, H. Freeman, E. Howard und Gerhard Hofmann (1988). *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Salais, Robert (2010a). Usages et mésusages de l'argument statistique: le pilotage des politiques publiques par la performance. *Revue française des affaires sociales* 1(1/2), S. 129–147.
- Salais, Robert (2010b). La donnée n'est pas un donné. Pour une analyse critique de l'évaluation chiffrée de la performance. *Revue française d'administration publique* 135, S. 497–515.
- Salais, Robert (2012). Quantification and the economics of convention. *Historical Social Research* 37(4), S. 55–63.
- Stockmann, Reinhard und Wolfgang Mayer (2014). *Evaluation. Eine Einführung*. 2. Aufl. Opladen: Barbara Budrich.
- Thévenot, Laurent (1984). Rules and implements: Investments in forms. *Social Science Information* 23(1), S. 1–45.
- Weiss, Carol H. (1974). *Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogrammen*. Opladen: Westdeutscher Verlag.
- Wottawa, Heinrich und Heike Thierau (1989). *Lehrbuch Evaluation*. Bern: Huber.

Nachrufe auf Kurt Imhof (1956–2015) / Nécrologies pour Kurt Imhof (1956–2015)



La Société suisse de sociologie (SSS) a la tristesse d'annoncer le décès de Kurt Imhof, ancien vice-président et membre très actif du Comité, survenu le 1^{er} mars 2015. Nous adressons à ses proches, à ses amis et collègues nos sincères condoléances.

Pour honorer cette figure de la sociologie suisse, nous avons tenu à publier dans ce Bulletin deux hommages écrits par Thomas Eberle et Ueli Mäder.

Le comité exécutif du SSS

Die Schweizerische Gesellschaft für Soziologie (SGS) betrauert den Tod von Kurt Imhof, eines ehemaligen Vizepräsidenten und sehr aktiven Vorstandsmitgliedes. Er verstarb am 1. März 2015. Wir übermitteln den Freunden und Kollegen unser Beileid.

Um diesen Menschen zu ehren, werden zwei Nachrufe in diesem Bulletin publiziert, die von Thomas Eberle und Ueli Mäder verfasst wurden.

Der Vorstand der SGS

Ein engagierter Intellektueller

Mit Kurt Imhof verlässt eine markante Persönlichkeit die Schweizer Öffentlichkeit. Der Professor für Publizistikwissenschaft und Soziologie hatte Substanz und Konturen. Er konnte auch über sich selbst lachen.

Substanziell lagen Kurt Imhof die Qualität und Vielfalt der Medien am Herzen. Das ist von ihm weithin bekannt. Medien sollten präzise und aufdeckend informieren, zum eigenen Denken anregen und ein kritisches Korrektiv zur Wirtschaft, Gesellschaft und Politik sein. Im dicken «Jahrbuch Qualität der Medien» analysierte und reflektierte Kurt Imhof die Entwicklungen. Harsch kritisierte er Tendenzen der Personalisierung und Ökonomisierung.

Kurt Imhof betrachtete Medienereignisse als zentrale Indikatoren für den sozialen Wandel. Sie dokumentieren, was sich gesellschaftlich tut. Und zwar in doppeltem Sinne. Medien konstruieren die Ereignisse mit, die sie beschreiben. Und wie sie das tun, oberflächlich oder geistreich, sagt viel über öffentliche Diskurse und Befindlichkeiten aus. Das vermittelte Kurt Imhof auch angehenden Journalistinnen und Journalisten als Kursleiter und Dozent am Schweizerischen Medien-Ausbildungszentrum (MAZ) in Luzern.

Zäsuren und thematische Vielfalt

Kurt Imhof erwähnte immer wieder zwei Zäsuren: 1968 und 1989. 1968 öffneten sich Horizonte. Dies im Sinne einer erweiterten Aufklärung: Sei realistisch und denke das Unmögliche! Kurt Imhof nahm diese freiheitliche Tradition auf. Und 1989? Da brach die Berliner Mauer auf. Und mit ihr der West-Ost-Gegensatz. Der Kalte Krieg schien beendet zu sein. Was für ein hoffnungsvolles Ereignis. Die Rüstungsausgaben sanken. Allerdings nur für kurze Zeit. Kurt Imhof wies im September 2007 am Basler Kongress der Schweizerischen Gesellschaft für Soziologie eindrücklich auf neue ideologische Kriegsgefahren und hegemoniale Ansprüche hin. Die Aktualität gibt ihm leider Recht. Mit der

Zäsur von 1989 setzte sich Kurt Imhof auch im Buch «Triumph und Elend des Neoliberalismus» (2005) auseinander. Das Kapital drängt seither offensiver dorthin, wo es sich optimal verwerten lässt. Ein finanzliberales Verständnis überlagert das politisch liberale, das mehr Wert auf ein ausgewogenes Verhältnis zwischen Kapital und Arbeit und auf einen sozialen Ausgleich legte.

Ein weiteres Thema: das Fremde. Kurt Imhof befasste sich in den 1990er-Jahren intensiv damit. So an Kongressen und in Studien. Zudem ab 1995 als Direktor des Forums für schweizerische Migrationsstudien. Er unterschied die Fremdenfeindlichkeit von der Fremdenangst. Wer diese verstehen will, muss sich auch – selbstreflexiv – dem zuwenden, was einem selbst im Vertrauten fremd bleibt. So entdecken wir auch eher das Vertraute im Fremden. Ab 1997 konzentrierte sich Kurt Imhof dann erfolgreich darauf, am Zürcher Institut für Soziologie den Forschungsbereich Öffentlichkeit und Gesellschaft aufzubauen und zu leiten.

Lebensnah und debattierfreudig

Kurt Imhof war ein lebensnaher Intellektueller. Bevor er, 25-jährig, die Matura nachholte, lernte er Bauzeichner. Als Werkstudent unterrichtete er dann Staats- und Wirtschaftskunde am Kaufmännischen Lehrinstitut Zürich (KLZ). Später wirkte er als Dozent an der Höheren Wirtschafts- und Verwaltungsschule (HWV). Zudem an der Berufsschule der Invalidenversicherung und an der Volkshochschule. Nach mehreren Gastprofessuren berief ihn die Universität Zürich im Jahr 2000 auf den Lehrstuhl für Publizistikwissenschaft und Soziologie. Damit zog er sich aber keineswegs in den Elfenbeinturm zurück. Im Gegenteil. Kurt Imhof bewegte sich noch motivierter auf dem öffentlichen Parkett. Er kommentierte an unzähligen Veranstaltungen das gesellschaftliche Geschehen aus soziologischer und medienwissenschaftlicher Sicht und nahm dabei kein Blatt vor den Mund. Erst recht nicht in der Arena oder im Club im

Schweizer Fernsehen. Die einen schätzten das, andere weniger.

Kurt Imhof engagierte sich auch in der Schweizerischen Gesellschaft für Soziologie. Er reiste als Vizepräsident selbst an kalten Wintertagen mit seinem Töff an unsere Vorstandssitzungen in Bern. Nachher genehmigten wir uns meistens noch eine Pizza und debattierten eifrig weiter. Kurt argumentierte inspirierend. Er stellte Fragen und wollte bei jeder Begegnung etwas Neues erfahren. Wie siehst Du dies oder jenes? Und er legte selbst gerne provokative Thesenpapiere vor. Kurt Imhof liebte Debatten und hörte andern gebannt zu, ohne sich einfach selbst zu inszenieren. Er wollte das aktuelle Geschehen ursächlich ergründen, suchte den Widerspruch und spitzte eigene Argumente gerne scharf zu.

Engagiert und lebensfroh

Vor ein paar Monaten nahmen wir eine gemeinsame Promotion ab. Zusammen mit einem konservativen Sicherheitsexperten. Die Doktorandin kam dabei auch zu Wort. Und die intensive Debatte dauerte noch länger über die offizielle Prüfungszeit hinaus: über die Folgen des Zerfalls der alten

Sowjetunion und die Wiederbelebung des Islams in Zentralasien. Kurt Imhof vertrat, religionssoziologisch fundiert, ein differenziertes Weltbild. Er promovierte selbst – anno 1989 – über Diskontinuitäten der Moderne. Und zwar in der Sozial- und Wirtschaftsgeschichte bei Hansjörg Siegenthaler, mit dem er den Club Helvétique begründete: aus Sorge um die Zukunft der Schweiz. Kurt Imhof engagierte sich persönlich und mit dem Club leidenschaftlich für einen soliden Rechtsstaat und dafür, die Demokratie zu stärken.

Kurt Imhof liebte das Leben. Und jetzt erlebt er nicht einmal mehr seinen sechzigsten Geburtstag. Ich befürchtete ab und zu, er könnte einmal mit dem Motorrad verunfallen, auf das er sich in seiner «verwetzten» Lederjacke jeweils in letzter Sekunde schwang. Nun starb er am 1. März 2015 an Krebs. Im Februar schrieb er uns noch: «Nun wird zurückgeschlagen, der Krebs wird verbrannt (Radiochirurgie), verstrahlt und in einer giftigen Chemiejauche ersäuft (Chemo). Wird ein harter unberechenbarer Kampf, aber ich will ihn zumindest auf Zeit gewinnen! On verra.»

Ueli Mäder (Universität Basel)

Kurt Imhofs Vermächtnis an die SGS

Im Gedenken an unseren langjährigen Vizepräsidenten (1998–2013).

Kennengelernt habe ich Kurt Imhof an einer Party. Ich weiss auch noch genau wann und wo. Wir waren von der Gastgeberin am selben grossen runden Tisch platziert worden. Er fiel mir auf, als er seiner Tischnachbarin dezidiert und eindringlich erklärte, warum er keine eigenen Kinder wolle: Er sei zu sehr mit sich selbst und der Wissenschaft beschäftigt. War er verheiratet mit der Wissenschaft? Zweifellos. Aber nicht nur: Ich habe auch viele andere Seiten von ihm kennengelernt, eine Zeit lang pflegten wir eine nahe Freundschaft und erzählten uns Privates und auch Intimes. Aber Wissenschaft und Forschung forderten unerbittlich ihren Tribut, und der Aufbau des Forschungsinstituts Öffentlichkeit und Gesellschaft (*fög*) verlangte seine volle Aufmerksamkeit.

Ein früherer Präsident der SGS sagte einmal, Soziologen studierten soziale Organisationen, weil sie selbst nicht organisieren können. Daran dachte ich immer wieder mal, stiess ich doch im Laufe meiner Berufsbiografie reichlich auf derartige Angehörige unserer Spezies. Kurt Imhof gehörte nicht dazu. Das *fög* verlangte Unternehmertum und Managementfähigkeiten, und Imhof war begierig, beides weiterzuentwickeln. Ich erinnere mich, wie er mal zu uns an die HSG kam und sich die Aufbau- und Prozessorganisation des *mcm* – des Instituts für Medien- und Kommunikationsmanagements, das neu gegründet wurde und innerhalb weniger Jahre auf einen Stab von 50 wissenschaftlichen Mitarbeitenden angewachsen war – erklären liess und wie er anschliessend begeistert, aber auch verblüfft und ein wenig resigniert zu mir sagte: «Also im Vergleich dazu sind wir an der Uni Zürich «einfach Weggliebue.»»

Während meiner Zeit als Präsident der SGS (1998–2005) hatte Kurt Imhof das Amt des Vizepräsidenten inne, und er blieb der Vize bis Ende 2013. Er wollte mitmachen und mitgestalten, hat

sich aber nie selbst in den Vordergrund gedrängt: Das *fög* war ihm Aufgabe genug, Bürde auch und gleichzeitig Freude und Stolz – auf jeden Fall war es genug Verantwortung und forderte viel von ihm ab. Das Vorhaben des damaligen Vorstands jedoch, die SGS etwas professioneller und nachhaltiger zu organisieren, trug er mit vollster Überzeugung mit. Wir schufen eine neue Corporate Identity, druckten erstmals professionell gestaltete farbige Plakate für unsere Kongresse als auch zum Zweck der Werbung für die SGS, kreierte mit professioneller Hilfe ein neues Logo und verpassten sowohl dem Bulletin als auch der Zeitschrift ein neues Design.

Wir gingen aber auch inhaltlich neue Wege. Kurt Imhof war stets federführend mit dabei, wenn es um die Bestimmung von Kongress Themen und das Verfassen von Calls for Papers ging. Mit grossem Interesse verfolgte er die gesellschaftlichen Entwicklungen und war dank des grossangelegten Medienmonitorings am *fög* nicht nur besonders gut informiert, sondern hatte auch ein ausgeprägtes Gespür für aktuelle Themen. Es faszinierte ihn, wie es der SVP seit den 1990er zunehmend gelungen war, bezüglich öffentlichem agenda-setting die Führung zu übernehmen. Es war Imhofs Idee, dass die SGS das agenda-setting für die Soziologie übernehmen müsse. In wissenschaftspolitischer Hinsicht hatten wir dies versucht, indem wir den SGS-Kongress 1999 den interpretativen Soziologien widmeten. Dies weckte eine bemerkenswerte Aufmerksamkeit bis weit über die Landesgrenzen hinaus, war doch eine derart spezifische Schwerpunktsetzung an einem DGS- oder ÖGS-Kongress damals undenkbar. Imhof war jedoch überzeugt, dass sich die Soziologie nicht wissenschaftsinternen Themen, sondern Themen mit grosser gesellschaftspolitischer Brisanz widmen solle – damit würde sie auch eine breite Resonanz in den Medien und der weiteren Öffentlichkeit gewinnen. So wurde es zum Ziel des damaligen Vorstands, die Soziologie aus ihrem Nischendasein herauszulösen und der SGS mehr Beachtung in der Öffentlichkeit zu verschaffen.

Diese Profilierung lief so richtig vom Stapel mit dem SGS-Kongress 2003 in Zürich, der dem Thema «Triumph und Elend des Neoliberalismus» gewidmet war. Zum ersten Mal versuchten wir, am Kongress-Ende mit einer Podiumsveranstaltung die Brücke zwischen Wissenschaft und Gesellschaft bzw. zwischen Soziologie und einer weiteren Öffentlichkeit zu schaffen. In der Tat stiess das von Imhof geleitete Schlusspodium zum Thema «Genese und Entwicklung des neoliberalen Wirtschafts- und Gesellschaftsmodells in den 90er Jahren» auf breites Interesse: Der grosse Hörsaal an der Universität Zürich war zum Bersten voll, und auch das Schweizer Fernsehen war zugegen. Wir hatten prominente Podiumsteilnehmer geladen, ehemalige Politiker, die aus dem Nähkästchen plaudern sollten: Heiner Geißler, Ex-Generalsekretär der CDU in Deutschland; Oskar Lafontaine, damals Ex-Parteivorsitzender der SPD; Franz Steinegger, Ex-Präsident der FDP Schweiz; und Peter Bodenmann, Ex-Präsident der SP Schweiz. In der ihm eigenen, unnachahmlichen Art erklärte Kurt Imhof, sichtlich amüsiert vor sich hin lachend, warum wir diese Personen eingeladen hatten: Politiker seien für Soziologen Forschungsobjekte ...

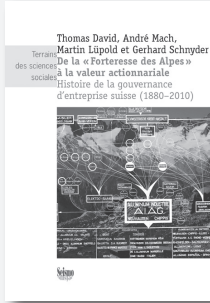
Die Brücke zur Öffentlichkeit wollte der damalige Vorstand der SGS erstens damit schlagen,

dass jeder Kongress einem einschlägigen, nicht nur wissenschaftlich sondern auch gesellschaftspolitisch brisanten Thema gewidmet wurde; zweitens aber auch damit, dass die besten Kongressbeiträge einer weiteren Öffentlichkeit zugänglich gemacht wurden in Form eines Kongressbandes, der beim SGS-eigenen SEISMO-Verlag publiziert wurde. In der Folge erschienen vier solcher Kongressbände: «Triumph und Elend des Neoliberalismus» (Imhof und Eberle 2005), «Sonderfall Schweiz» (Eberle und Imhof 2007), «Krieg!» (Maeder, Mäder und Schilliger 2009) und «Identität und Wandel der Lebensformen» (Burton-Jeangros und Maeder 2011). Heute, in der Rückblende aus dem Jahre 2015, haben sich diese Themensetzungen tatsächlich als klug gewählt und wegweisend erwiesen.

Dass die SGS anstreben soll, eine solche Themenführerschaft zu übernehmen und sowohl am Kongress als auch durch eine anschliessende Kongresspublikation eine Brücke zu einer weiteren und insbesondere auch der medialen Öffentlichkeit schlagen soll, könnten wir – in ehrendem, respektvollem und freundschaftlich-herzlichem Andenken an unseren langjährigen Vizepräsidenten – auch als Kurt Imhofs Vermächtnis an die SGS betrachten.

Thomas S. Eberle (Universität St. Gallen)

**Terrains
des
sciences
sociales**



Thomas David, André Mach, Martin Lüpold, Gerhard Schnyder

**De la «Forteresse des Alpes»
à la valeur actionnariale**
**Histoire de la gouvernance
d'entreprise suisse (1880-2010)**

504 pages, ISBN 978-2-88351-059-3, Sfr. 58.—/Euro 49.—

**Terrains
des
sciences
sociales**

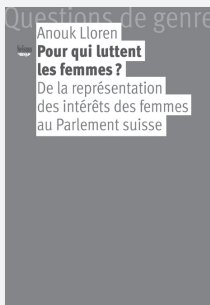


Sami Coll

Surveiller et récompenser
**Les cartes de fidélité qui nous
gouvernent**

348 pages, ISBN 978-2-88351-064-7, Sfr. 39.—/Euro 30.—

**Questions
de genre**



Anouk Lloren

Pour qui luttent les femmes?
**De la représentation des intérêts des
femmes au Parlement suisse**

188 pages, ISBN 978-2-88351-066-1, Sfr. 34.—/Euro 27.—

**Cohésion
sociale et
pluralisme
culturel**



Alexis Clotaire Némoyby Bassolé

L'agir clandestin
Agentivité de migrants ouest-africains

428 pages, ISBN 978-2-88351-063-0, Sfr. 53.—/Euro 41.—