

Qualität und Evaluation im System Hochschule

In: Stockmann, Reinhard (Hrsg.): Evaluationsforschung,
Opladen: Leske+Budrich (2. Aufl. 2003, S. 233-258)

1. "Evaluation" - Begriff und Funktionen

Was ist Evaluation? Was ist ihr Nutzen? Wie soll man es machen? Fragen dieser Art sind für das System Hochschule nicht nur deshalb schwer zu beantworten, weil die zu evaluierenden „Gegenstände“ so vielfältig, die mit Evaluation verfolgten Ziele so widersprüchlich und die für diesen Zweck eingesetzten Methoden und Verfahren nicht unstrittig sind. Schon der Begriff *Evaluation* selbst ist zu einem schillernden Allerweltswort geworden, mit dem je nach Kontext sehr Unterschiedliches verbunden wird. Sieht man einmal von der alltagssprachlichen Verwendung als „wohlklingendes“ Fremdwort für jede Form von Bewertung¹ ab, so findet sich die am wenigsten spezifische Verwendung in der politischen Diskussion. Hier bedeutet Evaluation die *Überprüfung und Beurteilung eines Programms oder einer Einrichtung*. Speziell mit dieser Aufgabe betraute Experten formulieren auf der Basis von Informationen, die zu diesem Zwecke gesammelt wurden, ein Evaluationsgutachten. Dieses kann gravierende Konsequenzen für die Zukunft der überprüften Programme oder Einrichtungen haben bis hin zu deren Einstellung oder "Abwicklung".

Von *Evaluation* wird häufig aber auch im Kontext "ganz gewöhnlicher" Umfrageforschung gesprochen. Gemeint ist dann – in Analogie zur Meinungsforschung – die *Erhebung, Aufbereitung und Auswertung bewertender* (also evaluativer) *Aussagen*. In diesem Sinne ist die Fragebogenerhebung der Urteile von Vorlesungsteilnehmerinnen und –teilnehmern über die von ihnen besuchte Veranstaltung eine "Evaluation der Lehre", ist das Sammeln der Auffassungen von Professorinnen und Professoren über das Image verschiedener Universitäten eine "Evaluation der Hochschulen" und ist die Befragung von Personalverantwortlichen in Verwaltung, Wirtschaft und Verbänden über ihre Vorlieben bei der Einstellung akademisch ausgebildeter Mitarbeiter eine „Evaluation der Qualifikation der Absolventen“.

In der Methodologie empirischer Sozialforschung schließlich bezeichnet Evaluation das *Design für einen spezifischen Forschungstyp*, durchgeführt im idealtypischen Fall als Feldexperiment mit Kontrollgruppen. Wo die Voraussetzungen dafür nicht in vollem Maße erfüllt sind – und das ist überwiegend der Fall –, behilft man sich mit "Ersatzlösungen" für diejenigen Designkomponenten, die nicht idealtypisch realisiert werden können: So tritt etwa das "matching"-Verfahren zur Konstruktion strukturäquivalenter Experimental- und Kontrollgruppen an die Stelle der häufig nicht möglichen Zuordnung durch Randomisierung oder ersetzt die statistische Kontrolle von "Störgrößen" deren nicht realisierbare Abschirmung in der Erhebungssituation. Charakteristisch an solchen quasi-experimentellen Anordnungen ist, dass die generelle Orientierung an der Vorgehens- und Argumentationslogik des Experiments erhalten bleibt. Je nach Erkenntnis- und Verwertungsinteresse kann so verfahrenende Evaluation unterschiedliche Schwerpunkte setzen: Sie kann sich als *Begleitforschung* auf den Implementationsprozess eines Programms beziehen oder als *Wirkungsforschung* auf die späteren Konsequenzen; ihre Ergebnisse können *formativ* direkt in den Prozess rückgekoppelt werden, oder sie können *summativ* im Nachhinein einen Gesamtüberblick vermitteln.

¹ "Alltagsevaluationen" können wie folgt charakterisiert werden: Irgendwer bewertet irgend etwas irgendwie unter irgend welchen Gesichtspunkten.

Gemeinsam bleibt bei allen Ausdifferenzierungen des Vorgehens der unmittelbare Bezug zu einem *Programm* (einem Bündel von Maßnahmen zur Erreichung definierter Ziele). Die eigentliche *Evaluation* besteht hier nicht in der Formulierung normativ wertender Urteile, sondern in “technologischen” (Vergleichs-)Aussagen, etwa: Ist das Programm so implementiert worden wie geplant? Hat es die gesetzten Ziele erreicht? Welche Maßnahmen waren besonders effektiv, welche ineffektiv? Welche ungeplanten (erwünschten oder unerwünschten) Nebenwirkungen traten auf?

Die dargestellten Begriffsverwendungen schließen eine zusätzliche und manchmal Verwirrung stiftende Dimension des Begriffs ein: Als Evaluation wird einerseits ein bestimmtes *Tun oder Handeln* bezeichnet: das *Evaluieren*, andererseits aber auch das *Resultat dieses Tuns*: die formulierte bewertende Aussage, das *Evaluationsurteil*. Beides hängt untrennbar zusammen: Damit das *Evaluationsurteil* nicht rein subjektiv bleibt, sondern den Anspruch intersubjektiver Geltung erheben kann, muss das *Evaluieren* nach objektivierbaren Regeln geschehen, muss also einer nachprüfbaren Evaluationsmethodik folgen. So existieren beispielsweise relativ genaue Verfahrensvorschläge für die Gewinnung entscheidungsrelevanter Informationen und für das Erstellen von Evaluationsgutachten durch Sachverständige im Hochschulbereich (vgl. etwa Wissenschaftsrat 1996). Ähnliches gilt für das Evaluieren per Befragung: Zunächst liefert die Methodenlehre empirischer Sozialforschung die Kriterien für die Gestaltung des Prozesses der Datenerhebung; ergänzend formulieren Anhänger dieser Evaluationsstrategie einige Axiome, um von den erhobenen *subjektiven Werturteilen* der je einzelnen Befragten zu *intersubjektiven Evaluationen* zu gelangen.²

Eine andere Perspektive auf die Evaluation und ihre Methodik eröffnet sich durch die Frage, *zu welchem Zweck* evaluiert werden soll. Nach einem Vorschlag von Eleanor Chelmsky (1997, 100 ff.) sind bei grober Einteilung drei *Funktionen von Evaluation* zu unterscheiden, denen methodisch drei “conceptual frameworks” entsprechen:

- Evaluation zur Verbreiterung der Wissensbasis (im folgenden als “Forschungsparadigma” der Evaluation bezeichnet),
- Evaluation zu Kontrollzwecken (im folgenden das “Kontrollparadigma”) und
- Evaluation zu Entwicklungszwecken (im folgenden das “Entwicklungsparadigma”).

Das “*Forschungsparadigma*” der Evaluation:

Insbesondere für Universitätswissenschaftler gelten Evaluationsprojekte als Chance, neben dem “eigentlichen” Evaluationszweck grundlagenwissenschaftliche Ziele zu verfolgen. Evaluation gilt in dieser Perspektive als angewandte Forschung, die sich mit der Wirksamkeit von sozialen Interventionen befaßt. Ihr kommt die Rolle eines Bindeglieds zwischen Theorie

² Für den Fall studentischer Lehrbewertungen wurden z.B. die folgenden - allerdings bestreitbaren - Axiome formuliert: (1) Urteile (Schätzungen) von Studenten über die Lehre sind verlässlicher als Urteile der Dozenten über die Leistung der Studenten und ebenso verlässlich wie Urteile von Kollegen über die Lehre. (2) Sie legen praktisch dieselben Kriterien für gute Lehre an wie die Dozenten selbst. (3) Sie sind - wenn man etwa 20-30 Studenten urteilen lässt - zuverlässig wie professionelle Testverfahren. (4) Sie sind von anderen Merkmalen der Studenten selbst und der Dozenten wenig beeinflusst (vgl. Schmidt 1980, 51-52). Bisher konnte nur die Geltung des zweiten Axioms belegt werden, wogegen sich die für die Qualität als intersubjektive Urteile entscheidenden Axiome 3 und 4 leider als empirisch falsch erwiesen haben (vgl. Kromrey 1994 und 1995b).

und Praxis zu (Weiss 1974, 11). Insbesondere staatliche Auftragsforschung eröffnet einen Weg, Zugang zu den internen Strukturen und Prozessen des politisch-administrativen Systems zu erhalten (Wollmann / Hellstern 1977, 456). Alle Anlässe, Aktionsprogramme zur Bewältigung sozialer Probleme zu implementieren, alle Situationskonstellationen, in denen durch neue gesetzliche Regelungen wichtige Randbedingungen geändert werden, alle Bemühungen, technische, organisatorische oder soziale Innovationen zu entwickeln, werfen zugleich sozialwissenschaftlich interessante Fragestellungen auf. Und im Unterschied zu forschungsproduzierten Daten zeichnen sich Untersuchungen unmittelbar im sozialen Feld durch einen ansonsten kaum erreichbaren Grad an externer Validität aus. Evaluationsforschung in diesem Sinne wird in erster Linie als Wirkungsforschung, die *Evaluation* selbst als wertneutrale technologische Aussage verstanden, die aus dem Vergleich von beobachteten Veränderungen mit den vom Programm angestrebten Effekten (den Programmzielen) besteht. Evaluatoren, die sich dem Forschungsparadigma verpflichtet fühlen, werden versuchen, wissenschaftlichen Gütekriterien so weit wie möglich Geltung zu verschaffen und Designs zu realisieren, die methodisch unstrittige Zurechnungen von Effekten zu Programm-Elementen durch Kontrolle der relevanten Randbedingungen erlauben.

Das “*Kontrollparadigma*” der Evaluation:

Im Unterschied zur Wirkungsforschung versteht sich der zweite Typus von Evaluation als Beitrag zur Planungsrationale durch Erfolgskontrolle des Programmhandelns. Planung, verstanden als Instrument zielgerichteten Handelns, um einen definierten Zweck zu erreichen, muß sich bestimmten Erfolgskriterien (Effektivität, Effizienz, Akzeptanz) unterwerfen. Evaluationen dieser Art werden argumentativ vertreten als eine weitere Kontrollform administrativen Handelns neben Rechtmäßigkeits-Kontrolle (Gerichte), politischer Kontrolle (Parlamente) und Wirtschaftlichkeits-Kontrolle (Rechnungshöfe). Eine schon früh vorgeschlagene, charakteristische Definition lautet: "Der Begriff Erfolgskontrolle impliziert ex-post-Kontrolle von Ausführung und Auswirkung von zu einem früheren Zeitpunkt geplanten Maßnahmen, und Erfolgskontrolle ist immer zugleich Problemanalyse für den nächsten Planungszyklus" (Hübener / Halberstadt 1976, 15). In welcher Weise der Erfolg kontrolliert wird und an welchen Kriterien der Erfolg gemessen wird, ob die Evaluation ihren Schwerpunkt auf output oder outcome des Programms legt oder auf dessen Implementation, hängt ab vom Informationsbedarf der programmdurchführenden und/oder der finanzierenden Instanz.

Das “*Entwicklungsparadigma*” der Evaluation:

Grundsätzlich anders gelagert sind Problemstellung und Erkenntnisinteresse bei diesem dritten Typus von Evaluationen. Am Beginn steht nicht ein bereits realisiertes oder in der Implementationsphase befindliches oder zumindest ausformuliertes Programm; vielmehr geht es darum, Konzepte und Vorstellungen zu entwickeln, die Fähigkeit von Organisationen zur Problemwahrnehmung und -bewältigung zu stärken, beim Strukturieren von Politikfeldern beratend und unterstützend mitzuwirken. So verstandene Evaluation ist im wörtlichen Sinne "formativ", also programmbeeinflussend. Sie ist wesentlicher Bestandteil des Gestaltungsprozesses, in welchem ihr die Funktion der Qualitätsentwicklung und Qualitätssicherung zukommt. Gelegentlich wird diese Konstellation auch als “offene” Evaluation bezeichnet, im Unterschied zu den zuvor geschilderten “geschlossenen” Evaluationen, in denen Problem-

und Fragestellungen, methodisches Vorgehen, Bewertungskriterien und die Zielgruppen der Evaluationsberichte von vornherein feststehen.

2. Evaluation im System Hochschule - wie und wozu?

Im Kontext der Qualitätsdiskussion im Bereich Hochschule schwingen - wenn Evaluationen eingefordert werden - fast immer die Erkenntnisinteressen aus allen drei „frameworks“ gleichzeitig mit. Natürlich möchte man neues empirisch abgesichertes Wissen darüber gewinnen, wovon erfolgreiches Lehren und Studieren abhängt und wie der Erfolg gefördert (wenn schon nicht garantiert) werden kann - insofern ist das Forschungsparadigma gefragt. Natürlich sollen zugleich Effektivität und Effizienz der Verwendung der in den Hochschulbereich fließenden öffentlichen Mittel kontrolliert werden, sollen die Hochschulen Rechenschaft über ihr Tun ablegen - also ist auch das Kontrollparadigma angesprochen. Und ebenso natürlich soll Evaluation dabei helfen, geeignete Maßnahmen zur Verbesserung der Qualität von Lehre und Studium zu konzipieren, zu implementieren und zu testen - womit schließlich das Entwicklungsparadigma zu seinem Recht kommt.

Trotz solcher unrealistisch hoher Erwartungshaltung wird dann aber nicht selten zugleich in aller Naivität gefordert, Evaluation müsse sich schnell, einfach, mit geringem Kosten- und Arbeitsaufwand realisieren lassen - denn Personalressourcen und Geld sind in den Hochschulen bekanntermaßen außerordentlich knapp. Außerdem darf die Evaluation den laufenden Betrieb nicht „stören“ - schließlich ist das eigentliche Ziel der Hochschule die Sicherstellung eines geregelten Angebots für ein ordnungsgemäßes Studium und nicht dessen Evaluation. Übersehen wird bei solchen Rufen nach simplen und belastungsfreien Verfahren in aller Regel die außerordentliche Komplexität des Gegenstands von Evaluationen im System Hochschule: Es existiert weder ein präzise beschreibbares „Programm“ mit klar definierten Zielen und ihnen zugeordneten Maßnahmen sowie eindeutig festgelegten Zielerreichungskriterien³ noch ein konkretes „Produkt“, dessen Qualität mit einem Satz von Qualitätsindikatoren durch standardisierte Messverfahren abgebildet werden kann⁴ (dazu später mehr). Zudem muss die Evaluation unmittelbar im aktiven Feld durchgeführt werden und kann - anders als etwa bei politischen Pilotprojekten - nicht einen Teilbereich abgrenzen und (zumindest teilweise) von der Umwelt isolieren. Nicht zuletzt ist sie dabei mit zahlreichen Akteuren mit je unterschiedlichen Zielen und Vorstellungen konfrontiert, deren Handeln sämtlich über Erfolg und Misserfolg des zu evaluierenden Programms wie auch der Evaluation selbst mitentscheidet.

³ Ein „Studiengang“ zeichnet sich selbst im Falle extrem starker „Verregelung“ durch rigide Prüfungs- und Studienordnungen dadurch aus, daß er nur als formaler Rahmen existiert, der in jedem Semester aufs neue empirisch interpretiert und „in die Welt gesetzt“ werden muß: durch die jeweiligen Lehrangebote und Prüfungsthemen, durch das jeweilige Lehr-, Beratungs-, Betreuungs- und Prüfungsverhalten des Lehrpersonals, durch die jeweils zum Studieren bereitgestellte Infrastruktur des Instituts oder Fachbereichs und natürlich durch das jeweilige Lern- und Arbeitsverhalten der Studierenden.

⁴ Ein Produkt „Qualifizierung der Studierenden“ ist zwar *allgemein* mit Hilfe von Leerformeln beschreibbar, aber nur *individuell* bei jedem einzelnen Studierenden zu konkretisieren - und zudem auch noch in allererster Linie durch die Qualität und die Eigenschaften des „Rohmaterials“ determiniert, nämlich durch die individuell eingebrachten Vorkenntnisse und Fertigkeiten sowie das individuelle Studierverhalten.

Will Evaluation im System Hochschule dieser Komplexität gerecht werden, ist sie extrem zeit- und ressourcenaufwendig. Soll sie nicht lediglich Selbstzweck sein, sondern Veränderungen (Qualitätsverbesserungen) in Gang setzen, ist sie trotz des mit ihr verbundenen Aufwands bei allen Beteiligten auf *aktive Akzeptanz* - auf Mitwirkungsbereitschaft - angewiesen. Soll diese erwartbar sein, muss sich der Aufwand lohnen: Die Evaluation muss für die Beteiligten einen erkennbaren Nutzen bringen; andernfalls würde jeder von ihr angestoßene Prozess schnell wieder zum Stillstand kommen. Akzeptanz ist darüber hinaus aber auch eine wesentliche Voraussetzung dafür, dass Evaluation überhaupt gültige - und somit verwertbare - Ergebnisse liefern kann. Es ist daher vorab zu klären und für alle Beteiligten erkennbar zu machen, zu welchem Zweck evaluiert werden soll, was mit den zu erhebenden und auszuwertenden Daten geschehen soll. Evaluation darf nicht als „Evaluations-Ritual“ erscheinen. Nicht selten geschieht allerdings genau dies: Es wird evaluiert, um eine an die Universität gestellte Forderung nach Evaluation zu erfüllen. Dass in einem solchen Fall möglichst wenig Aufwand getrieben wird und lediglich leicht zugängliche Informationen gesammelt werden, dass zudem vor allem „unproblematische“ Daten präsentiert werden (etwa in Lehrberichten, die nicht zum Gegenstand fakultäts-/fachbereichs-interner Diskussion über Lehre und Studium werden), darf nicht verwundern.

Eine geringe Akzeptanz ist auch dann zu erwarten, wenn Evaluation lediglich als *Kontrollinstrument* verwendet werden soll, um – seien die zu Bewertenden nun Lehrpersonen oder ganze Fächer – die „Guten“ von den „Schlechten“ zu sonders und daran Sanktionen zu knüpfen. Sofern eine solche Evaluation nicht zu umgehen ist, haben die Evaluierten unendlich viele Möglichkeiten, kritische Informationen zu verschleiern und positive Informationen überdimensioniert in den Vordergrund zu rücken. Als Argument für die Verpflichtung zu regelmäßigen Kontrollen wird u.a. vorgetragen, Evaluation sei der Preis, den die Hochschule und ihre Angehörigen für größere Autonomie (etwa in Form von Globalhaushalten) zu zahlen habe. An die Stelle abnehmender staatlicher Kontrolle müsse zunehmende interne Kontrolle und öffentliche Berichterstattung über die effiziente Verwendung der zufließenden Steuermittel treten. Daxner kritisiert diese Tendenz als einen „Weg in die Rechtfertigungsgesellschaft“ (1999, 41 ff.). Der Wissenschaftsrat erkennt zwar durchaus ein „berechtigtes öffentliches Interesse“ an Erhöhung der Transparenz, Rechenschaftslegung und Kontrolle an, sieht aber zugleich einen Zielkonflikt zum (ebenfalls öffentlichen) Interesse an „Qualitätsentwicklung und -sicherung im Lehrbetrieb“ sowie der „Verbesserung von Organisationsformen und -strukturen“ (Küchler 1996, 6-8); denn ein erfolgversprechendes Verfolgen des letztgenannten Ziels setzt „eine schonungslose Selbstkritik und Selbstanalyse der Fachbereiche“ voraus: „Hier kann die öffentliche Dimension der Evaluation eher zu einer Verschleierung als zu einer Behebung bestehender Defizite führen“ (a.a.O., 7).⁵

Weniger strittig als die (öffentliche) Kontrollfunktion der Evaluation ist ihr Einsatz als *hochschulinternes Steuerungsinstrument*, zum Teil verknüpft mit „incentives“ z.B. für gute Lehrorganisation und Forschungsleistungen. In manchen Bundesländern wird ein Teil der

⁵ Konsequenterweise räumt der Wissenschaftsrat in diesem Zielkonflikt der Qualitätsentwicklung Priorität ein und empfiehlt ein Evaluationsverfahren in Eigenverantwortung der Hochschulen mit den Zielsetzungen: „Verbesserung der Transparenz in Studium und Lehre; Förderung der institutionellen Verantwortung der Fachbereiche für die Lehre; Herausbildung von Qualitätssicherungsstrategien in Studium und Lehre ...“; aber auch: „Verstärkung des Wettbewerbs im deutschen Hochschulsystem sowie der Profilbildung an Hochschulen und Fachbereichen nicht nur in der Forschung, sondern auch in der Lehre“ (Küchler 1996, 9).

universitären Sachmittel „nach Leistungs- und Belastungskriterien“ vergeben. Zu diesem Zweck ist – soll dies in der Universität routinemäßig und flächendeckend geschehen – ein Raster von möglichst wenigen Indikatoren zu entwickeln, die regelmäßig verfügbar sind und möglichst objektiviert Leistungen und Belastungen eines Fachs oder auch von kleineren Einheiten abbilden: etwa Zahl der Studienanfänger und Betreuungsrelation Studierende / Lehrende (als Belastungsindikatoren), Zahl der Zwischenprüfungen, Examensarbeiten, Studienabschlüsse, Studiendauer der Absolventen, Promotionen, Studienabbrüche etc. (als Leistungsindikatoren). Methodisch sind solche Verfahren nicht unstrittig: Indikatoren können immer nur einen Ausschnitt aus dem gesamten Problemfeld abbilden und auch dies immer nur mit zweifelhafter Gültigkeit. Des weiteren besteht – wenn an die Ausprägung der verwendeten Indikatoren die Verteilung von Mitteln geknüpft wird – die Möglichkeit (und damit die Gefahr), lediglich die durch Indikatoren abgebildeten Bereiche zu „optimieren“ und anderes zu vernachlässigen; ganz abgesehen von der Möglichkeit der Umdefinition von Kriterien, um „bessere“ Ergebnisse zu erzielen (mehr Studienabschlüsse in kürzerer Zeit kann man auch dadurch erreichen, dass man das Anspruchsniveau senkt). Indikatorensysteme – das ist aus den Erfahrungen der Sozialindikator-Bewegung bekannt – funktionieren nur so lange, wie sie lediglich zu Deskriptions- und Erklärungszwecken (allenfalls auch noch als prognostisches Frühwarnsystem) genutzt werden, so lange also an die Indikatorenwerte keine Sanktionen für diejenigen geknüpft werden, die die Ausprägungen durch ihr Handeln beeinflussen können.

Einen etwas anderen Zungenschlag erfährt die Diskussion um Evaluation als Steuerungsinstrument im Kontext der Forderung nach stärkerer *Wettbewerbsorientierung der Hochschulen*: „Auch ein Hochschulsystem, das staatlich globalgesteuert, aber zunehmend von Wettbewerb und Profilbildung gekennzeichnet ist, muss sich Marktgesetzmäßigkeiten stärker öffnen. Auch wenn sie nicht auf Gewinnerzielung hin orientiert sind, müssen Hochschulen sich in mancher Hinsicht wie Unternehmen verhalten lernen. Das heißt unter anderem, bei der Planung und Ausgestaltung von Lehrangeboten rascher auf Nachfrageänderungen zu reagieren und auch Studierende als ‚Kunden‘ ernster zu nehmen“ (Landfried 1999, 10). Evaluation schafft in diesem Zusammenhang „ein Stück Markt-Ersatz, eine Art Quasi-Wettbewerb“ (ders., 11). Mit wem die Hochschulen über das Medium Evaluation in welcher Form um welche knappen Ressourcen konkurrieren, bleibt allerdings ebenso unbeantwortet wie die Frage, auf wessen Nachfrageänderungen – und dann in welcher Weise – rascher zu reagieren sei. Auch die „Kundenrolle“ von Studierenden bleibt diffus.

An der Schnittstelle von Kontroll- und Wettbewerbsargumentation finden wir die Vorstellung von *Evaluation als Instrument globaler Qualitäts-„Messung“*. Wenn es gelänge, die Qualität der Leistungen der Institution Hochschule und ihrer Gliederungen umfassend, detailliert, gültig und zuverlässig zu messen, dann stünde damit einerseits ein „objektives“ Kontrollinstrument zur Verfügung; andererseits existierte in Gestalt der Qualitätsmaße auch eine Art „Währung“, die für einen funktionierenden Wettbewerb (etwa um Reputation, aber auch um öffentliche Finanzmittel, um Forschungsförderung, sogar um besonders leistungswillige Studierende) notwendig scheint. Die wiederholt unternommenen Versuche, „Rankings“ von Hochschulen, Hochschulfächern bis hin zu Lehrveranstaltungen zu erstellen, sind u.a. auch als Bemühung zu verstehen, Transparenz auf einem solchen Wettbewerbsmarkt zu schaffen.⁶

⁶ „Vorreiter“ der Hochschulrankings in Deutschland war der SPIEGEL mit einer 1989 veröffentlichten ersten Rangreihe, der 1993 eine zweite folgte („Welche Uni ist die beste?“). „Gerankt“ wurde im Hinblick auf „Service-Qualität“; Maßstab für die Qualitätsaussagen waren studentische Urteile (also

Beziehen wir das Argument auf den Sektor Lehre und Studium, so gilt zwar einigermaßen unbestritten als letztlisches Kriterium für die Leistungsqualität der Hochschule der positive Effekt bei den Adressaten, also der „Qualifizierungserfolg“ bei den Studierenden. Doch ist hier die unerschütterliche Annahme weit verbreitet, dass gute Servicequalität bereits eine weitgehende Gewähr für solchen Erfolg sei. Somit gehört es zu den ersten Aufgaben der Evaluation, die qualitätsrelevanten Dimensionen des Serviceangebots zu bestimmen und zu deren Beurteilung Qualitätsindikatoren zu begründen und zu operationalisieren. So wird die Evaluation gleich zu Beginn mit einem zentralen theoretischen und methodologischen Problem konfrontiert: der Unbestimmtheit des Begriffs „Qualität“. Zwar kann an dieser Stelle nicht im Detail hierauf eingegangen werden⁷ – die Sozialwissenschaft befasst sich im Rahmen der Sozialindikatorenbewegung seit Jahrzehnten damit; für das Gesundheitswesen hat Donabedian ein differenziertes und weit verbreitetes Qualitätskonzept entwickelt (ausführlich Donabedian 1980). Das Fazit jedenfalls ist, dass Qualität angesichts der Vieldimensionalität dieses Konstrukts und seiner unterschiedlichen Bedeutung in unterschiedlichen Kontexten und für unterschiedliche Zielgruppen nicht in einer Weise empirisch abbildbar ist, wie es das wissenschaftliche Konzept „Messen“ verlangt.

Dennoch gilt gemeinhin als unbestrittenes Ziel von Evaluation, einen *Beitrag zur Verbesserung von Qualität* zu leisten, hier also: zur Qualität von Studium und Lehre. Kann Qualität schon nicht „gemessen“ werden, so kann von der Empirie doch gefordert werden, diejenigen Informationen zur Verfügung zu stellen, die es erlauben, die Güte der zu evaluierenden Sachverhalte aus unterschiedlichen Perspektiven einzuschätzen. Für das Ziel Qualitätsentwicklung und/oder Qualitätssicherung ist allerdings allein mit dem Bereitstellen solcher Informationen durch die Evaluation noch nicht viel gewonnen - denn Qualität entwickelt sich nicht von selbst. Informationen sind allenfalls die notwendige (aber noch nicht hinreichende) Voraussetzung dafür, gezielte Veränderungen dort in Gang zu setzen, wo der evaluierte Sachverhalt verbesserungsbedürftig und verbesserungsfähig erscheint. Selbst wenn alle methodischen und organisatorischen Fragen der Evaluation geklärt sein sollten (s. dazu den folgenden Abschnitt), bleibt daher noch eine wesentliche Entscheidung zu treffen: Wer ist *Träger des Qualitätsentwicklungs-Vorhabens*? Nur in Ausnahmefällen wird dies auch der *Träger des Evaluationsvorhabens* sein (= „interne Evaluation“). Wo Qualitätsentwicklungs-Akteure und Evaluationsinstanz sich unterscheiden (= „externe Evaluation“), ist ein auf gegenseitigem Vertrauen basierendes Verhältnis beider Instanzen die Voraussetzung sowohl für gültige Evaluationsresultate (keine Unterdrückung „problematischer“ Informationen, zuverlässige „Schwachstellen“-Analyse) wie für gelingende Umsetzung der Resultate in Maßnahmen zur Qualitätsentwicklung (Formulierung *konstruktiver* und *realisierbarer* Empfehlungen durch die Evaluation, Zusicherung der Vertraulichkeit erlangter interner Kenntnisse, Veröffentlichung nur im gegenseitigen Einvernehmen).⁸

„Kundenzufriedenheit“). Mittlerweile erstellt das Centrum für Hochschulentwicklung (CHE) gemeinsam mit der Stiftung Warentest „Studienführer“, die neben den subjektiven Urteilen von Studierenden und Hochschullehrern auch „harte“ Fakten aufbereiten (für eine Darstellung der SPIEGEL-Rankings s. Hornbostel/Daniel 1995, zum Studienführer s. Hornbostel 1999, für eine erste Bilanzierung auf der Basis von Sekundäranalysen s. Müller-Böling u.a. (2001); eine vergleichende Analyse von Hochschulrangreihen liefert Klostermeier 1994, kritisch Kreuz u.a. 1992); generell zum Thema „Qualität und Wettbewerb s. den Tagungsbericht DAAD/HRK 2002.

⁷ Der Abschnitt 4 greift das Thema noch einmal auf.

⁸ Ein Beispiel für die Verknüpfung von Evaluation und Organisationsentwicklung hat die Hamburger Hochschule für Wirtschaft und Politik (HWP) vorgestellt (Künzel/Nickel/Zechlin 1999).

3. Welches „Paradigma“ der Evaluation ist das beste?

Gegen Schluß des Abschnitts 1 wurden drei „Paradigmen“ für die Evaluation skizziert; und es stellt sich nach dem Abriß der Funktionen und Zwecke von Evaluationen im vorigen Abschnitt die Frage, ob eines der drei Paradigmen als besonders geeignet charakterisiert und daher als „Königsweg der Hochschulevaluation“ empfohlen werden könnte. Bei diesen Überlegungen ist zu berücksichtigen, daß wir es bei dem zu evaluierenden Leistungsspektrum von Hochschulen nicht mit einem konkret faßbaren „Gegenstand“ zu tun haben - anders als im Falle der Güterproduktion, wo sich Effizienz und Effektivität des Produktionsprozesses sowie die Qualität des Produkts (output) relativ leicht beurteilen und in standardisierter Form messen lassen. Hier geht es vielmehr um die Bereitstellung von Dienstleistungen, bzw. noch eingeschränkter: von Humandienstleistungen, die die aktive Mitwirkung der Klienten (hier: der Studierenden) voraussetzen, sollen sie einen „Erfolg“ bewirken, also „Qualität“ haben, effizient und effektiv sein. Was ist in diesem Fall das „Produkt“? Was ist der „Produktionsprozeß“? Ist es das Vorhalten einer Dienstleistungs-Infrastruktur (geregelter Lehrangebot und die Informationen darüber in kommentierten Vorlesungsverzeichnissen, Personal für Beratungen und Prüfungen, PC-Räume und Bibliotheken) oder die einzelne Dienstleistung selbst (die Lehrveranstaltung, Prüfung, das Beratungsgespräch)? Oder interessiert nicht eher, was durch die vorgehaltene und realisierte Dienstleistung bewirkt wird (outcome anstelle von output)? Schließlich: Wenn es sich - wie hier - um eine Dienstleistung handelt, die auf die Akzeptanz und das aktive Mitwirken der Adressaten angewiesen ist: Wer oder was ist dann eigentlich zu evaluieren - der Anbieter, der Nachfrager oder beide? Und nicht zuletzt: Wer evaluiert wen?

Das „*Entwicklungsparadigma*“ ist von seiner ganzen Konzeption her auf Innovationsprozesse zugeschnitten. Anzuwenden wäre es also für die Qualitätsentwicklung in einzelnen Lehrveranstaltungen (für Beispiele s. Knäuper/Kroeger 1999, Peter/Wawrzinek 1995, Kromrey 1996) oder bei der Reform eines Curriculums oder bei Vorhaben der Organisationsentwicklung. Hierfür können offene Evaluationsansätze geeignete Konzepte bereitstellen, etwas das Konzept der responsiven Evaluation (Beywl 1991) oder das einer nutzenfokussierten Evaluation (Patton 1997, Beywl/Joas 2000). Solche Strategien sind jedoch nicht gemeint, wenn die regelmäßige und (in angemessenen Zeitabständen) flächendeckende Evaluation von Lehre und Studium eingefordert wird. Diese Forderung bezieht sich gerade nicht auf wissenschaftlich kontrollierte Innovationsprozesse, sondern auf den Regelbetrieb der Hochschule und ihrer Einrichtungen, u.a. auch um anhand der Ergebnisse einschätzen zu können, ob Reformmaßnahmen und Veränderungen notwendig erscheinen.

Das „*Kontrollparadigma*“ in seiner evaluationsmethodisch ausgearbeiteten Form als Erfolgskontrolle (vgl. etwa Eekhoff/Muthmann/Sievert 1977) setzt klar formulierte und operationalisierbare Ziele und (im allgemeinen output-orientierte und quantifizierbare) Zielerreichungskriterien sowie ihre Zurechenbarkeit zum Handlungsvollzug der zu evaluierenden Einrichtung voraus. Die obigen Ausführungen dürften jedoch gezeigt haben, daß diese Voraussetzungen für das Handlungsfeld Humandienstleistungen im System Hochschule nicht gegeben sind. Der von der Politik gewählte Ausweg ist die Ersetzung von Erfolgskontrolle durch ein formalisiertes Berichtswesen: Fast alle Bundesländer verlangen mittlerweile die regelmäßige Erstellung von Lehrberichten in Ergänzung der schon länger

üblichen Forschungsberichte.⁹ Solche Lehrberichte sind zwar eine Sammlung beschreibender Informationen, die das System Hochschule transparenter und somit auch für Außenstehende einschätzbarer (also „evaluierbarer“) machen, sind aber selbst keine Evaluationen.¹⁰

Noch anspruchsvoller in den Voraussetzungen für seine Anwendbarkeit ist das „*Forschungsparadigma*“, sofern es nach dem Standardmodell der Programmforschung unter Rückgriff auf quasi-experimentelle Designs (s.o., Abschnitt 1) verfahren soll. Dass dieses Modell in der Hochschule flächendeckend nicht eingesetzt werden kann, ist offensichtlich. Die Universität kann – sowohl aus grundsätzlichen wie aus pragmatischen Gründen – nicht zum Experimentallabor umfunktioniert werden, nur um dadurch evaluierbar zu sein.¹¹ Aber auch aus weiteren methodologischen Gründen eignet sich diese Form der Evaluierung nicht.

So könnte beispielsweise das zu bewertende *Programm der Diplomstudiengang eines Fachs* sein. Als Ziele kämen die an die Studierenden zu vermittelnden Qualifikationen, als Maßnahmen Studienordnung, Studienverlaufspläne, Lehrveranstaltungen, Studieninfrastruktur sowie Betreuung und Beratung durch das Lehrpersonal, außerdem Prüfungsordnung, Prüfungen und andere Leistungskontrollen in Betracht. Für die Messung der Zielerreichung böte sich der Zeitpunkt der Beendigung des Studiums (Diplomprüfung oder Studienabbruch) bei den einzelnen Studierenden an.

Ein erstes Bündel von Problemen ergäbe sich hier bereits bei der empirischen Beschreibung der für die Studierenden bis zum Examen relevant gewordenen Maßnahmen. Studienordnung, Studienverlaufspläne und Prüfungsordnung wären für alle im Verlaufe ihres Studiums konstant und somit (im Hinblick auf *Unterschiede* in den erworbenen Qualifikationen) ohne Wirkung. Lehrveranstaltungen dagegen – mit Ausnahme einiger standardisierter Vorlesungen und Übungen insbesondere im Grundstudium – sind auch von ihren Inhalten häufig so stark variierend, dass zusätzlich zu den im Prinzip statistisch kontrollierbaren Unterschieden der studentischen Veranstaltungsauswahl (feststellbar etwa durch Auswertung der Studienbücher) eine zusätzliche Variation in nicht kontrollierbarem Ausmaß hinzukäme. Beratung, Betreuung und Prüfungen schließlich ergeben sich in Interaktionen zwischen einzelnen Studierenden und einzelnen Mitgliedern des Lehrpersonals und wären bei Studienabschluss überhaupt nicht mehr rekonstruierbar.

Als ähnlich problematisch erwiese sich die Erfolgsmessung. Die im Studium zu vermittelnden Qualifikationen sind üblicherweise in den Studiengangsdokumenten (Studien- und Prüfungsordnung) nur sehr vage – falls überhaupt – definiert. Ersatzweise kämen die in Klausuren und

⁹ Einen Auflistung gesetzlicher Grundlagen zu Evaluation und Lehrberichterstattung findet sich bei el Hage 1996, 27 ff. Die Dokumentation gibt auch einen instruktiven Überblick über staatlich initiierte und/oder geförderte Projekte zur Lehrevaluation seit den 1990er Jahren (13 ff.), informiert über Methodenfragen und Instrumentenentwicklung und stellt einige der verwendeten Instrumente studentischer Veranstaltungskritik vor.

¹⁰ Obwohl durch Landesgesetze vorgeschrieben und ihrem Wesen nach nicht Evaluation, sondern Dokumentation, können Lehrberichte dennoch über ihre öffentliche Transparenzschaffungsfunktion hinaus auch hochschulintern von Nutzen sein, als “Instrument des Controlling” und als “Mittel einer von der Hochschule selbst gesteuerten Studienreformarbeit” (Habel 1995, 12); und sie können natürlich auch Evaluationen *enthalten*: Selbstbeurteilungen der Fachbereiche und Institute, Ergebnisse studentischer Befragungen. Als “Handreichung für die Praxis” finden sich bei Habel (a.a.O., 14 ff.) Leitfäden für Interviews und Informationserhebungen für Lehrberichte.

¹¹ Allenfalls Lehrexperimente *neben* dem laufenden Lehrbetrieb könnten in dieser Weise evaluiert werden; hier wären z.B. für die Vermittlung des herkömmlichen Stoffs Doppelangebote bei Realisierung unterschiedlicher Lernsettings denkbar.

Prüfungen erbrachten Leistungen der Absolventen (gemessen in den erzielten Noten) in Betracht. Diese wären allerdings keine direkten Maße der Qualifikationen, sondern lediglich Indikatoren für eine Teilmenge von ihnen. Erfolge/Misserfolge auf anderen Dimensionen blieben unerkannt. Außerdem wäre zu fragen, wie es um die Gültigkeit dieser Indikatoren bestellt wäre, wenn die Träger des zu evaluierenden Programms die Indikatoreausprägungen selbst festlegten (nämlich in Prüfungen und Klausurbenotungen).

Ganz unmöglich schließlich wäre die Zurechnung der Beiträge einzelner Maßnahmen zum festgestellten Studienerfolg der jeweiligen Absolventen. In welcher Weise das Studium verläuft sowie ob und in welchem Ausmaß es erfolgreich abgeschlossen wird, hängt nach allen vorliegenden empirischen Erkenntnissen aus der Bildungsforschung in hohem Maße von Merkmalen in der Individualsphäre der Studierenden ab: wie Lebenssituation, Interesse und Leistungsmotivation, Studienstil und -intensität. Die von den Trägern des Studiengangs beeinflussbaren Gegebenheiten – Studieninfrastruktur, Lehre und Betreuung – können lediglich (wenn sie von schlechter Qualität sind) das Studium erschweren oder (bei guter Qualität) erleichtern; den individuellen Erfolg *bewirken* können sie nicht. Um also den relativen (fördernden oder hemmenden) Beitrag der angebotenen Maßnahmen zum Studienerfolg abschätzen zu können, müsste zunächst der individuelle Eigenbeitrag des jeweiligen Studierenden bekannt sein – eine, wie leicht einsehbar, völlig unrealistische Anforderung, deren Nichterfüllbarkeit in diesem Bereich jede Evaluation im Sinne von Zielerreichungskontrolle prinzipiell unmöglich macht.

4. Der Ausweg: Evaluation durch Umfrageforschung

Wenn Evaluation nach dem Modell der Programmforschung nicht möglich ist, liegt es nahe, das Fällen von Urteilen – also die *Tätigkeit des Evaluierens* – auf dafür geeignet erscheinende Dritte zu verlagern (auf Experten, auf Kunden, auf Betroffene) und die Funktion der Forschung auf das systematische Einholen und Auswerten solcher “Fremd-Evaluationen” zu beschränken. Dies wird in der Tat überall dort so gehandhabt, wo Lehrevaluation betrieben wird.

In besonders systematischer, formalisierter und nachprüfbarer Form geschieht dies in dem *Verfahren der zweistufigen (internen und externen) Evaluation*, wie es im Verbund Norddeutscher Universitäten und von der Zentralen Evaluationsagentur (ZEVA) in Niedersachsen angewendet wird und das auf Empfehlungen der Hochschulrektorenkonferenz (1995) und des Wissenschaftsrates (1996) beruht. Die so durchgeführte Evaluation erfasst als Gegenstand die Organisation und Durchführung der Lehre und des Studiums innerhalb einer Hochschuleinheit (Fakultät/Fachbereich oder Institut) und hat explizit nicht die Bewertung einzelner Lehrveranstaltungen zum Ziel. Die Hauptelemente des Verfahrens sind (s. HRK 1998a):

- Der Lehrbericht eines Fachbereichs / einer Fakultät als kontinuierliche Sammlung von Basisdaten und Leistungsindikatoren.
- Die interne Evaluation (...), die von einer internen Arbeitsgruppe vorbereitet wird und auf der Analyse der in den Lehrberichten erfassten Daten und auf Interviews mit Studenten und Personal basiert. Sie führt zu einem kritisch-abwägenden Bericht über die Selbsteinschätzung der erreichten Resultate im Hinblick auf die selbstgesteckten Ziele; sie enthält eine Beschreibung möglicher Hindernisse und Defizite sowie von Maßnahmen zu ihrer

Beseitigung, Vorschläge für die Kontrolle und Verbesserung der Qualität der Lehre und die Verteilung von Mitteln für Forschung und Lehre. (...)

- Der Vor-Ort-Besuch der Sachverständigen (Peers), der von der betreffenden Agentur vorbereitet wird, welche die Selbstbewertungsberichte an die Mitglieder der Peer-Group weiterleitet und, falls nötig, den zu evaluierenden Fachbereich um weitere Informationen bittet. Der in der Regel zweitägige Vor-Ort-Besuch schließt Gespräche mit der Universitätsleitung, dem Dekan und den Lehrenden und Studierenden ein, (...).
- Der Evaluations-Bericht der Peers schließt eine kritische Würdigung der internen Evaluation und ihrer tatsächlichen Bedeutung als Mittel der Qualitätssicherung ein, weist auf Probleme hin und gibt Hinweise auf mögliche Lösungen. Vor der Veröffentlichung des Abschlussberichts erhält der evaluierte Fachbereich Gelegenheit, den vorläufigen Bericht zu bearbeiten, um Irrtümer und Missverständnisse zu korrigieren. Dies findet im Rahmen einer gemeinsamen Sitzung statt, an der die Mitglieder der Sachverständigengruppe (Peer-Group), Vertreter der evaluierten Einrichtung und der Evaluationsagentur teilnehmen. (...)
- Das „follow up“ umfasst eine Vereinbarung bzw. einen Vertrag zwischen dem Fachbereich und der Universitätsleitung über zu ergreifende Maßnahmen zur Verbesserung von Lehre und Studium, zur Optimierung der Ergebnisse bzw. zur Sicherstellung bestimmter zu erreichender Standards innerhalb eines definierten Zeitraums. (...)“ (a.a.O., S. 11 f.)¹².

Die Evaluierung geschieht in diesem Modell – wie ersichtlich – nicht *durch* die Umfrageforschung, wohl aber (unter anderem) *mit* Umfragen, und wird ergänzt um andere Erhebungen sowie um Daten aus der Hochschulstatistik und um Beobachtung und Diskussion. Für die Evaluation dieses Typs erfüllt die empirische Forschung und deren Methodik nicht die Funktion einer Instanz der Qualitätsentscheidung mittels „objektiver“ Daten. Vielmehr finden wir hier ein Beispiel für das Prinzip der „Objektivierung durch Verfahren“. Die Sicherung der Intersubjektivität der Ergebnisse wird durch ein darauf zugeschnittenes Verfahrensmodell angestrebt: Die Einbeziehung aller Beteiligten und Betroffenen in den Prozess soll gewährleisten, dass das für den Zweck der Evaluation relevante Informationsspektrum erfasst wird. Die Gültigkeit der Ergebnisse, wie sie der Evaluationsbericht dokumentiert, wird durch die Möglichkeit zur Korrektur sowie durch eine gemeinsame Abschlussdiskussion zwischen Evaluatoren und Evaluierten angestrebt (kommunikative Validierung). Damit die Evaluation nicht ins Leere läuft, sondern Anstöße zu Qualitätsverbesserungen gibt, mündet das Verfahren in konkrete Zielvereinbarungen (Festlegung nachprüfbarer Maßnahmen mit expliziten Terminen für die Realisierung). Und um es nicht bei einem einmaligen Anstoß bewenden zu lassen, sondern einen Prozess kontinuierlicher Qualitätsverbesserung in Gang zu setzen, sind schließlich in regelmäßigen Abständen (von mehreren Jahren) „follow ups“ vorgesehen.¹³

¹² Im selben Heft findet sich auf S. 15 f. ein Muster zur Gliederung des Lehrberichts sowie auf S. 17 f. ein Vorschlag für eine Gliederung des Evaluationsberichts. „Crucial Points“ des Verfahren diskutiert Bülow-Schramm (1995).

¹³ Die Realisierung dieses Modells im „Verbund norddeutscher Hochschulen“ wird ausführlich dargestellt und diskutiert von Fischer-Blum (1995) und Schierholz/Vocke (1997). – Im Auftrag der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung führt die Hochschulrektorenkonferenz seit 1998 ein „Projekt Qualitätssicherung“ durch, das u.a. den Auftrag hat, die Weiterentwicklung von Standards der Evaluation zu unterstützen und auf gemeinsame Standards hinzuwirken, Informationen für Hochschulen bereitzustellen sowie zwischen Hochschulen und Öffentlichkeit zu transportieren. Eines der Instrumente ist die Unterstützung und Durchführung von Tagungen und Workshops an verschiedenen Hochschulstandorten sowie die Veröffentlichung der Tagungsbände in der Reihe „Beiträge zur Hochschulpolitik“ (s. Schreier 1998). Auf einem „Nationalen Expertenseminar“ wurde am 29.5.1998 ein umfassender Überblick über den Stand der Evaluation in Deutschland gegeben (HRK 1998b). Die

Es ist leicht nachvollziehbar, dass ein solches Evaluations- und Qualitätssicherungsmodell mit einem hohen Aufwand an Kosten, Zeit und Personal verbunden ist.

Wesentlich weniger anspruchsvoll – sowohl hinsichtlich des Verfahrens als auch hinsichtlich des Bemühens um Objektivierung – ist demgegenüber die weit verbreitete Strategie, die Lehrevaluation allein auf die *Befragung Studierender* zu stützen. Dem liegt offenbar die Gleichsetzung von Betroffen-Sein mit Expertentum zugrunde. Für die Beurteilung der Qualität der Lehre etwa lässt sich die folgende einfache (und auf den ersten Blick auch durchaus plausibel erscheinende) Argumentation rekonstruieren: „Ein aufwendiges Verfahren der Qualitätsbeurteilung durch Evaluationsforschung ist entbehrlich. Mit den Studierenden verfügt die Hochschule bereits über *die* Experten, die die Lehre aus erster Hand – als tagtäglich von ihr Betroffene – fundiert und zuverlässig beurteilen können. Deren Wahrnehmungen und Bewertungen brauchen nur in standardisierter Form erhoben und pro Lehrveranstaltung in geeigneter Form ausgewertet zu werden, um aussagekräftige Qualitätsindikatoren zu erhalten“ (s. oben). Manche Lehrende gehen noch einen Schritt weiter und vertreten unter Verweis auf „jahrzehntelang bewährte Praxis in den USA“¹⁴ die Auffassung, hierzu werde nicht einmal ein detailliertes Instrumentarium benötigt. Vielmehr reichten kurze und damit schnell ausfüllbare Fragebögen aus, in denen von den Studierenden auf wenigen zentralen Dimensionen (typischerweise Didaktik, Angemessenheit von Stoffmenge und Schwierigkeitsgrad, Auftreten der Lehrperson und soziales Klima, Lernerfolgseinschätzung) zusammenfassende Bewertungen erbeten werden. Studierende seien durchaus kompetent, solche Urteile zu fällen, wird – vermeintlich studentenfreundlich – argumentiert. Damit erübrigten sich zugleich auch aufwendige Auswertungsverfahren; Auszählungen und Durchschnittsberechnungen seien hinreichend. Ein Beispiel für diesen Typ von „Einfach-Evaluation“ ist das an der Freien Universität Berlin in regelmäßigen Abständen eingesetzte „FU-Studienbarometer“, das für die standardisierte Beurteilung eines ganzen Studiengangs mit einer einzigen Fragebogenseite auskommt (s. Kromrey 1999, 62 ff.).

In dieser Form eingesetzt, sind mit einer Strategie der Erhebung studentischer Wahrnehmungen und Bewertungen *als* Evaluation von Studium und Lehre allerdings gleich mehrere *Fehlschlüsse* verbunden.

Im Unterschied zur Expertenevaluation anhand vorgegebener Kriterien und auf der Basis systematisch ausgewerteter Informationen mit anschließender kommunikativer Validierung (wie beim Verfahren der peer-Evaluation) sind die befragten Studierenden „Alltags-Evaluatoren“ (vgl. Fußnote 1): Jeder einzelne von ihnen bewertet *irgend etwas* (was er mit dem in der Frage angesprochenen Sachverhalt ad hoc assoziiert) *irgendwie* („alles in allem“ oder „aus aktueller Erfahrung“ oder „mit Blick auf das Wesentliche“ oder ...) *unter irgendwelchen Gesichtspunkten* (Nutzen für sein Studium oder vermuteter Nutzen für den angestrebten Beruf oder aktuelles persönliches Interesse oder abstrakt-verallgemeinertes

Darstellung des Projekts Universitätsentwicklung an der Universität Hamburg durch Bülow-Schramm (2003) illustriert den Prozess der Umsetzung von Evaluation und follow-up an einem konkreten Fallbeispiel.

¹⁴ Übersehen wird dabei, dass mit Studierendenbefragungen in US-Universitäten bewusst Akzeptanzmessung betrieben wird (schließlich sind die Studierenden „Kunden“ der Universität, nämlich Abnehmer einer Dienstleistung, die durch zum Teil hohe Studiengebühren die Institution mitfinanzieren). Für die Evaluation der Lehrenden durch die Hochschule ist allerdings die per Befragung ermittelte Akzeptanz nur *ein* Baustein in einem detaillierteren Bewertungsverfahren.

Interesse *der* Studierenden oder ...). Die Bedeutung der im standardisierten Erhebungsbogen gegebenen Antworten ist so nicht mehr rekonstruierbar.

Werden *Globalaussagen* verglichen mit *differenziert erhobenen Beurteilungen* (unter Verwendung von Itemlisten, mit denen vor der Erhebung zusammenfassender Urteile zunächst Detail-Aspekte eingeschätzt werden), so zeigt sich, dass unter den Studierenden – grob zusammengefasst – zwei geradezu gegensätzliche Evaluierungsweisen existieren. Nahezu die Hälfte der Veranstaltungsteilnehmer urteilt so pauschal, dass in der Tat die Verwendung einfacher und kurzer Erhebungsinstrumente angemessen wäre: Die Tendenz der Einschätzungen auf *allen* Detail-Items einer Dimension stimmt überein mit dem zusammenfassenden Gesamturteil, d.h.: Man ist entweder in jeder Hinsicht zufrieden oder in jeder Hinsicht unzufrieden oder empfindet das Angebot durchweg als mittelmäßig. Die andere Hälfte der Befragten urteilt in den Details differenziert (man ist mit dem einen Teilaspekt zufrieden, mit dem anderen weniger, mit dem dritten unzufrieden) und bildet dann für die Gesamtbewertung der Dimension einen subjektiven Mittelwert. Bei diesem Teil der Studierenden gingen bei Verwendung kurzer „Alles-in-allem“-Fragebögen nicht nur wesentliche Informationen verloren, vielmehr würde dadurch auch der unzutreffende Eindruck einer einfachen, in sich widerspruchsfreien Urteilsstruktur erweckt – methodologisch ausgedrückt: Es würde ein *Erhebungsartefakt* produziert.

Ähnlich problematisch ist die Empfehlung „einfacher Auswertungen“, insbesondere in Form isolierter Auszählungen der Antworten auf die einzelnen Fragen und/oder durch Berechnung von Mittelwerten. Auch hier zeigt die komplexe Analyse differenziert erhobener studentischer Bewertungen die Unangemessenheit solchen Vorgehens: Zum einen werden von den Befragten die Einschätzungen hinsichtlich der verschiedenen Dimensionen und Teildimensionen des Evaluationsgegenstands (z.B. Lehrveranstaltung oder Lehrperson) nicht unabhängig voneinander vorgenommen, sondern sie stehen – selbstverständlich – in einem subjektiv sinnvollen Zusammenhang. Daraus folgt, dass sich die Einzelurteile jedes Befragten zu einem für seine Wahrnehmung *typischen Urteilsprofil* verbinden und dadurch sozusagen „Gestalt annehmen“. Die isolierte Auszählung einzelner Variablen aber lässt solche Profile gar nicht erst sichtbar werden. Zum anderen sind sich die Teilnehmer ein und derselben zu evaluierenden Veranstaltung – eigentlich ebenfalls selbstverständlich – in ihren Beurteilungen nicht einig. Das liegt nicht nur daran, dass ihnen für ihre „Alltags-Evaluationen“ keine intersubjektiven Vergleichsstandards vorgegeben wurden, sondern insbesondere auch daran, dass es sich bei den Befragten nicht um austauschbare Exemplare *der* Gattung Studierende handelt, sondern um Individuen: mit unterschiedlichen Sozialisationserfahrungen und von daher unterschiedlichen Vorkenntnissen, Interessen und Lernstilen, mit unterschiedlichen Präferenzen und Sympathien/Antipathien für die Lehrperson, mit unterschiedlichen Standorten in ihrem Studiengang, mit unterschiedlicher Einschätzung der Brauchbarkeit ihres Studiums und des zu Lernenden für das Leben außerhalb der Hochschule usw. Das heißt: Die Gesichtspunkte, unter denen beurteilt wird, sind sehr verschiedenartig; sie *müssen* demgemäß – wenn der Fragebogen ernsthaft und kompetent ausgefüllt wird – zu unterschiedlichen Urteilen führen. Die Berechnung von Mittelwerten, die die studentischen Individualurteile zu Qualitätskennziffern *der* Teilnehmer kondensieren, produziert *Auswertungsartefakte*.

Für Evaluationen durch Befragung führt das zu dem Fazit: Ein komplexer Sachverhalt kann angemessen auch nur durch hinreichend komplexe empirische Erhebungen valide abgebildet werden; und komplexe Interdependenzen im abzubildenden Sachverhalt werden erst durch hinreichend komplexe Analyseverfahren sichtbar. Der Verweis auf die o.g. Gefahren von

Fehlschlüssen sollte allerdings nicht als Argument gegen die Verwendung von „Alltags-evaluationen“ Betroffener missverstanden werden. Sie liefern wichtige Informationen darüber, wie das „Dienstleistungsangebot Lehre“ bei den Adressaten „ankommt“. Um diese aber als gültige Informationen nutzen zu können, ist zuvor im Zuge der Analyse das Kriteriensystem der Evaluierenden zu rekonstruieren. In methodisch angemessener Form kann die Befragung von studentischen Veranstaltungsteilnehmern als wertvolles Informationsinstrument zur Entwicklung von Lehrqualität genutzt werden.¹⁵

5. Andere Formen des Einsatzes von Befragungen und Erhebungen

Weitere für die Qualitätsentwicklung nutzbare Formen der Umfrageforschung in der Hochschule sind in Lehrveranstaltungen eingesetzte und von Didaktikern schon seit langem empfohlene (kürzere oder längere) Fragebögen als Instrument der *Rückmeldung an die Lehrperson*. Sie dienen nicht der Evaluation, sondern der *Kommunikation* über Lehre und sind vor allem in größeren Veranstaltungen hilfreich, in denen eine direkte Interaktion zwischen Lehrenden und Lernenden nicht mehr ohne weiteres möglich ist. Im Unterschied zu Befragungen *als* Evaluationsverfahren sind der Differenzierungsgrad und die methodische Qualität der Fragebögen ebenso wie die Form der Erhebung zweitrangig. Die Ergebnisse sollen der Lehrperson einen Eindruck von der Sichtweise der Teilnehmer vermitteln, *und* sie sollen der Ausgangspunkt für die Diskussion zwischen Lehrenden und Studierenden über die Lehre sein. Für diesen Zweck sind kurze Fragebögen mit durchaus auch groben Kategorien und zusammenfassend vorzunehmenden Bewertungen sogar von Vorteil: Sie bieten mehr Raum für die Interpretationsphantasie und damit auch mehr Ansatzpunkte für eine engagierte Diskussion.

Gegenstand der Befragung und Bewertung müssen nicht in jedem Fall Lehrveranstaltungen sein. Sinnvolle Fragestellungen können sich richten auf das Curriculum und die übergreifende *Studiensituation* im Fach (Institut, Fakultät/Fachbereich): Wie nehmen die Studierenden die durch Studien- und Prüfungsordnung vorgenommene Definition des Fachs wahr? Wie einleuchtend sind ihnen Struktur und Inhalte des Lehrangebots? Wird der Zusammenhang zwischen Lehre und Prüfungen als hinreichend erkannt? Wie wird die Betreuung empfunden? und vieles mehr. Ebenso sind aktuelle *Kenntnisse über die Studierenden* und ihre Art und Weise des Studierens für die Träger des Curriculums von Bedeutung: Unter welchen persönlichen Bedingungen und wie intensiv wird studiert (etwa Berufstätigkeit neben dem Studium, Anzahl der besuchten Veranstaltungen, zeitlicher Aufwand für das Studium)? Welche inhaltlichen Schwerpunkte setzen die Studierenden dort, wo sie Wahlmöglichkeiten haben? Welche Studierstile sind im Grund-, welche im Hauptstudium anzutreffen? Wie ist das Informationsverhalten der Studierenden? usw. In diesem Zusammenhang kann auch eine Vollerhebung der *Teilnehmerstruktur* in allen Veranstaltungen eines Semesters wichtige Informationen liefern: nicht nur darüber, ob und in welchem Maße eine Veranstaltung ihre definierte Zielgruppe tatsächlich erreicht, sondern auch darüber, ob und in welchem Ausmaß das Fach mit seinen Angeboten Dienstleistungen für andere Fächer liefert (für Nebenfächler,

¹⁵ Als Beispiel für die Entwicklung und den systematischen Test eines differenzierten Befragungsinstruments sei auf die Arbeiten von Heiner Rindermann verwiesen (zusammenfassend: Rindermann 2001). Mit anschließenden Beratungsgesprächen der Lehrenden kombiniert, können solche studentischen Evaluationen einen hohen Stellenwert in Konzepten zur Entwicklung von Lehrqualität erhalten (s. Rindermann 2003). Darauf wird in Abschnitt 6 eingegangen.

aber auch durch Teilnehmer anderer Fächer, die lediglich spezielle Angebote wahrnehmen und dort erbrachte Leistungen im eigenen Fach anerkennen lassen).

Befragungen müssen sich nicht lediglich an einen Querschnitt der aktuell Studierenden richten. Auch *spezifischere Auswahlen* können nützlich sein: Studienanfänger, Studierende im Grundstudium vor der Zwischenprüfung, Studierende bei Beginn des Hauptstudiums, in der Examensphase. Darüber hinaus werden zunehmend *Absolventenbefragungen* durchgeführt, entweder zur ex-post-Evaluation des Studiums aus der späteren Perspektive von Berufstätigen und/oder als Verbleibstudien ehemaliger Studierender. Schließlich kommen auch „*Abnehmer*“-Befragungen vor, insbesondere mit dem Ziel, in potentiellen Berufsfeldern Profile von Anforderungen an das Qualifikationsprofil der Bewerber zu ermitteln. Nicht zuletzt – wenn auch (warum eigentlich?) ganz selten durchgeführt – könnten (und sollten) auch *die Lehrenden* eine Zielgruppe von Erhebungen sein. Eine Konfrontation der Wahrnehmung von Lehre und Lehrpersonen aus der Perspektive der Studierenden mit der Wahrnehmung der Studierenden und ihres Studienengagements durch die Lehrenden dürfte interessante Ergebnisse bringen.

6. Zu guter Letzt: Qualitätsentwicklung ohne Qualitätsbegriff?

Bei aller Unterschiedlichkeit der Argumente der Befürworter von Evaluationen scheint über ein Ziel ihres Einsatzes Einmütigkeit zu herrschen: Sie werden durchgeführt, um die Qualität von Lehre und Studium zu verbessern, um aus den Evaluationsresultaten Hinweise zu gewinnen, *wo* etwas verbesserungsbedürftig ist und *wie* es verbessert werden kann. Wird dieses Ziel akzeptiert, unterstreicht es noch einmal die Notwendigkeit einer *differenzierten* Herangehensweise an das Evaluationsvorhaben: Pauschale Bewertungen und leerformelhafte Aussagen mögen zwar ein geeigneter Ansatzpunkt für das Formulieren von Hypothesen und für das Inangsetzen von Diskussionen sein; für das Erkennen *konkreten* Veränderungsbedarfs und für die Ableitung *konkreter* Maßnahmen sind sie jedoch keine zuverlässige Basis. Ist man sich auch darüber einig, bleibt aber dennoch eine weitere, ganz zentrale Frage unbeantwortet: Was *ist* eigentlich Qualität von Lehre und Studium?

Zunächst: Eine *Qualität* „*alles in allem*“ existiert nicht. Ein Sachverhalt kann zugleich in einer Hinsicht von ausgezeichneter Qualität, in anderer Hinsicht dagegen fehlerbehaftet sein. Es sind also verschiedene Aspekte *oder* „*Dimensionen*“ von *Qualität* – in der Fachdiskussion des Qualitätsmanagements „*Kriterien*“ genannt – zu unterscheiden. Zum anderen: Qualitätsaussagen sind Werturteile. Sollen sie intersubjektiv gefällt werden, sind Vergleichsmaßstäbe – Fachausdruck: „*Standards*“ – notwendig. Mit der Festlegung, aus wessen Perspektive Kriterien und Standards ausgewählt und formuliert werden, ist dann bereits eine wesentliche Vorentscheidung getroffen. Soll nun die Qualität des Sachverhalts „gemessen“ werden – was als Erwartung des öfteren an Evaluationsvorhaben herangetragen wird –, ist das Qualitätskonzept präzise zu definieren und sind die anzulegenden Kriterien und Standards durch geeignete, gültige Indikatoren zu operationalisieren.

Um es an einem einfachen Beispiel zu veranschaulichen: Zu beurteilen sei die Qualität von Autoreifen. Als *Qualitätskriterien* kämen wesentliche Eigenschaften des Objekts selbst in Frage. *Ein* Qualitätskriterium wäre etwa die Haltbarkeit des Produkts, gemessen an der Laufleistung in Kilometern; ein bei der Beurteilung anzulegender *Standard* könnte lauten: mindestens 30.000 km auf glatten Straßen. Andere Kriterien könnten sein: die Bodenhaftung (auf trockener sowie auf nasser Straße), die Sicherheit (bei Überbeanspruchung sowie bei

Außeneinwirkung) u.ä. Auch dazu sind messbare Standards und zuverlässig durchführbare Qualitätstests relativ leicht definierbar.

Nun ist aber - wie oben bereits ausgeführt - die Lehre kein Sachverhalt, dessen Merkmale als Eigenschaften des „Objekts“ direkt ablesbar und in diesem Sinne „objektiv“ messbar wären. Vielmehr ist sie eine *Dienstleistung*, deren Produkt sich erst in der Interaktion von Lehrenden und Lernenden herstellt. Bemühungen, die Qualität von Lehre kontextunabhängig verbindlich zu definieren, sind somit von vornherein zum Scheitern verurteilt. Qualität ist hier keine „objektive“, dem Gegenstand (dem „Objekt“) zurechenbare, sondern eine *relationale* Eigenschaft. Wo dennoch der Versuch unternommen wird, Merkmale „guter Lehre“ aufzulisten, setzt dieser – unabhängig vom Lehr-Inhalt – an der didaktischen Oberfläche an (Webler 1991, S. 246)¹⁶; und selbst da fällt es schwer, Einigkeit über einen Kriterienkatalog für „gute Didaktik“ zu erzielen. Für Einführungsveranstaltungen mit Pflichtcharakter, in denen ein bei Studierenden eher unbeliebter Stoff vermittelt werden soll, wird eine andere Didaktik angemessen sein als in Hauptstudienseminaren zu Spezialthemen mit ausschließlich freiwillig teilnehmenden und interessierten Studierenden oder als in Trainings zur Vermittlung fachübergreifender Schlüsselqualifikationen – um nur wenige unterschiedliche Lehr-Lern-Situationen zu benennen. Und welche Didaktik in diesen Situationen jeweils als angemessen gelten kann, dürfte von verschiedenen Lehrenden ebenso unterschiedlich eingeschätzt werden wie von Studierenden ohne oder mit Vorkenntnissen, ohne oder mit Leistungsmotivation, mit passiv-konsumierendem oder mit aktiv-entdeckendem Lernstil. Eine rein formale Definition – als Qualität der Darbietung – geht jedoch auch *prinzipiell* am Ziel der „Dienstleistung Lehre“ vorbei. Lehre soll ja nicht stromlinienförmig nach Rezeptbuch abgespult werden, ihr Ziel ist auch nicht lediglich das Sich-Wohlfühlen oder die gute oder gar spannende Unterhaltung der Teilnehmer von Lehrveranstaltungen. Sie soll vielmehr Anregungen, Orientierung und – wo nötig – auch Anstöße zum aktiven Studieren geben. Ihr Ergebnis kann nicht in „Einschaltquoten“ oder Zufriedenheits-Kennziffern gemessen werden.

Es bleibt also nur der Ausweg *relativer* Qualitätsdefinitionen, wie dies in der Diskussion um Qualitätsentwicklung und Qualitätssicherung von Dienstleistungen geschieht. Für Ingenieurwissenschaftler liegt es nahe, auf Qualitätsdefinitionen aus der Industrie zurückzugreifen und sie analog auch für die Organisation Hochschule anzuwenden (z.B. Weule 1999). So findet sich etwa in der DIN/ISO-Norm 8402 eine inhalts- und ergebnisbezogene Definition: „Qualität ist die Beschaffenheit einer Einheit bezüglich ihrer Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen.“ Für welche Zwecke die Leistung geeignet sein soll, welche und wessen Erfordernisse festzulegen und vorauszusetzen sind, müsste demnach zunächst ermittelt werden, bevor eine Evaluation beginnen kann. Qualität der Lehre – so ist bis jetzt zu resümieren – kann nicht adressatenunabhängig, sondern kann nur zielgruppenorientiert bestimmt und realisiert werden. Von Studienanfängern und Fortgeschrittenen, von gegenwärtig Studierenden und künftigen Absolventen, von Arbeitgebern und fachwissenschaftlicher community werden unterschiedliche, teils sogar gegensätzliche Erfordernisse geltend gemacht. Die Vorstellung von Lehre als Dienstleistung hat konsequenterweise zur Übernahme des oben bereits genannten Begriffs der Kundenorientierung in die Qualitätsdiskussion geführt – hier allerdings nicht in Analogie zum Wettbewerbsmarkt, sondern als Bezugspunkt für die Definition von Leistungsanforderungen. Soll Lehre ihrem Charakter als

16

Oder es werden recht abstrakte und damit kaum intersubjektiv prüfbare „erfolgsrelevante Persönlichkeitsmerkmale der Lehrenden“ genannt (ders., S. 247 f.).

Dienstleistung gerecht werden, kann somit ihre Qualität und können Qualitätskriterien nicht extern (von wem auch immer) und auch nicht ein für allemal festgesetzt werden, sondern sie müssen den jeweiligen Gegebenheiten angepasst und – wo keine direkte Marktabstimmung durch Angebot und Nachfrage wirksam wird – zwischen den Beteiligten „ausgehandelt“ werden. Dies findet seinen Niederschlag in einem weiteren, an den DIN/ISO-Normen orientierten Definitionsversuch: „Qualität ist die Erfüllung der gemeinsam (Kunde – Lieferant) vereinbarten Anforderungen – einschließlich der Erwartungen und Wünsche“ (Rühl 1998, S. 22). Die Grundtendenz dieser Definition aus dem Produktionsbereich wird inhaltlich auch auf das Qualitätsmanagement von Dienstleistungen übertragen (DIN/ISO 9001 sowie 9004/2, wo als Anwendungsfall ausdrücklich u.a. auf die Wissenschaft verwiesen wird; ausführlicher dazu Stock 1994).

Es stellt sich die Frage: Ist unter diesen Bedingungen „Qualität der Lehre“ in der Universität überhaupt evaluierbar?

Dass für eine flächendeckende Evaluation der Lehre in der Hochschule das methodologische Konzept der Programm- und Wirkungsforschung aus grundsätzlichen wie aus pragmatischen Gründen nicht in Frage kommt, wurde bereits begründet. Es verbleiben damit nur die oben genannten verschiedenen Erhebungsformen vom Typ Befragung, um die für Evaluierungen benötigten empirischen Informationen zu beschaffen. Das Instrumentarium der Umfrageforschung wird daher notgedrungen weit verbreitet zu Lehrevaluationszwecken eingesetzt; es führt aber zugleich – wo dies nicht im Bewusstsein der begrenzten Reichweite von Umfragedaten geschieht – über die oben bereits genannten Schwierigkeiten hinaus zu zahlreichen Problemen der Erhebung und Deutung.

Erhebungs- und Deutungsprobleme, die auf zu geringem Differenzierungsgrad der Frageformulierungen beruhen, wurden bereits hinreichend angesprochen. Hinzu kommen – noch nicht thematisiert – *Probleme der Auswahl* der zu befragenden Informanten. Teilnehmerbefragungen in Lehrveranstaltungen werden üblicherweise in der Mitte des Semesters durchgeführt. Dies hat zur Konsequenz, dass die Ergebnisse in Wahl- und Wahlpflichtveranstaltungen durch den bis dahin wirksam gewordenen Prozess der Selbstselektion einen positiven Bias aufweisen: Diejenigen anfänglichen Teilnehmer, die die Veranstaltung und/oder die Lehrperson besonders negativ einschätzen, sind nicht mehr anwesend. Pflichtveranstaltungen ohne Wahlmöglichkeit zwischen Alternativangeboten erscheinen dadurch im studentischen Urteil systematisch negativer. Noch ungünstiger wirkt es sich auf die Gültigkeit der Resultate aus, wenn die Fragebögen nicht unmittelbar in der Veranstaltung ausgefüllt und wieder eingesammelt, sondern den Studierenden mit der Bitte um Rückgabe mitgegeben werden. Die Konsequenz dieses Vorgehens ist eine Tendenz zur Polarisierung in den erfassten Urteilen. Die Ursache ist leicht nachzuvollziehen: Die Befragung verlangt ein aktives Beteiligungsverhalten von den Befragten; dazu sind vor allem diejenigen motiviert, die „etwas mitteilen“ wollen – sei es ein besonderes Lob (dieser Fall findet sich besonders in Wahlveranstaltungen), sei es explizite Kritik (vor allem unzufriedene Teilnehmer in Pflichtveranstaltungen). Die Mehrheitsgruppe der „einigermaßen Zufriedenen“ ist demgegenüber mangels Motivation zu aktiver Mitwirkung in den Daten deutlich unterrepräsentiert. Eine analoge Verzerrung tritt bei der Erhebung studentischer Aussagen zur Studien-, Betreuungs- und Prüfungssituation in Instituten oder Fakultäten/Fachbereichen auf, wenn sie nicht direkt zu Semesterbeginn, sondern in der Semestermitte anhand einer Stichprobe der im Institut bzw. Fachbereich Anwesenden erfolgt. Diesmal fällt der Bias zugunsten der kontinuierlich Studierenden aus. Wer – aus unterschiedlichen Gründen – die Lehrangebote nur selektiv nutzt

(etwa weil er berufstätig ist und sich lediglich zu Semesterbeginn umfassender informiert), wird in diesem Fall gar nicht erfasst. Damit fallen die Zugehörigen zu einer hinsichtlich des Studierverhaltens gegenwärtig zentralen Gruppe als Informanten vollständig aus.

Ein zweites Problem, das häufig nicht zur Kenntnis genommen wird, hat seine Ursache im (zu Beginn bereits behandelten) ungenauen Sprachgebrauch: Die Erhebung und Auszählung bewertender (also „evaluierender“) Aussagen *ist* noch keine Evaluation im wissenschaftlichen oder methodologischen Sinne, sondern ein Verfahren des Sammelns der persönlichen „Alltags-Evaluationen“ der Befragten. Die *begriffliche Gleichsetzung des Einsatzes der Umfrageforschung zum Zwecke von Evaluation und Qualitätsentwicklung in Lehre und Studium mit Evaluation* verursacht unnötige Verwirrung. Umfrageforschung ist nicht mehr – aber auch nicht weniger – als ein bewährtes Instrument der Informationserhebung; und sie kann als solches ein wichtiger Baustein in einem Konzept von Evaluation und Qualitätsentwicklung sein.

Sinnvollerweise wird man also bei Evaluationsvorhaben innerhalb der Hochschule von einem Konzept ausgehen, das ich zu Beginn als „politischen Evaluationsbegriff“ bezeichnet habe: Ein Programm, eine Maßnahme, eine Einrichtung wird auf der Basis zielgerichtet gesammelter und aufbereiteter Informationen von einem Evaluator (z.B. einem ausgewiesenen Evaluationsexperten oder einem Evaluationsgremium) beurteilt, der abschließend ein Evaluationsgutachten mit Empfehlungen erstellt. Für das Gelingen dieses Konzepts ist es wesentlich, zu Beginn Klarheit darüber zu schaffen, welche Ziele verfolgt werden, welcher Informationsbedarf besteht, wer welche Kompetenzen hat. Mindestens die folgenden Fragen sind eindeutig und verbindlich zu beantworten:

Wer ist zuständig für die Koordination und Durchführung des Evaluationsvorhabens? Diese organisatorisch verantwortliche Evaluationsinstanz hat die Aufgabe der unbeeinflussten, methodisch kontrollierten Sammlung und Aufbereitung der für die Urteilsbildung notwendigen Informationen. Sie erhebt Daten, befragt Informanten, sichtet vorhandene Statistiken, macht sich durch eigene Anschauung ein Bild vom Gegenstand der Beurteilung.

Wer evaluiert? Das heißt: Welche Instanz nimmt auf der Basis der zielgerichtet gesammelten Informationen die Bewertungen vor, trifft Qualitätsentscheidungen? Das kann dieselbe Instanz sein, die das Evaluationsvorhaben koordiniert und durchführt. Es kann aber auch z.B. eine vom Fachbereichsrat eingesetzte Kommission aus Vertretern aller universitärer Gruppen sein.

Welche Informationen sind die Basis für die vorzunehmenden Bewertungen (z.B. Befragungsergebnisse und Hochschulstatistiken und per Beobachtung gewonnene Daten)? Und wer sind die Informanten (z.B. Studierende und Lehrende und die Hochschulverwaltung)?

Was soll evaluiert werden und warum? Ist es die Lehre in Lehrveranstaltungen? Oder die Struktur des Lehrangebots? Sind es die Rahmenbedingungen für das Studieren und für die Lehre? Ist es das Betreuungsverhalten der Lehrenden und/oder das Arbeitsverhalten der Studierenden? Und aus welchen Gründen besteht Interesse an Beurteilungen dieser Gegebenheiten?

Und schließlich: *Zu welchem Zweck soll evaluiert werden?* Was soll mit den Daten geschehen? Evaluation ist kein Selbstzweck (auch wenn es in manchen landesgesetzlichen Regelungen und Erlassen so erscheint). Evaluation ist aufwendig; der Aufwand muss sich lohnen. Durch die Evaluation *allein* wird noch nichts besser. Sie liefert lediglich die

notwendigen Informationen als Voraussetzung dafür, Verbesserungsbedarf zu erkennen und erforderliche Veränderungen in Gang zu setzen. Das aber muss *nach* der Evaluation – als „follow up“ – auch geschehen, wenn nicht die Motivation der Beteiligten erlahmen soll.

Die Evaluation muss münden in konkrete Zielvereinbarungen, oder anders formuliert: in ein Qualitätsentwicklungsprojekt.

Daraus leitet sich die Notwendigkeit einer weiteren Entscheidung ab, die im Abschnitt 2 schon angesprochen wurde: *Wer ist Träger des Qualitätsentwicklungs-Projekts?* Anders formuliert: Wer ist verantwortlich dafür, dass die gelieferten Evaluations-Informationen in konkretes Handeln umgesetzt werden? Dies wird nur in Ausnahmefällen auch der Träger des Evaluationsvorhabens sein.

Wenn Qualität von Dienstleistungen nicht absolut, sondern nur relativ bestimmt und somit auch nur relativ sichergestellt werden kann – nämlich relativ zu den Adressaten (oder „Kunden“) der Dienstleistung –, dann hat der Träger des Qualitätsentwicklungs-Projekts zu entscheiden und zu begründen, *für welche Zielgruppe* er die Dienstleistung optimieren will. Das bedeutet immer zugleich eine Entscheidung *gegen* andere potentielle Adressaten. Der Versuch, einem imaginären ‚Durchschnitt‘ heterogener Zielgruppen mit heterogenen Bedürfnissen und Ansprüchen gerecht zu werden, führt nahezu zwangsläufig zu dem Resultat, dass die Leistung für keine Gruppe von großem Nutzen ist. Die Dienstleistung wird immer auch unter einschränkenden – angesichts knapper werdender Ressourcen im Hochschulbereich: unter zunehmend stärker eingeschränkten – Rahmenbedingungen erbracht. Der Träger des Qualitätsentwicklungs-Projekts wird sich also klarzumachen haben – und wird dies seiner Zielgruppe zu vermitteln haben –, *welche Rahmenbedingungen gestaltbar sind* und welche außerhalb der Möglichkeiten einer Realisierung liegen. Ein Idealkonzept, das Utopie bleiben muss, nützt weniger und schadet mehr als ein Bündel kleiner Schritte in die gewünschte Richtung.

7. Literatur

Bülow-Schramm, Margret, 1995: „Wer hat Angst vor den Evaluatoren?“ Der Umgang mit Akzeptanzproblemen von Evaluationsverfahren. In: Handbuch Hochschullehre, D 1.6, Bonn: Raabe

Bülow-Schramm, Margret, 2003: Evaluation auf dem Weg zum Qualitätsmanagement. In: Hochschulrektorenkonferenz (Hg.): Evaluation und ihre Konsequenzen. Beiträge zur Hochschulpolitik 2/2003. Bonn: HRK, S. 13-34

Chelimsky, Eleanor, 1997: Thoughts for a new evaluation society. „Keynote speech“ at the UK Evaluation Society conference in London 1996. In: Evaluation, 3/1, 97-109

Daxner, Michael, 1999: Evaluation, Indikatoren und Akkreditierung. Auf dem Weg in die Rechtfertigungsgesellschaft. In: Hochschulrektorenkonferenz (Hg.): „Viel Lärm um nichts?“ Evaluation von Studium und Lehre und ihre Folgen. Beiträge zur Hochschulpolitik 4/1999, Bonn: HRK, 41-49

Deutscher Akademischer Austauschdienst / Hochschulrektorenkonferenz (Hg.), 2003: Tagungsbericht Hochschulmarketing im Aufbruch. Qualität und Wettbewerb: Trio Verlag

- Donabedian, A., 1980: Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment, Ann Arbor, MI
- Eekhoff, Johann; Muthmann, R.; Sievert, O., 1977: Methoden und Möglichkeiten der Erfolgskontrolle städtischer Entwicklungsmaßnahmen, Bonn-Bad Godesberg, Schriftenreihe „Städtebauliche Forschung“, Bd. 03.060
- el Hage, Natalija, 1996: Lehrevaluation und studentische Veranstaltungskritik. Projekte, Instrumente und Grundlagen, Bonn: BMBF
- Fischer-Bluhm, Karin, 1995: „Gemeinsam geht es besser!“ Evaluationsprojekte im Verbund norddeutscher Hochschulen. In: Handbuch Hochschullehre, D 3.3, Bonn: Raabe
- Habel, Edna, 1995: „Hochschulen zum Rapport??“ Erfahrungen mit internen Lehrberichten und Lehrberichten nach Universitätsgesetz an der Universität Dortmund. In: Handbuch Hochschullehre, D 1.5, Bonn: Raabe
- Hochschulrektorenkonferenz, 1995: Zur Evaluation im Hochschulbereich unter besonderer Berücksichtigung der Lehre. Entschließung des 176. HRK-Plenums vom 3.7.1995
- Hochschulrektorenkonferenz (Hg.), 1998a: Evaluation. Sachstandsbericht zur Qualitätsbewertung und Qualitätsentwicklung in deutschen Hochschulen. Dokumente & Informationen 1/1998, Bonn: HRK
- Hochschulrektorenkonferenz (Hg.), 1998b: Evaluation und Qualitätssicherung an den Hochschulen in Deutschland - Stand und Perspektiven. Beiträge zur Hochschulpolitik 6/1998, Bonn: HRK
- Hornbostel, Stefan, 1999: Evaluation und Ranking. Führen sie zu mehr Transparenz und Vergleichbarkeit? In: Hochschulrektorenkonferenz (Hg.): „Viel Lärm um nichts?“ Evaluation von Studium und Lehre und ihre Folgen. Beiträge zur Hochschulpolitik 4/1999, Bonn: HRK, 91-95
- Hornbostel, Stefan; Daniel, H.-D., 1995: Das SPIEGEL-Ranking. In: Mohler, P. (Hg.): Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung, Münster: Waxmann, 29-44
- Hübener, A.; Halberstadt, R. 1976: Erfolgskontrolle politischer Planung - Probleme und Ansätze in der Bundesrepublik Deutschland, Göttingen
- Klostermeier, Johannes, 1994: Hochschul-Ranking auf dem Prüfstand. Hochschuldidaktische Arbeitspapiere Nr. 26, Hamburg: IZHD
- Knäuper, Bärbel; Kroeger, Matthias und Studierende, 1999: Qualitätssicherung und -verbesserung im Intensivstudium Psychologie: Ein Werkstattbericht zur Lehrevaluation, Berlin: FU Studiengang Psychologie (der Bericht ist auf der Webseite <http://userpage.fu-berlin.de/~sciencec/iStudium/> einsehbar).
- Kreutz, Henrik; Tarnai, Ch.; Wolff, K. E.; Zezula, S., 1992: Universitätsrangreihen und die Bewertung der Qualität von Hochschulen, 2 Bde., Erlangen-Nürnberg, Wien: IAS
- Kromrey, Helmut, 1994: Wie erkennt man „gute Lehre“? Was studentische Vorlesungsbefragungen (nicht) aussagen. In: Empirische Pädagogik, Jg. 8, H. 2, 153-168

- Kromrey, Helmut, 1995a: Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeiten ihrer Realisierung. In: ZSE Zeitschrift für Sozialisationsforschung und Erziehungssoziologie, Jg. 15, Heft 4, 313-336
- Kromrey, Helmut, 1995b: Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In: Peter Ph. Mohler (Hg.): Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung, Münster, New York, 2. Aufl., 105-127
- Kromrey, Helmut, 1996: Qualitätsverbesserung in Lehre und Studium statt sogenannter Lehrevaluation. Ein Plädoyer für gute Lehre und gegen schlechte Sozialforschung. In: Zeitschrift für Pädagogische Psychologie, Jg. 10, H. 3/4, 153-166
- Kromrey, Helmut, 1999: Von den Problemen anwendungsorientierter Sozialforschung und den Gefahren methodischer Halbbildung. In: SuB Sozialwissenschaften und Berufspraxis, Jg. 22, H. 1, 58-77
- Küchler, Tilmann, 1996: Für eine verbesserte Qualität der Lehre. Empfehlungen des Wissenschaftsrates zur Evaluation der Lehre. In: Handbuch Hochschullehre, D 1.7, Bonn: Raabe
- Künzel, Ellen; Nickel, Sigrun; Zechlin, Lothar, 1999: Organisationsentwicklung an Hochschulen. Was geschieht mit den Evaluationsergebnissen? In: Hochschulrektorenkonferenz (Hg.): „Viel Lärm um nichts?“ Evaluation von Studium und Lehre und ihre Folgen. Beiträge zur Hochschulpolitik 4/1999, Bonn: HRK, 105-119
- Kultusminister- und Hochschulrektorenkonferenz, 1994: Umsetzung der Studienstrukturreform, Bonn: HRK
- Landfried, Klaus, 1999: Qualitätssicherung als Aufgabe wettbewerblicher Hochschulen. In: HRK (Hg.): Ein Schritt in die Zukunft. Qualitätssicherung im Hochschulbereich. Beiträge zur Hochschulpolitik 3/1999, Bonn: HRK, 7-13
- Müller-Böling, Detlef / Hornbostel, S. / Berghoff, S. (Hg.), 2001: Tagung Hochschulranking. Aussagefähigkeit, Methoden, Probleme. Gütersloh: CHE
- Peter, Lothar; Wawrzinek, Andreas, 1995: Dialogische Evaluation. Ein studentisches Evaluationsverfahren. In: Handbuch Hochschullehre, D 3.1, Bonn: Raabe
- Rindermann, Heiner, 2001: Lehrevaluation. Einführung und Überblick. Landau: Verlag Empirische Pädagogik
- Rindermann, Heiner / Kohler, J., 2003: Lässt sich Lehrqualität durch Evaluation und Beratung verbessern? Überprüfung eines Evaluations-Beratungs-Modells. In: Psychologie in Erziehung und Unterricht, Jg. 50, H. 1, S. 71-85
- Rühl, Werner J., 1998: ISO 9000 – Erfahrungsbericht aus einem technischen Entwicklungszentrum. In: Hochschulrektorenkonferenz: Qualitätsmanagement in der Lehre. TQL 98. Beiträge zur Hochschulpolitik 5/1998, Bonn: HRK, 21-46
- Schierholz, Petra; Vocke, Christina, 1997: Selbstkritik und Aufbruch. Evaluation im Verbund Norddeutscher Universitäten - Erfahrungen und Konsequenzen. In: Handbuch Hochschullehre, D 3.4, Bonn: Raabe
- Schmidt, Jörn, 1980: Evaluation als Diagnose. HDZ-Dozentenkurs, Essen

Schreier, Gerhard, 1998: Das HRK-Projekt Qualitätssicherung - Konzepte und Ziele. In: Hochschulrektorenkonferenz (Hg.): Evaluation und Qualitätssicherung an den Hochschulen in Deutschland. Beiträge zur Hochschulpolitik 6/1998, Bonn: HRK, 13-17

SPIEGEL Spezial, 1990: „Welche Uni ist die beste?“, Hamburg

Stock, Wolfgang G., 1994: Wissenschaftsevaluation. Die Bewertung wissenschaftlicher Forschung und Lehre. ifo Diskussionsbeiträge 17, München: ifo Institut für Wirtschaftsforschung

Webler, Wolff-Dietrich, 1991: Kriterien für gute akademische Lehre. In: Das Hochschulwesen, Jg. 39, Heft 6, 243-249

Weiss, Carol H., 1974: Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogrammen, Opladen

Weule, Hartmut, 1999: Praktische Probleme der Qualitätssicherung an Hochschulen. In: Hochschulrektorenkonferenz (Hg.): Ein Schritt in die Zukunft. Qualitätssicherung im Hochschulbereich. Beiträge zur Hochschulpolitik 3/1999, Bonn: HRK, 45-54

Wissenschaftsrat 1996: Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation. In: ders.: Empfehlungen und Stellungnahmen 1996, Band I, Köln

Wollmann, Hellmut; Hellstern, G.-M., 1977: Sozialwissenschaftliche Untersuchungsregeln und Wirkungsforschung. Zur Chance kritischer Sozialwissenschaft im Rahmen staatlicher Forschungsaufträge. In: P. Haungs (Hg.): Res Publica. Dolf Sternberger zum 70. Geburtstag, München 1977, 415-466

Autor:

Prof. Dr. Helmut Kromrey
Kufsteiner Str. 12, 10825 Berlin
kromrey@t-online.de