

Friedhelm Neidhardt:

Forschungsrating der deutschen Soziologie durch den Wissenschaftsrat

(in: Soziologie, 35 Jg., Heft 3, 2006, S. 303-308)

Im November 2004 hat der Wissenschaftsrat nach Aufarbeitung der einschlägigen Erfahrungen, die in anderen Ländern schon entstanden sind, "*Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung*" verabschiedet. Wenig später wurde von ihm ein Verfahren eingeleitet, das zunächst Pilotprojekte einer flächendeckend gedachten Bewertung von zwei Disziplinen, nämlich neben der Chemie auch der Soziologie, vorsieht. Es soll in beiden Fächern ein Forschungsrating (nicht -ranking) von universitären und außeruniversitären Forschungseinrichtungen geprobt werden. Dabei sollen in den Dimensionen Forschung, Nachwuchsförderung und Wissensvermittlung nach ursprünglicher Planung zu insgesamt neun Kriterien Bewertungen ermittelt werden. Die Erfahrungen, die bei den Pilotprojekten entstehen werden, will der Wissenschaftsrat benutzen, um endgültige Entscheidungen zur Durchführung von Forschungsratings zu treffen.

Verfahren

Zur Vorbereitung und Kontrolle des Verfahrens wurde eine *Steuerungsgruppe* eingesetzt, der nicht nur Wissenschaftler aus Universitäten, sondern auch von außeruniversitären Einrichtungen (Max Planck-Gesellschaft, Leibniz-Gemeinschaft, Helmholtz-Gemeinschaft, Fraunhofer-Gesellschaft) sowie einige Vertreter aus den Bereichen der Wissenschaftsverwaltungen angehören. Der Steuerungsgruppe wurde die Aufgabe zugewiesen, zusammen mit der Geschäftsstelle des Wissenschaftsrats für die praktische Umsetzung der Empfehlungen des Wissenschaftsrats zu sorgen. Das soll durch Klärungen und Spezifizierungen dieser Empfehlungen sowie durch Einsetzung und Kontrolle fachspezifischer Bewertungsgruppen geschehen, denen für das Forschungsrating eine zentrale Rolle zugeordnet ist.

Die *Bewertungsgruppen* sollen zunächst die fachspezifische Anwendbarkeit der vom Wissenschaftsrat vorgegebenen allgemeinen Leistungskriterien prüfen und die Daten festlegen, die für die Einschätzung dieser Kriterien und der auf sie beziehbaren Indikatoren gebraucht werden. Die Daten sollen von der Geschäftsstelle, in der Soziologie beraten von ZUMA und hinsichtlich der Publikationsdaten unterstützt vom Informationszentrum Sozialwissenschaften (IZ), über die betroffenen Forschungseinrichtungen erhoben werden. "*Kernaufgabe der Bewertungsgruppe ist, nach Abschluss der Erhebungen die Daten ... zu sichten und die eigentlichen Bewertungen vorzunehmen. Jede Leistungsbewertung ist ausschließliches Produkt der Begutachtung durch die Mitglieder der Bewertungsgruppe...*" Im Übrigen sind die Bewertungsgruppen der Pilotstudie aufgefordert, "*der Steuerungsgruppe in einem Abschlussbericht ihre Bewertung der Pilotstudie mitzuteilen.*"

Von der Steuerungsgruppe wurde Friedhelm *Neidhardt* gebeten, den Vorsitz der Bewertungsgruppe zu übernehmen. Nach seiner Ernennung wurden unter Nutzung von Vorschlagslisten, die von der Deutschen Gesellschaft für Soziologie und den Wissenschaftsorganisationen der sogenannten "Allianz" eingeholt worden waren, die weiteren Mitglieder der Bewertungsgruppe berufen. Es handelt sich um Andreas *Diekmann* (Zürich), Birgit *Geissler* (Bielefeld), Jürgen *Gerhards* (FU Berlin), Ronald *Hitzler* (Dortmund), Hans *Joas* (Erfurt/ Chicago), Karl-Dieter *Keim* (Berlin), Siegfried *Lammek* (Eichstätt-Ingolstadt), Richard *Münch* (Bamberg), Werner *Rammert* (TU Berlin), Bernhard *von Rosenblatt* (Infratest München), Uwe *Schimank* (Hagen), Heike *Solga* (Göttingen), Peter *Weingart* (Bielefeld - stellvertr. Vorsitzender), Helmuth *Wiesenthal* (HU Berlin) und Reinhard *Wippler* (Utrecht). Als Vertreter der Wissenschaftlichen Kommission des Wissenschaftsrats und fachfremdes Mitglied gehört auch Peter *Lichter* (Krebsforschungszentrum Heidelberg) der Bewertungsgruppe Soziologie an. In der Geschäftsstelle des Wissenschaftsrats sind für die Belange der Pilotstudie und der Bewertungsgruppe die Herren Dr. Rainer *Lange* sowie Moritz *Mälzer* zuständig.

Zum Zeitpunkt dieser Berichterstattung konzentriert sich die Bewertungsgruppe auf die Vorbereitung eines Pretests, in den nur eine sehr kleine Zahl von Forschungseinrichtungen einbezogen wird. In der zweiten Jahreshälfte 2006 soll die Haupterhebung aller universitären und außeruniversitären Forschungseinrichtungen der Soziologie durchgeführt werden. Angestrebt ist, im Frühjahr 2007 erste Ergebnisse der Evaluationen an die Steuerungsgruppe zu vermitteln. Ob dies bei der außerordentlich umfangreichen Erhebung und im Hinblick auf die schwierigen Analyse- und Bewertungsaufgaben erreicht werden kann, bleibt abzuwarten.

Aufgaben und Motive

Soziologen kann die Art der Aufgabe, die sich mit der geplanten Recherche über das eigene Fach stellt, im Prinzip nicht fremd sein. Es geht um eine Massenerhebung, die mit einer großen Menge systematischer und pragmatischer Entscheidungen über große und kleine Fragen verbunden ist. Erforderlich ist die Bestimmung von Erhebungs- und Analyseeinheiten, die Definition und Operationalisierung von Indikatoren, die Festlegung von Skalierungen und Gewichtungen, auch die Normierung qualitativer Abwägungsregeln für die Berücksichtigung kontextspezifischer Rahmenbedingungen - und um mancherlei mehr. Dabei gilt auch für das Ratingprojekt, dass die fälligen Entscheidungen in großer Zahl nicht nach den eindeutigen Unterscheidungen von richtig versus falsch abzuleiten sind. Für die Erhebungspraxis müssen systematische Ableitungen mit Machbarkeitsgesichtspunkten vereinbart werden, und für die Gewinn- und Verlustkalkulationen alternativer Verfahren bedarf es pragmatischer Akzeptanzkriterien, über die sich endlos streiten ließe, wenn man Zeit dazu hätte und Dezsionen ausweichen will.

Der Erfolg der Bewertungsgruppe ist neben allem Sonstigen von dem Funktionieren der Geschäftsstelle des Wissenschaftsrats abhängig, die die Erhebungen durchführt, sowie von der Kooperationsbereitschaft der betroffenen Forschungseinrichtungen, die die Daten liefern müssen. Das Ziel ist, im Zusammenspiel belastbare Befunde zu erreichen, für die

nach Lage der Dinge die Fehlertoleranzen gering sind. Dies ergibt sich nicht nur aus dem Umstand, dass ein Projekt zur Bestimmung von Forschungsqualität selbstwidersprüchlich wäre, würde es den Qualitätsstandards der Forschung nicht genügen. Ein besonderer Qualitätsbedarf dieses Projekts ergibt sich vor allem aus den praktischen Zielsetzungen des Forschungsratings. Es geht um eine Massenerhebung zur Gewinnung von Einzelfallbewertungen - und zwar von Bewertungen, aus denen sich praktische Folgen, möglicherweise auch ausgesprochen unangenehme Folgen, ergeben können. Praktische Folgen sind mit umso größerer Wahrscheinlichkeit anzunehmen, als später die Ergebnisse im öffentlichen Umlauf mit dem Prestige des Wissenschaftsrats ausgezeichnet werden. Es geht hier also nicht um harmlose Forschung - auch wenn die Einzelfallbewertungen namentlich nur auf die jeweilige "Gesamtsoziologie" der universitären und außeruniversitären Forschungseinrichtungen bezogen und weder deren organisatorische Untereinheiten noch gar einzelne Personen benannt werden.

Warum sollte die Soziologie bei diesem Unternehmen mitmachen? Und was mag auch den Mitgliedern der Bewertungsgruppe gute Gründe dafür liefern, sich mit einem starkem Arbeitseinsatz daran zu beteiligen?

Zuerst einmal muss man davon ausgehen, dass auch unser Fach der sich zunehmend aufdrängenden Zumutung vergleichender Bewertungen gar nicht ausweichen kann. Flächendeckende Evaluationen, in vergleichbaren Ländern schon durchgesetzt und praktiziert, werden auch in Deutschland die Regel werden - ob wir wollen oder nicht. Und wenn das so ist, muss die Soziologie ein professionspolitisches Interesse daran haben, sich gerade den Pilotprojekten solcher Art von Evaluationen nicht zu entziehen. Im vorliegenden Fall operiert die Bewertungsgruppe zwar in einem Rahmen, der durch Entscheidungen des Wissenschaftsrats und seiner Steuerungsgruppe im Grundsätzlichen schon gesteckt ist; die Freiheitsgrade sind insoweit eingeschränkt. Aber die Bewertungsgruppe beeinflusst die Ausgestaltung der Untersuchung und den Ausweis der Ergebnisse dennoch in nicht geringem Maße. Ihr Einfluss ergibt sich vor allem daraus, dass es sich um ein Pilotprojekt handelt, bei dem es nicht nur um die Ermittlung von Urteilen, sondern auch und ganz wesentlich um die Prüfung der Bedingungen ihrer Ermittlung geht. Es besteht insoweit die Chance, dass Soziologen und Soziologinnen selber zu einer akzeptablen Standardisierung vergleichender Bewertungen beitragen.

Nur wenn dies tatsächlich gelingt, wird die Evaluationspraxis, deren lokales "trial and error" gegenwärtig für viel zu viel Unruhe in der Forschung sorgt, durch professionelle Routinen abgelöst werden können. Dass auch diese, da sie Leistungsprüfungen darstellen und für die Transparenz ihrer Ergebnisse sorgen, mancherorts sicher als lästig empfunden werden, wird nicht die Ablehnung des Verfahrens selber rechtfertigen können. Auch Wissenschaftler und Wissenschaftlerinnen können nicht selber ständig Noten verteilen und gleichzeitig meinen, es seien keine Maßstäbe dafür zu finden, dass man sie selber benotet. Alles kommt darauf an, solche Maßstäbe zu entwickeln bzw., da sie in der scientific community immer schon im Gebrauch sind, zu explizieren. Man kann allerdings nicht annehmen, dass dieser Prozess nach einem ersten Anlauf zu Ende geht. Er gehört in die Dauerreflexion auch der Soziologie.

"Indikatorenunterstütztes Expertenrating"

Bei flächendeckenden Ratings, die den Vorteil mit sich bringen, systematische Vergleiche für ihre Bewertung von Einzelfällen einsetzen zu können, sind keine "Begehungen" der zahlreichen Forschungseinrichtungen möglich, in denen soziologische Forschung betrieben wird. Solche Projekte müssen mit standardisierten Leistungsindikatoren arbeiten, die das Urteil der Gutachter orientieren können. Es ist deshalb vor Beginn der Erhebungen notwendig, solche Indikatoren systematisch abzuleiten und mit Blick auf die fälligen Datenerhebungen valide zu operationalisieren. Die Bewertungsgruppe ist zum Zeitpunkt dieser Berichterstattung noch damit beschäftigt, diesen für alles weitere wichtigen Schritt zu erledigen, bevor der Pretest durchgeführt wird.

Für das Verfahren des Wissenschaftsrats ist nun aber entscheidend, dass "*Indikatoren ... lediglich der Information der Gutachter*" dienen. Die Einschätzungen, mit denen die Forschungseinrichtungen auf den diversen Leistungsdimensionen beurteilt werden, ergeben sich aus "*indikatorengestütztem Expertenrating*" ("*informed peer review*"). Indikatoren besitzen für die Urteilsbildung der Gutachter einerseits wichtige Steuerungs- und Kontrollfunktionen, sie bedürfen andererseits aber der Validierung durch diese Gutachter, bevor sie brauchbar werden. Für sich genommen sind Indikatoren nur "Deskriptoren". Was diese Deskriptoren tatsächlich indizieren, bedarf einer Interpretation von Experten, in die auch eine Mehrzahl qualitativer Gesichtspunkte eingeht.

Die Notwendigkeit, solche qualitativen Gesichtspunkte zu berücksichtigen, führt zu einer erheblichen Arbeitsbelastung der Bewertungsgruppe. Sie muss zum Beispiel arbeitsteilig ein beachtliches Lesepensum erledigen, wenn sie Publikationsniveaus nicht nur mit standardisierten Verfahren aus den Publikationslisten von Forschungseinrichtungen bestimmen will. Darüber hinaus ist es erforderlich, für jeden Einzelfall einen Mindestsatz an Kontextfaktoren berücksichtigen zu können. Informationen, welche die Peers in anderen Fällen durch Begehung erfahren, müssen hier mit den Erhebungsbögen ersatzweise wenigstens zu wichtigen Teilfragen eingeholt werden. Dabei sind nicht zuletzt auch die Antworten auf einige offene Fragen instruktiv, mit denen die Betroffenen Selbsteinschätzungen ihrer Forschung vortragen sollen. Eine besondere Herausforderung wird darin bestehen, den Forschungseinrichtungen gerecht zu werden, deren Kontexte durch interdisziplinäre und / oder angewandte Forschung bestimmt sind. Die üblichen Evaluationsverfahren leiden mit ihrer institutionell verankerten Disziplinenorientierung unter einem akademischen Bias, da Indikatorenbildungen für interdisziplinäre und angewandte Forschung vergleichsweise schwierig sind.

Aus den genannten Bedingungen ergibt sich, dass die Qualität des Evaluationsverfahrens entscheidend von der Kompetenz und Integrität der Gutachter abhängt. Zumindest folgendes müsste im Hinblick auf diese Peers so gut wie möglich gewährleistet sein:

- Die Experten müssen fachliche *Expertise* besitzen. Das sollte durch die Zusammenstellung der Bewertungsgruppe sowie die Wahl der für Einzelfälle tätigen Berichterstatter im Prinzip gesichert werden. Ist dies in Einzelfällen nicht hinreichend gesichert, sollte es möglich sein, dass ein externes Gutachterurteil eingeholt wird.

- Die Experten müssen im Hinblick auf die von ihnen zu beurteilenden Fälle ein hinreichendes Maß an *Unbefangenheit* besitzen. Im Hinblick darauf hat die Bewertungsgruppe nach den Vorgaben des Wissenschaftsrats strenge Befangenheitsregeln beschlossen und individuelle Ausschlussentscheidungen förmlich festgehalten. Diese werden zu Nichtbefassung und Nichtanwesenheit bei Beratungen in schon jetzt bestimmten Fällen führen.

- Expertise und Unbefangenheit sollen dadurch gesteigert werden, dass alle Ratingentscheidungen nach dem Vorschlag der jeweiligen Berichterstatter am Ende von der gesamten *Bewertungsgruppe* (abzüglich jener Mitglieder, deren jeweilige Befangenheit festgestellt wurde) getroffen werden.

Umgang mit Ungewissheit

Es ist auch bei sorgfältiger Anlage und gewissenhafter Durchführung eines "indikatorengestützten Expertenratings" davon auszugehen, dass nicht für alle Forschungseinrichtungen und in jeder Hinsicht Bewertungen möglich sind, die als hinreichend qualifiziert erscheinen und ihm Hinblick auf ihre praktischen Folgen dann auch verantwortet werden können. Die Bewertungsgruppe wird sich noch damit zu befassen haben, welche Konsequenzen in solchen Fällen gezogen werden sollen. Mehrere Optionen sind verfügbar:

- Prinzipiell gibt die Bewertungsgruppe am Ende ihrer Recherche nicht nur Befunde über Forschungsleistungen, sondern auch allgemeine Urteile über Qualität und Belastbarkeit dieser Befunde ab. Dabei können Unsicherheitseinschätzungen, wo es nötig erscheint, differenziert dargestellt und Folgerungen im Hinblick auf ihre praktische Verwendung angezeigt werden.

- In Einzelfällen sollte es möglich sein, Unsicherheiten durch Bewertungsmargen auszudrücken, also zum Beispiel dadurch, dass beim Rating mit Skalen keine bestimmten Skalenwerte, sondern mehr oder weniger breite Skalenbereiche verwendet werden.

- Wenn bestimmte Fehlertoleranzen so stark überschritten werden, dass auch Bewertungsmargen nicht als hinreichender Ausdruck von Unsicherheit erscheinen, sollte die Bewertungsgruppe entscheiden können, auf Angaben zu bestimmten Kriterien und Indikatoren ganz zu verzichten.

- Vorher aber könnte noch geprüft werden, ob in Einzelfällen gezielte Nacherhebungen über die Geschäftsstelle des Wissenschaftsrats mit vertretbarem Aufwand durchgeführt werden sollen, um den wünschbaren Informationsstand signifikant zu verbessern.

Für eine zumindest nachträgliche Außenkontrolle der Evaluationsentscheidungen werden sehr wahrscheinlich auch Sekundäranalysen eingesetzt werden können. Der

Wissenschaftsrat sieht entsprechende Datennutzungen vor: *"Die Ergebnisse der Erhebung sollten für Zweitauswertungen verfügbar gemacht und nutzerfreundlich aufbereitet werden."*

Diesem Vorsatz kommt eine erhebliche professionspolitische Bedeutung zu. Das umfangreiche Material, das mit den anstehenden Erhebungen anfällt, gestattet über die jetzt durchzuführenden Evaluationen der einzelnen Forschungseinrichtungen hinaus vielfältige Analysen über den Zustand der deutschen Soziologie insgesamt - Analysen, die von der Bewertungsgruppe selber sicher nicht durchgeführt werden können. Dass sie aber durchgeführt werden, müsste auch im Interesse der Profession liegen - und die DGS sollte sich früh genug darum kümmern, dass etwas geschehen wird.