

Evaluation oder Selbstevaluation? Dilemmata im peer-review-Konzept

Helmut Kromrey

Vortrag auf der DgFE-Tagung "Evaluation in den Erziehungswissenschaften
am 4. Juli 2003 in Göttingen

Ich werde ich Folgenden nicht von *Evaluationsforschung*, sondern von Problemen der Evaluation - bzw. des *Evaluierens* - im Hochschulbereich sprechen. Es geht also einerseits um ein breiteres Thema "Evaluation" (wozu selbstverständlich die Forschung zum Zwecke des Evaluierens immer mit dazugehört). Zugleich ist mein Thema enger, es beschränkt sich auf das Modell der mehrstufigen Evaluation mit den Bestandteilen Selbstbeschreibung, peer-Evaluation und follow up, wie es sich für die Beurteilung von Studiengängen oder Instituten oder Fakultäten mittlerweile durchgesetzt hat. Dabei greife ich *fünf Fragen* auf, die in der einen oder anderen Form auftauchen, wenn von Evaluation die Rede ist, und die konkrete Antworten verlangen, damit Evaluation praktisch betrieben werden kann.

- (1) In dem Maße, in dem zunehmend fast alle von Evaluation reden, ist es zu einem regelrechten Begriffswirrwarr gekommen. Es ist also zu klären: *Worüber reden wir eigentlich?*
- (2) Personen, die als Durchführende, als Betroffene, als Beteiligte, als "Gegenstände" mit Evaluation konfrontiert sind, stellen sich die Frage nach den Anforderungen, die erfüllt sein müssen, damit die Resultate brauchbar sind; also: *Evaluation - wie geht das?*
- (3) Evaluation bereitet viel Aufwand, der oft zu Lasten anderer wichtiger Aufgaben geht. Evaluation darf also nicht lediglich Selbstzweck sein; ihre Zwecke und Ziele müssen klar - für alle Beteiligten nachvollziehbar begründet - sein. Ihre Akzeptanz setzt eine überzeugende Antwort auf die Frage voraus: *Wozu der ganze Aufwand?*
- (4) Evaluationsresultate können - insbesondere gegenwärtig in Zeiten extrem knapper Ressourcen - gravierende Konsequenzen für die Evaluierten haben. Es ist also zu fordern, dass der Prozess des Evaluierens professionellen Standards gerecht wird: *Welche Standards sind dies?*
- (5) Faktisch allerdings wird das Evaluieren überwiegend von Evaluationslaien betrieben. Da drängt sich die Frage auf: Ist das verantwortbar. Daher am Schluss ein paar Ideen: *Was kann man ändern?*

I. Worüber reden wir eigentlich?

Zunächst also ein paar notwendige Vorbemerkungen zum - mittlerweile - Allerweltsbegriff "Evaluation". Es kann ja Verschiedenstes gemeint sein, wenn von "Evaluation" gesprochen wird:

- Im allgemeinsten Sprachgebrauch ist es lediglich ein vermeintlich anspruchsvollerer Ersatz für das gute deutsche Wort "Bewertung": *Irgend jemand bewertet irgend etwas irgendwie unter irgendwelchen Gesichtspunkten.*
- In politischen Diskussionszusammenhängen ist häufig gemeint: Ein zu diesem Zweck berufenes *Expertengremium* überprüft ein Programm oder eine Einrichtung, erstellt ein

Evaluationsgutachten und formuliert Empfehlungen, ob und wie es mit dem Programm / mit der Einrichtung weiter gehen soll.

- Im Kontext der Methodologie empirischer Sozialforschung wird mit “Evaluation” eigentlich ein spezifischer *Typ angewandter Forschung* gekennzeichnet, der ein mehr oder weniger präzise definiertes Handlungsprogramm oder dessen Entwicklung zum Gegenstand hat und der im Idealfall ein experimentelles oder zumindest quasi-experimentelles Design zu realisieren trachtet. In unserem Zusammenhang - Bewertung und Qualitätsentwicklung von Studiengängen - ist dieses methodologische Leitmodell jedoch niemals anwendbar.
- Leider wird im Kontext der Sozialforschung oft auch die *gewöhnliche Umfrageforschung* mit Evaluation gleich gesetzt, nämlich dann, wenn - analog zur Meinungsforschung - evaluierende (also bewertende) Aussagen erhoben und ausgewertet werden. Und da Befragungen (insbesondere dem Methoden-Laien) *immer* anwendbar erscheinen, treffen wir überall quasi selbstverständlich auf ein Einsatz von Fragebögen.

Im mehrstufigen peer-review-Verfahren haben wir es im Kern mit der zweiten Alternative zu tun (ein zu diesem Zweck berufenes Gutachtergremium überprüft ein Programm oder eine Einrichtung - hier: ein Curriculum oder eine Fakultät). Zugleich aber wird in diesem Konzept auch die Umfrageforschung eingesetzt: z.B. für Evaluationsbefragungen bei Studierenden und/oder Absolventen.

Es gibt jedoch *nicht nur unterschiedliche Verständnisse des Begriffs “Evaluation”*. Wir haben es auch mit *unterschiedlichsten Gegenständen* oder Sachverhalten zu tun, die evaluiert werden sollen; und es existieren verschiedene Zwecke, zu denen dies geschehen soll; etwa

- Wirkungsanalyse komplexer politischer Großprogramme,
- Effektivitäts- und Effizienzkontrolle spezifischer, genau abgrenzbarer Vorhaben,
- begleitende Analyse und Bewertung eines Innovationsprozesses zur Optimierung des angestrebten Resultats und seiner Transferfähigkeit,
- Angemessenheitskontrolle des Handelns von Dienstleistungspersonal (z.B. Lehrende) sowie Qualitätskontrolle der Resultate ihres Handelns,
- Analyse der Akzeptanz von Produkt- oder Dienstleistungsangeboten durch die Kunden oder Klienten,

um nur ein paar Beispiele zu nennen.

Im Falle der peer-review-Verfahren sind Gegenstände und Zwecke der Evaluation problematischerweise nicht immer ganz klar. Eigentlich trifft von allem ein bisschen zu. Und damit stehen wir schon vor einem ziemlich gravierenden Dilemma: Eine umfassende Evaluation, die allen Gesichtspunkten gerecht werden könnte, ist weder pragmatisch (da viel zu aufwändig) noch sachlogisch (da zueinander in Konflikt stehend) realisierbar. Es müssen Schwerpunkte gesetzt werden; man muss wissen, worum es überhaupt geht; klare Zielvorgaben sind unabdingbar. Leider scheinen gerade im Hochschulbereich operationalisierbare Ziele ein besonderes Problem zu sein.

Schließlich existieren auch noch *unterschiedliche Methoden und Verfahren des Evaluierens*, und es gibt *verschiedene Instanzen* oder Personen, denen die Kompetenz zum Evaluieren zugesprochen wird. Das Spektrum reicht

- von der Beauftragung professioneller Fachexperten, die aus der Perspektive ihrer Disziplin ein Evaluationsgutachten erstellen (etwa über die zu erwartenden Beschäftigungswirkungen eines Ausbaus des Flughafens Frankfurt/Main),
- über die wissenschaftliche Erhebung empirischer Informationen als Basis für Bewertungen
- oder die Messung von Erfolgsindikatoren, deren Werte-Ausprägungen zugleich Bewertungen darstellen, bis zur schon genannten
- Erhebung, Aufbereitung und Analyse von erfragten Bewertungen aus Populationen, deren Urteilen eine besondere Relevanz für den zu evaluierenden Gegenstand zugesprochen wird (in der Marktforschung etwa die potentiellen Kunden für ein kommerzielles Produkt, in der sozialpolitischen Praxis die Klienten einer Humandienstleistung, bei einer in Planung befindlichen Infrastrukturmaßnahme die potentiellen Betroffenen oder eben die Studierenden als Teilnehmer von Lehrveranstaltungen oder die Absolventen eines Ausbildungsgangs).

Auch in dieser Hinsicht finden wir in unserem Fall der peer-reviews eigentlich wieder von allem etwas. Auch diese diffuse Situation macht es allen Beteiligten schwer: Wer hat welche Kompetenzen? Wessen Aussagen haben welchen Stellenwert. Welche Daten sind Informationen, welche sind Evaluationen?

Womit nun endlich ein bisher noch nicht angesprochenes - aber im Vortragstitel genanntes - besonderes Bonbon angetippt wäre, die Verbindung von interner und externer Evaluation; deutlicher formuliert: die *Mischung aus Fremdbewertung* (durch unabhängige und unbeteiligte Außenstehende) und *Selbstbewertung* (durch die direkt Beteiligten und Betroffenen). Ist dies ein Vorzug des Konzepts? Oder sind hier spezifische Probleme versteckt? Können - z.B. - die peers wirklich unbeteiligte Außenstehende sein? Sind die Studierenden, die Absolventen, das wissenschaftliche Personal im zu evaluierenden Fach Beteiligte oder sind sie Betroffene?

Auf diese Thematik und die damit im Zusammenhang stehende Frage der Legitimation zum Evaluieren werde ich aber erst später zu sprechen kommen. Zunächst möchte ich mich dem Thema "methodische Anforderungen an Evaluationen - und was davon bei peer-reviews nicht einlösbar ist" zuwenden.

II. Evaluation - wie geht das?

Ich hatte zu Beginn gesagt: Im Alltags-Sprachgebrauch heißt "Evaluation" nichts weiter als: *Irgend etwas wird von irgend jemandem unter irgendwelchen Gesichtspunkten in irgend einer Weise bewertet.*

Im wissenschaftlichen Sprachgebrauch ist jeder dieser Punkte zu präzisieren. In unserem Fall:

Nicht irgend etwas ist zu bewerten, sondern ein Curriculum und seine Realisierung oder ein Fachgebiet und seine Leistungen für einen Studiengang oder eine ganze Fakultät mit mehreren Fachgebieten und mehreren Studiengängen.

Evaluiert wird dieser Gegenstand auch *nicht von irgend jemandem*, sondern von gleich mehreren als besonders kompetent angesehenen Instanzen: dem wissenschaftlichen Personal als Selbstevaluatoren, den Studierenden und Absolventen als Beteiligten und Betroffenen, einer externen Evaluationskommission als Fremdevaluatoren.

Beurteilt werden soll auch *nicht unter irgendwelchen* (gruppenspezifisch beliebigen) *Gesichtspunkten*, sondern nach explizit auf den Sachverhalt bezogenen und begründeten Kriterien und Standards. Grundlage sollen sein: explizierte Ziele des Curriculums und verbindlich vereinbartes Organisationsleitbild bei der Selbstevaluation durch das Personal; Grad der Förderung bzw. Behinderung des Lernens bei der Betroffenenevaluation durch die Studierenden; Passung der vermittelten Ausbildung für Anforderungen der Berufspraxis bei der Beurteilung durch Absolventen; Erreichen übergreifender fachlicher Standards unter Berücksichtigung der Besonderheiten des jeweiligen Falles beim peer-review.

Und natürlich soll dies *nicht irgendwie* / nicht in irgend einer Weise geschehen, sondern in einem objektivierten Verfahren und im Rahmen eines im Detail geplanten Evaluationsdesigns. So existieren mittlerweile inhaltlich weitgehend übereinstimmende Leitfäden für die Selbstevaluation sowie Empfehlungen darüber, wie der Prozess des Erstellens des Selbstberichts zu organisieren ist. Ähnliche Anleitungen sind für die Arbeit der peers entwickelt worden. Ebenso gibt es standardisierte Musterfragebögen für Studierende und Absolventen. All dies ist Resultat vielfältiger Erfahrungen mit Konzepten mehrstufiger Evaluation im Ausland, von HRK- und Wissenschaftsrats-initiierten Pilotprojekten und darauf fußenden Empfehlungen sowie von mittlerweile langjähriger Praxis mehrerer Evaluationsagenturen in Deutschland.

Unter *methodologischen* Gesichtspunkten betrachtet scheint das Konzept also außerordentlich gut abgesichert zu sein, besser als manche Großprojekte der Programmevaluation. Ziehen wir zur Kontrolle die Qualitätsstandards der Deutschen Gesellschaft für Evaluation heran¹ und gehen die dort genannten Kriterien durch, dann fällt das Ergebnis ebenfalls positiv aus: Identifizierung der Beteiligten und Betroffenen, Klärung der Evaluationszwecke, Glaubwürdigkeit und Kompetenz der Evaluatoren, Auswahl und Umfang der Informationen, Transparenz von Werten, Vollständigkeit und Klarheit der Berichterstattung, Rechtzeitigkeit, Nutzung der Evaluation, angemessene Verfahren und so weiter und so weiter. Alles das ist in den Verfahrensvorschriften und Leitfäden *berücksichtigt* – aber ist es in der Praxis auch *erreichbar*?

Wo liegt das - im Untertitel meines Vortrags angesprochene - Dilemma? Wir finden es gleich auf drei Dimensionen:

- *Erstens*. Die unterstellten Ziele und Organisationsleitbilder als Bezugs-Eckpunkte für die Selbstevaluation existieren entweder gar nicht oder allenfalls in Form von Leerformeln, also nicht in einer auch nur näherungsweise operationalisierbaren Formulierung. Daher ist ihr Herausarbeiten eine der Aufgaben des Selbstevaluationsprozesses. Wenn sie aber erst in diesem Prozess entstehen, können sie fairerweise nicht zur Basis von Bewertungen für eine Situation herangezogen werden, die ohne die Möglichkeit der Orientierung an ihnen zustande gekommen ist. Zwar kann es im Hinblick auf den Verwertungszweck der Evaluation (Organisationsentwicklung und Qualitätssicherung) durchaus als positiv

¹ Deutsche Gesellschaft für Evaluation e.V. (Hg.). 2002: Standards für Evaluation, Köln: DeGEval

angesehen werden, dass (eventuell erstmals) überhaupt Zieldiskussionen geführt werden. Aber der Evaluation selbst fehlt in diesem Fall ihr Fundament, d.h. die Selbstevaluation reduziert sich auf Selbstbeschreibung.

- *Zweitens*. Die Kriterien, die zur Beurteilung tatsächlich angewendet werden, dürften von Gruppe zu Gruppe verschiedene sein und sind (da sie nicht expliziert und dokumentiert werden) nicht rekonstruierbar. Worauf bezieht das wissenschaftliche Personal seine Urteile, worauf die einzelnen Studierenden, worauf die Absolventen? Welche Vergleichsmaßstäbe, welche Qualitätsstandards wenden sie an? Diese Unbestimmbarkeit gilt ähnlich auch für die externen peers. In solchen Fachdisziplinen, für die ein Konsens über generelle Zielorientierungen nicht existiert (das dürfte für die meisten gelten), müssen an dessen Stelle die Vorstellungen der jeweiligen peer-Evaluatoren treten; und die dürften je nach Zusammensetzung der Kommissionen variieren. Damit wäre dann doch ein wichtiger DeGEval-Standard für Evaluationen verletzt: der der Fairness der Beurteilungen.
- *Drittens*. Das geschilderte Evaluationskonzept ist außerordentlich umfassend und entspricht vom Aufwand durchaus dem eines kleinen Forschungsprojekts. Spezielles Personal aber existiert - in Form der Evaluationsagenturen - nur für das Projektmanagement. Die übrigen Beteiligten müssen ihre Leistungen zusätzlich zu ihren sonstigen Verpflichtungen erbringen – und dies innerhalb einer vorgegebenen relativ kurzen Frist. Das Problem ist auch nicht durch Zurückschrauben des Anspruchsniveaus lösbar; denn: Angesichts zu erwartender Konsequenzen der Resultate der Evaluation wäre es zumindest für das wissenschaftliche Personal der evaluierten Einrichtung eine riskante Strategie, den Aufwand dadurch zu minimieren, dass die Selbstevaluation nur oberflächlich betrieben wird. Was soll dann zugunsten der Evaluation in diesem Zeitraum zurückstehen? Vielleicht die Qualität der Lehre?

III. Wozu der ganze Aufwand?

Damit komme ich nun zu einem weiteren zentralen Punkt des Konzeptes: Die *Evaluation darf nicht "Selbstzweck" sein*. Dies wäre eine Verschwendung knapper Ressourcen. Sondern sie muss erkennbar und wirksam in einen Prozess der Organisationsentwicklung und Qualitätsentwicklung eingebettet sein.² Und der Aufwand muss sich für alle Beteiligten lohnen. Geht das? Durch die Einbettung in einen Prozess der Organisationsentwicklung wird ja *einerseits* der Aufwand noch einmal vergrößert. *Zum anderen* ist für die Beteiligten zu Beginn eines Entwicklungsprozesses nicht ohne weiteres vorhersehbar, wer zu den Gewinnern und wer zu den Verlierern gehören wird. Um dennoch für möglichst viele eine aktive Mitwirkung als Ergebnis einer rationalen Entscheidung zu ermöglichen, ist ein Höchstmaß an Transparenz unbedingte Voraussetzung. Es muss präzise nachvollziehbar sein, was tatsächlich der Gegenstand der Evaluation ist, auf welche Teilaspekte der Schwerpunkt gelegt wird und wie die Resultate verwendet werden. Es muss verbindliche Klarheit bestehen über die Entwicklungs-

² Hier ist also ein Unterschied zu einer "reinen" Evaluationsforschung zu konstatieren, bei der es "nur" darum geht, gegenstands- bzw. programmbezogene Daten über Maßnahmendurchführung und Veränderung von Zielindikatoren zu messen sowie diese Veränderungen den Maßnahmen "zuzurechnen", d.h. Indikatorendaten in evaluationsrelevante Informationen zu transformieren. Evaluation als Bestandteil von Qualitätsentwicklung bedeutet, dass das Konzept um (mindestens) den weiteren Schritt der "Umsetzung" in Praxis zu verlängern ist.

ziele, über die Rahmenbedingungen und die Risiken. Ohne solche Transparenz ist ein Fehlschlag des gesamten Vorhabens vorprogrammiert.³

Die *Forderung* “*Es muss sich lohnen, Arbeit in das Vorhaben zu investieren*” gilt natürlich auch für die *Mitglieder der externen Evaluationskommission*. Nur wenn die Mitarbeit in peer-Kommissionen nicht den Status einer lästigen und nach Möglichkeit abzuwehrenden Pflicht hat, kann mit Ergebnissen gerechnet werden, die professionellen Standards genügen. Und nur wenn die Resultate professionellen Standards genügen, ist es auch verantwortbar, auf sie Entscheidungen zu stützen, die für die vom Vorhaben Betroffenen mit zum Teil weitreichenden Konsequenzen verbunden sein können. Damit aber die Ergebnisse professionellen Standards genügen können, ist von den für das Evaluationsprojekt Verantwortlichen (insbesondere den Mitgliedern der peer-Kommission) professionelle Kompetenz zu fordern. Im herkömmlichen Konzept der Evaluation als Programmwirkungsforschung wird von den Evaluatoren lediglich diejenige Methodenkompetenz gefordert, die auch für jede andere Forschung notwendig ist. Die eigentliche Evaluation des Erfolgs oder Misserfolgs eines Programms geschieht ja über die empirische Erhebung von Zielerreichungsindikatoren und deren Zurechnung zu den getroffenen Maßnahmen. Das ist zwar keine einfache Aufgabe, aber sie ist doch unter Rückgriff auf bewährte empirische Instrumente und Verfahren objektivierbar und anhand der traditionellen Gütekriterien der Forschung überprüfbar.

Peer-Evaluatoren können und dürfen sich darauf allein nicht zurückziehen. Wie schon angesprochen, ist der Gegenstand der Beurteilung viel zu komplex, um ihn auch nur annähernd durch Indikatoren operationalisieren zu können (sie sind natürlich eine der Grundlagen der Bewertung und daher Bestandteil der Selbstbeschreibung). Zudem fehlt es zum Teil an klaren und operationalisierbaren Zielen, über deren Erreichung die Indikatoren Auskunft geben könnten. Vieles ist also bei der Vor-Ort-Begehung zu ergänzen, ist in Gesprächen mit den Beteiligten zu konkretisieren. Durch gemessene Indikatorwerte objektivierbare Beurteilungen sind durch subjektive Einschätzungen der peers abzurunden, gegebenenfalls auch zu korrigieren. Und aus all dem sind Empfehlungen abzuleiten, Alternativen aufzuzeigen und deren Erfolgswahrscheinlichkeiten zu prognostizieren. Hierfür gibt es kein in Empirie-Lehrbüchern dokumentierbares Methoden-Arsenal. Auch bei den vorhin genannten Leitfäden für die Selbstevaluation und Selbstbeschreibung oder für die peer-Gutachter handelt es sich lediglich um orientierende Vorgehensregeln, nicht um “harte Methoden”; es sind allenfalls Sammlungen von “Kunstregeln”, die im Rahmen des Ermessens der Selbstevaluatoren und der Gutachter berücksichtigt werden sollen.

IV. Standards für peer-Evaluationen?

Wenn also an eine Objektivierung der Ergebnisse durch Methoden und Verfahren nicht ernsthaft zu denken ist, dann kann die Lösung nur in einem generell zu realisierenden Prinzip “Professionalität” liegen, d.h. dann muss an die Stelle objektivierbarer Methoden/Verfahren das Vertrauen in die Professionalität und Unabhängigkeit der Evaluatoren treten.

³ Die hochschulpolitische Großwetterlage ist dafür zur Zeit allerdings nicht sonderlich günstig. Denn wie soll beispielsweise eine Hochschulleitung diese Anforderung für ein Curriculum oder für eine Fakultät erfüllen können, wenn nicht einmal auf der Ebene der Gesamtuniversität oder gar der Hochschulpolitik eines Bundeslandes Planungssicherheit gegeben ist?

Eine zentrale Frage, die sich hierbei stellt, lautet: Handelt es sich bei den peers wirklich um unvoreingenommene Außenstehende, die bei ihrem Urteil nicht von persönlich-subjektiven Interessen gelenkt werden? Oder haben wir es beim peer-review nicht doch eher mit einer weiteren Selbstevaluations-Komponente zu tun (wenn auch auf "höherem" Niveau: Fachvertreter evaluieren im Interesse ihres eigenen Faches)? Sie kennen das Alltagsurteil: Eine Krähe hackt der anderen nicht die Augen aus. Voraussetzung für die Akzeptanz der peer-Urteile ist daher, dass die aufgeworfene Frage eindeutig und begründbar zugunsten von "unabhängige Fremdevaluation" beantwortbar ist. Das bedeutet, dass den Gutachtern und ihrer Auswahl für das Gelingen des Vorhabens und für die Akzeptanz seiner Resultate eine zentrale Bedeutung zukommt.

Entsprechend werden in der Diskussion um das peer-review-Verfahren hohe Anforderungen an die Mitglieder der Evaluationskommission gestellt. Ein paar Zitate:⁴

- Die Gutachter sollen angesehene Vertreter ihres Faches sein.
- Sie müssen sowohl unabhängig als auch unbefangen urteilen können.
- Sie dürfen in den evaluierten Fakultäten keinerlei Entscheidungsbefugnisse haben und dürfen auch nicht von den Ergebnissen einer Evaluation betroffen sein.
- Die Gutachter müssen eine nachfragende und skeptische Haltung einnehmen.
- Sie benötigen Kraft zur Analyse und zu einem klaren Urteil.
- Sie sollen über eine persönliche Autorität verfügen und Präsenz demonstrieren, die mit der Fähigkeit verbunden wird, effektiv im Gutachterteam mitzuwirken.
- Notwendig ist die Fähigkeit, angemessene Beurteilungen auch im Umfeld außerhalb der eigenen Kerndisziplin zu fällen.
- Peers sollen darüber hinaus Erfahrungen in Organisation und Management mitbringen.
- Zu fordern sind ein hohes Niveau in mündlicher und schriftlicher Kommunikation sowie Erfahrungen in der termingerechten Abfassung schriftlicher Berichte auf der Basis formaler Vorgaben und ein gutes zeitökonomisches Management.

Auch wenn es vielleicht so klingen mag -: Die hier angesprochene Problematik ist nicht besonders spezifisch für Evaluationen, d.h. sie besteht nicht etwa in besonders erschwerender Weise für den Bereich gesellschaftlich relevanter Evaluation. In einem erheblich relevanteren Sektor der Gesellschaft, nämlich dem der Rechtsprechung, war schon seit eh und je ein ähnlich gelagertes Problem zu lösen. Auch hier können Urteile selten auf unbestreitbare Tatsachefeststellungen gegründet werden. Auch hier müssen - zwar regelgeleitete, aber doch subjektive - Abwägungen von mit der Kompetenz des Urteil-Fällens betrauten Personen ("Richter") Lücken schließen. Die Lösung besteht in diesem Fall darin, dass deren subjektiven Urteilen aufgrund ihrer Professionalität Vertrauen entgegengebracht werden soll und kann.

⁴ Zusammengestellt nach *Hermann Reuke*, Vortrag auf der DeGEval-Jahrestagung 2000 im Forum 2 "Gutachter auswählen und qualifizieren" (<http://www.degeval.de>), sowie *Stifterverband für die Deutsche Wissenschaft* (Hg.): Positionen, Mai 2003 (<http://www.stifterverband.de>)

Wie steht es jedoch mit der Professionalität bzw. mit der Möglichkeit zur Professionalisierung im Bereich der Hochschulevaluation durch peers?

Dass erfahrene Evaluatoren in ausreichender Zahl für die zahlreichen peer-review-Kommissionen zur Auswahl stehen, entspricht bekanntermaßen nicht der Alltagssituation. Erst recht ist es undenkbar, dass die zur Mitwirkung bereiten Personen zugleich auch noch die meisten der o.g. Zusatzqualifikationen aufweisen.

Die benötigten Erfahrungen könnten zwar durch learning on the job erworben werden, sofern die peers wiederholt zur Verfügung stünden und die Kommissionen hinsichtlich des Grades an Evaluationserfahrung gemischt zusammengesetzt werden könnten. Über ein solches Modell müssen wir jedoch - wie Sie mir sicher zustimmen werden - unter den derzeit gegebenen Bedingungen gar nicht erst nachdenken. Jeder "angesehene Vertreter seines Fachs" wird sich vielleicht der Mitwirkung an einem Verfahren nicht entziehen können, bei weiteren Anfragen jedoch mit Hinweis auf diese bereits geleistete akademische Pflicht dankend abwinken. Lehre, Betreuung von Examensarbeiten, Prüfungen, Selbstverwaltung verlangen angesichts knapper gewordener Ressourcen in den Universitäten ohnehin schon einen zunehmenden Anteil am Zeitbudget, der viel zu oft zu Lasten der Forschung geht.

Eine Alternative wäre ein fundiertes Evaluationstraining für die externen Gutachter. Dieses könnte zwar nicht die Erfahrung ersetzen, die durch mehrfache Teilnahme zu gewinnen ist, könnte aber doch dazu beitragen, dass die wirksam werdenden Standards zwischen verschiedenen Gutachtergruppen nicht zu stark variieren (immerhin wäre dies ein wichtiger Beitrag zur Fairness des Verfahrens). Verschiedentlich wird eine solche Gutachterschulung explizit gefordert. So heißt es z.B. in den Empfehlungen einer im Auftrag des Stifterverbandes für die Deutsche Wissenschaft durchgeführten Meta-Evaluation der peer-review-Praxis von Nordverbund und Zentraler Evaluationsagentur Hannover:

"Für die Phase der externen Evaluation wird es als notwendig angesehen, die Gutachter/-innen ausführlich auf die Evaluation vorzubereiten. Eventuell sollten dafür anderthalb bis zwei Tage vorgesehen werden."⁵

Ist dies realisierbar? Ganz abgesehen davon, dass nach meinem Eindruck aus Gesprächen mit Kollegen ein solches Evaluationstraining die Assoziation von Interviewerschulung hervorzurufen und auf wenig Akzeptanz zu stoßen scheint, würde eine so ausführliche Vorbereitung die geforderte Vor-Ort-Zeit verdoppeln (für die Begehung sind üblicherweise ca. 1,5 Tage vorgesehen; jetzt kämen noch einmal 1,5 bis 2 Tage hinzu). Dies scheint als Anforderung an die Evaluatoren wenig realitätsnah. Und selbst wenn die Bereitschaft dazu bestünde, würde der Plan vermutlich an der Schwierigkeit scheitern, dafür einen Zeitraum zu finden, der in den Terminkalendern aller Mitglieder der Kommission untergebracht werden kann.

Mir scheint daher die einzig realisierbare Lösung für die *Deckung des Bedarfs an Evaluatoren* in einer grundlegenden Veränderung der Rahmenbedingungen für die Mitwirkung zu liegen. Es handelt sich ja hierbei schließlich nicht um einen Bedarf in einer nur kleinen Nische des Wissenschaftsalltags. Vielmehr steigt der Bedarf z.Z. geradezu rapide dadurch, dass in unüberschaubar großer Zahl Akkreditierungsverfahren für neu konzipierte und/oder schon implementierte Bachelor- und Master-Studiengänge anstehen, die ebenfalls als wesentlichen

⁵ Stifterverband: Positionen Mai 2003, a.a.O., Empfehlung VII.6)

Bestandteil das peer-review-Gutachten enthalten. Und wenn die Akkreditierungen - was ja diskutiert wird - auch noch in regelmäßigen Abständen zu wiederholen sind, wird der Bedarf an professionell arbeitenden Evaluatoren dauerhaft bestehen bleiben.

V. Was kann man ändern?

Ich will hier zum Abschluss nicht versuchen, umfassende Überlegungen vorzutragen und darüber einen Grundsatzstreit vom Zaun zu brechen. Aber ein paar Ideen möchte ich doch skizzieren.

Eine Veränderung der Rahmenbedingungen hätte zunächst einmal darin zu bestehen, dass die Mitwirkung an peer-review-Kommissionen - sei es in Evaluations-, sei es in Akkreditierungsverfahren - als ein *wichtiger Beitrag zur Entwicklung der Qualität von Lehre und Studium* anerkannt wird, der nicht in Form einer Freizeit-Hobby-Tätigkeit ausgeübt werden kann. Evaluation (sei sie bezogen auf Lehre und Studium, wie hier thematisiert, sei es Forschungsevaluation etwa für die DFG) sollte als gleichwertige *wissenschaftliche* Leistung neben herkömmlicher Forschung und als einer der Pfeiler des Spektrums *dienstlicher Pflichten* von Professoren neben Lehre, Forschung und Selbstverwaltung gelten.

Wenn aber die Mitwirkung an peer-Evaluationen ein bedeutsamer Beitrag für die Qualität von Lehre und Studium ist, dann müssen Evaluatoren *für die zeitliche Belastung* im laufenden Semester eine *gleichwertige Entlastung* erhalten, z.B. durch entsprechende Verringerung des Lehrdeputats im jeweiligen Semester.

Über den Ausgleich der mit Evaluation verbundenen Zusatzbelastung hinaus muss sich die Mitwirkung aber auch für die *Evaluatoren als Wissenschaftler* lohnen: beispielsweise könnte die Mitgliedschaften in peer-Kommissionen als Leistungsindikator in Modellen leistungsorientierter Mittelverteilung (analog zur Einwerbung von Drittmitteln) berücksichtigt werden.

Die Mitwirkung sollte sich natürlich nicht nur für den einzelnen Wissenschaftler lohnen, sondern auch für die Hochschulen, die Evaluatoren (durch Gewährung von Lehrdeputatsverringerung) "freistellen": Warum sollte nicht die Stellung von Gutachtern in Hochschulrankings in ähnlicher Weise als *Indikator des Renommees der Hochschulen* zu Buche schlagen, wie dies für Wahrnehmung anderer "ehrevoller" Aufgaben der Fall ist?

Die von mir skizzierten Ideen zur Änderung der Rahmenbedingungen für eine bessere Hochschulevaluation haben in Zeiten knapper Ressourcen - wie ich meine - den Charme, dass sie kostenneutral wären. Falls auch Sie Gefallen an ihnen finden sollten, helfen Sie bitte dabei mit, sie zu verbreiten. Ich verspreche: Ich werde keine Lizenzgebühren verlangen!