

Studierendenbefragungen als Evaluation der Lehre?

Anforderungen an Methodik und Design

Helmut Kromrey

veröffentlicht in:

Uwe Engel (Hg.): "Hochschulranking. Zur Qualitätsbewertung" von Studium und Lehre", Frankfurt/M., New York 2001: Campus, 11-47

Vorbemerkungen:

Qualitätsentwicklung in Studium und Lehre ist unbestreitbar ein wichtiges Ziel für an den Hochschulen verantwortlich Handelnde – nicht nur wegen knapper gewordener und knapp bleibender Ressourcen. Dies dokumentiert sich nicht nur darin, dass seit 1998 im Auftrag der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung bei der Hochschulrektorenkonferenz ein „Projekt Qualitätssicherung“ u.a. mit dem Ziel der Förderung der Zusammenarbeit zwischen den Hochschulen durchgeführt wird. Ebenso unbestreitbar ist in diesem Bemühen Qualitätsbewertung – oder „Evaluation“ – ein bedeutsames Instrument. Als solches muss es verlässliche und für den Zweck seines Einsatzes brauchbare Ergebnisse liefern.

Mit dieser Forderung ist bereits ein zentrales Dilemma angesprochen: Verlässliche Resultate liefert Evaluation vor allem dann, wenn sie als methodisch kontrollierte empirische *Evaluationsforschung* realisiert wird. Dann aber bindet sie ihrerseits soviel Ressourcen, dass sie dem eigentlichen Ziel – Qualitätsentwicklung in Studium und Lehre – im Wege stehen kann. Außerdem benötigt sie als *Evaluationsforschung* relativ viel Zeit, so dass das Warten auf Resultate die ohnehin langwierigen Entscheidungsprozesse in den Hochschulen noch weiter in die Länge ziehen kann.

Es ist daher nicht verwunderlich, wenn vor allem einfache und wenig zeitaufwendige Verfahren der Qualitätsbewertung gefordert werden. Verwunderlicher dagegen ist es, wenn in diesem Zusammenhang immer wieder und anscheinend völlig unreflektiert zwei Strategien unangemessener *Vereinfachung durch Reduktion der Problemsicht* vorgeschlagen werden:

- zum einen die Reduktion des komplexen Wirkungszusammenhangs universitärer Bildung und Ausbildung – also von Studienbedingungen, Lernen und Studieren *in einem Fach* – auf die Lehrpräsentation *in einzelnen Veranstaltungen*,
- zum anderen die Reduktion von Evaluation als verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens entscheidungsrelevanter Informationen auf herkömmliche Umfrageforschung, auf die Befragung einer selbstselektiven Teilmenge von Studierenden, nämlich von Teilnehmerinnen und Teilnehmern in einzelnen Lehrveranstaltungen.

Von Einstein wird die Aussage kolportiert, man solle alles so einfach wie möglich tun – aber nicht einfacher. Darauf bezieht sich das Fragezeichen im Titel meines Beitrags: *Wie einfach können wir Evaluation (von Studium und Lehre) realisieren?*

1. Was bedeutet der Begriff “Evaluation”?

Evaluation ist in der allgemein-öffentlichen wie in der hochschulpolitischen Diskussion zu einem vielfältig schillernden Modewort geworden, das je nach dem Verwendungskontext sehr Unterschiedliches bedeutet.

Alltagssprachlich ist es lediglich ein anspruchsvoller klingendes Fremdwort für *Bewertung*: Irgendwer bewertet irgend etwas irgendwie unter irgendwelchen Gesichtspunkten.

Anders *in der politischen Diskussion*. Hier bedeutet es 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". Evaluiert werden in dieser Weise etwa die Institute der "blauen Liste" und wurden nach der deutschen Vereinigung die Hochschuleinrichtungen der ehemaligen DDR.

Auch im Kontext "ganz gewöhnlicher" *Umfrageforschung* wird häufig von Evaluation 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" und ist das Sammeln der Auffassungen von Professorinnen und Professoren über das Image verschiedener Universitäten eine "Evaluation der Hochschulen".

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 Zuordnung durch Randomisierung oder ersetzt die statistische Kontrolle von "Störgrößen" deren nicht mögliche Abschirmung in der Erhebungssituation (vgl. für einen zusammenfassenden Überblick Frey/Frenz 1982). 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 die 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. Gemeinsam bleibt bei allen Ausdifferenzierungen des Vorgehens dennoch 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?

Doch selbst im Kontext angewandter Sozialwissenschaft steht der Begriff „Evaluation“ nicht lediglich für ein *methodologisches* Konzept, sondern umfasst nicht selten darüber hinaus einen komplexen *Handlungszusammenhang*, der sich neben der Forschung (dem Sammeln und Auswerten von Informationen über die Implementation und die Auswirkungen von Programmen) auch auf die wissenschaftliche Beratung oder gar die aktive Teilnahme am Prozess der Entwicklung, Implementation und Optimierung erstreckt. Zum Teil wird hierfür der Terminus „*offene Evaluation*“ verwendet, im Unterschied zum eben skizzierten „geschlossenen“ Konzept, bei dem für die Evaluation ein in der Umsetzung befindliches Programm, zumindest jedoch eine explizit ausformulierte Planung vorgegeben ist. Akteure in „offenen“ Evaluationsvorhaben können in die gesamte Programm-Historie eingebunden sein: von der Aufarbeitung und Präzisierung von Problemwahrnehmungen und Zielvorstellungen über eine zunächst vage Programmidee, über die Entwicklung geeigneter erscheinender Maßnahmen und deren Erprobung bis hin zu einem auf seine Güte und Eignung getesteten

(endgültigen) Konzept. Zugleich darf ihr Blickwinkel nicht auf den Sachverhalt „Programm“ beschränkt bleiben, sondern muss explizit auch die Beteiligten und Betroffenen – die „stakeholders“, wie der schwer übersetzbare amerikanische Fachterminus lautet – einbeziehen. Die Aufgabe besteht nicht lediglich im Zurverfügung-Stellen von Informationen, sondern auch in deren Vermittlung, zielgruppengerechter „Übersetzung“, im Entwickeln von Umsetzungs-Ideen, im Moderieren, im Vermitteln zwischen unterschiedlichen Perspektiven. „Evaluation“ wird so zu einem umfassenden Informations- und Prozessmanagement (für eine ausführlichere Darstellung der Konzepte s. Kromrey 1995a).

Aus den bisher dargestellten Begriffsverwendungen kann bei genauerem Hinsehen noch eine weitere, zusätzliche Verwirrung stiftende Dimension des Begriffs erkannt werden. 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*. Damit das Evaluationsurteil nicht rein subjektiv bleibt, sondern den Anspruch intersubjektiver Geltung erheben kann, muss das Evaluieren nach objektivierbaren Regeln geschehen, muss es einer nachprüfbaren Evaluationsmethodik folgen. So existieren beispielsweise relativ genaue Verfahrensvorschläge für die Gewinnung der entscheidungsrelevanten Informationen und für das *Erstellen von Evaluationsgutachten durch Sachverständige im Hochschulbereich* (vgl. etwa Wissenschaftsrat 1996). Gleiches 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 aus den erhobenen subjektiven Werturteilen der je einzelnen Befragten zu intersubjektiven Evaluationen zu gelangen; so z.B. für den Fall studentischer Lehrbewertungen:

- “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,
 - sie legen praktisch dieselben Kriterien für gute Lehre an wie die Dozenten selbst,
 - sie sind – wenn man etwa 20-30 Studenten urteilen lässt – zuverlässig wie professionelle Testverfahren,
 - sie sind von anderen Merkmalen der Studenten selbst und der Dozenten wenig beeinflusst”.
- (Schmidt 1980, 51-52)¹

Ein erstes Fazit: Schon auf der Ebene des Argumentierens gilt die Mahnung: so einfach wie möglich, *aber nicht einfacher!* Jedenfalls verbietet es sich angesichts der Vieldeutigkeit des Evaluationsbegriffs, ohne weitere Präzisierung von *der* Evaluation zu sprechen. Vor jeder inhaltlichen Argumentation ist zunächst klarzustellen, auf welchen Typ von Evaluation (als wertende Aussage wie auch als Verfahren) sich das Argument beziehen soll.

2. Was wird an Hochschulen unter dem Label “Lehrevaluation” getan?

2.1 Zur Unmöglichkeit von Evaluation als Erfolgskontrolle

Dass das Modell Evaluation als Programmforschung 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

¹ Bisher konnte allerdings nur die Geltung von Axiom 2 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).

um dadurch evaluierbar zu sein.² 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 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 Indikатораusrprägungen 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.

² 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.

2.2 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 1998):

- „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.; 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).

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 (stakeholder-Modell). 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.

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, Abschnitt 1). Manche Lehrende gehen noch einen Schritt weiter und vertreten unter Verweis auf „jahrzehntelang bewährte Praxis in den USA“³ 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“: Jeder einzelne von ihnen bewertet *irgend etwas* (was er mit dem in der Frage

³ 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.

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 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 (s.o.), 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*.

Das führt zum zweiten 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. Die „ärgerlich komplizierte“ soziale Realität wird nicht dadurch einfach, überschaubar und leicht handhabbar, dass man sich Scheuklappen anlegt. Lediglich die Gefahr, Wichtiges zu übersehen und fehlerhafte Entscheidungen zu treffen, wächst.

Der Verweis auf die o.g. Gefahren von Fehlschlüssen sollte allerdings nicht als Argument gegen die Verwendung von „Alltagsevaluationen“ Betroffener missverstanden werden. Um diese als gültige Informationen nutzen zu können, muss jedoch im Zuge der Analyse das Kriteriensystem der Evaluierenden rekonstruiert werden. So werden auch Akzeptanzaussagen zu wichtigen Informationen. In aufwendigerer Form kann die Befragung von studentischen Veranstaltungsteilnehmern sogar als wertvolles Informationsinstrument zur Entwicklung von Lehrqualität genutzt werden. Auf ein Beispiel wird später (Abschnitt 6) noch eingegangen.

2.3 Umfrageforschung als Rückmeldung

Weitere 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. Eine lange Liste statistischer Werte, die den Eindruck erwecken, an ihnen gebe es "nichts mehr zu deuteln", blockiert dagegen von vornherein jede Diskussion. Gute Dienste leisten z.B. an jeden Teilnehmer verteilte Blätter (DIN A 5 oder kleiner) mit folgendem Text:

Mini-Rückmeldung

Wenn Ihnen während des Seminarverlaufs etwas Anmerkenswertes auffällt, bitte sofort notieren (*Anmerkenswert* ist alles, was Sie stört, was Ihnen besonders gefällt, was Sie an Ideen für Veränderungen haben). Bitte immer nur eine Mitteilung pro Blatt!

Intensitätsskala:

0 1 2 3 4 5

Eigentlich
nicht so
wichtig

Sehr
bedeutsam

Bewertung:

+ (oder) -

Regelmäßig eingesetzt und am Ende von Veranstaltungen eingesammelt, ist ein solches Mini-Rückmeldeinstrument selbst in großen Vorlesungen ein wertvoller Seismograph, der sofort anzeigt, was im Auditorium ansonsten vom Lehrenden unbemerkt abläuft. Wichtig ist jedoch bei jeder Form eingesetzter Rückmeldeinstrumente, dass sie tatsächlich für das Ingangsetzen einer Diskussion über die Lehre genutzt werden und nicht lediglich als eine lästige Pflicht erscheinen.

2.4 Andere Formen des Einsatzes von Befragungen und Erhebungen

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, 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.

3. Warum / wozu „Evaluation“?

Evaluation von Studium und Lehre – wenn sie in der dem Gegenstand angemessenen Komplexität realisiert wird – ist zeit- und ressourcenaufwendig. Dieser Aufwand muss sich lohnen. Die Evaluation muss für die Beteiligten einen erkennbaren Nutzen bringen, soll sie auf die erforderliche Mitwirkungsbereitschaft treffen, um überhaupt mit Aussicht auf Erfolg durchführbar zu sein. Evaluation braucht Akzeptanz, um gültige Ergebnisse zu liefern. Es ist

also 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 Selbstzweck, als „Evaluations-Ritual“ erscheinen. Nicht selten jedoch geschieht 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 immer auch dann zu erwarten, wenn *Evaluation 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, S. 41 ff.).

Weniger strittig ist mittlerweile die *Evaluation als hochschulinternes Steuerungsinstrument*, zum Teil verknüpft mit „incentives“ z.B. für gute Lehrorganisation und Forschungsleistungen. In manchen Bundesländern wird ein Teil der 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 objektiv 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 (gesamt und relativ), Studiendauer der Absolventen, Promotionen, Studienabbrüche etc. (als Leistungsindikatoren). Verbunden sind damit wieder methodische Probleme, u.a. dass Indikatoren immer nur einen Ausschnitt aus dem gesamten Problemfeld abbilden können und dass auch dies immer nur mit zweifelhafter Gültigkeit geschehen kann. Eine hohe scheinbare „Studienabbruchquote“ im Grundstudium eines Fachs kann auch dadurch zustande kommen, dass dieses Fach von Studierenden gewählt wurde, die in einem numerus-clausus-Fach keinen Studienplatz erhalten haben, nun einen „Quereinstieg“ suchen und ihn nach zwei oder drei Semestern auch erreichen. 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 (die Zahl erfolgreicher Studienabschlüsse kann man auch maximieren, indem man die Prüfungsanforderungen minimiert). Indikatorensysteme – das weiß man aus den Erfahrungen der Sozialindikator-Bewegung – 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 mit 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, S. 10). Evaluation schafft in diesem Zusammenhang „ein Stück Markt-Ersatz, eine Art Quasi-Wettbewerb“ (ders., S. 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 der 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. Beziehen wir das Argument auf den Sektor Lehre und Studium, so gilt zwar einigermaßen unbestritten als letztl. 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 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“ (darauf wird im folgenden Abschnitt näher einzugehen sein). Im Detail kann an dieser Stelle darauf hier nicht eingegangen werden – 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 hat die Empirie doch 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 nichts gewonnen. 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. Damit stoßen wir auf zwei Fragen, deren Zusammenhang häufig nicht beachtet wird. Wer ist *Träger der Evaluation*? Und wer ist

Träger des Qualitätsentwicklungs-Vorhabens? Anders formuliert: Wer ist verantwortlich dafür, dass die gelieferten Evaluations-Informationen in Handeln umgesetzt werden? 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).

Für die Evaluation von Lehrveranstaltungen mit dem Ziel, diese für die Studierenden zu verbessern, ist in der Regel eine interne Evaluation (Lehrperson als Träger) empfehlenswert, ggf. unter Hinzuziehung eines Methodenexperten und/oder Didaktikers als Berater. Ist der Gegenstand der Evaluation und Qualitätsentwicklung ein Curriculum und/oder eine Struktureinheit der Universität (Institut, Fachbereich/Fakultät), hat sich die externe Evaluation durch „peers“, die im Einvernehmen mit der Struktureinheit als Evaluatoren bestellt wurden, bewährt (s. Abschnitt 2.2)⁴. In diesem Fall ist eine formale Ergebnissicherung der Evaluation durch die Formulierung von Zielvereinbarungen ebenso vonnöten wie ein „follow up“ im Abstand von einigen Jahren, um – wiederum gemeinsam – die Angemessenheit von Zielen und getroffenen Maßnahmen zu überprüfen.

4. Was ist „Qualität“?

Das Fazit der obigen Antworten auf die Frage „Warum/wozu ‚Evaluation‘?“ lautete: 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. Dies unterstreicht noch einmal die Notwendigkeit einer *differenzierten* Herangehensweise an das Evaluationsvorhaben: Pauschale Bewertungen und leerformelhafte Aussagen mögen ein geeigneter Ansatzpunkt für das Formulieren von Hypothesen und für das Ingangsetzen von Diskussionen sein; sie sind jedoch keine zuverlässige Basis für das Erkennen *konkreter* Veränderungsbedarfs und für die Ableitung *konkreter* Maßnahmen.

Nicht beantwortet ist damit aber eine weitere, ganz zentrale Frage: 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.

⁴ In einigen Bundesländern existieren bereits Zentrale Evaluationsagenturen der Hochschulen (z.B. in Niedersachsen und Nordrhein-Westfalen), die die Funktion als externe Träger von Evaluationsvorhaben übernehmen.

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 die Lehre kein Sachverhalt, dessen Merkmale als Eigenschaften des „Objekts“ direkt ablesbar und in diesem Sinne „objektiv“ messbar wären. Im Unterschied zu gegenständlichen Produkten – wie dem o.g. Autoreifen – ist Lehre eine *Dienstleistung*, deren Produkt (Lernservice für Studierende) 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)⁵; 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

⁵ Oder es werden recht abstrakte und damit kaum intersubjektiv prüfbare „erfolgsrelevante Persönlichkeitsmerkmale der Lehrenden“ genannt (ders., S. 247 f.).

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 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 Marktstimmung 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).

5. Ist „Qualität der Lehre“ in der Universität 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 im Abschnitt 2.1 begründet. Es verbleiben damit nur die verschiedenen Erhebungsformen vom Typ Befragung, um die für Evaluierungen benötigten empirischen Informationen zu beschaffen. Das Instrumentarium der Umfrageforschung wird daher auch weit verbreitet zu Lehrevaluationszwecken eingesetzt; es führt aber zugleich – wo dies nicht im Bewusstsein der begrenzten Reichweite von Umfragedaten geschieht – 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.⁶ Sie liefert lediglich die

⁶ Außer vielleicht in den Fällen, wo *wegen* einer anstehenden Evaluation die Organisation sich erst ihrer Aufgaben bewußt wird und – um nicht zu schlecht abzuschneiden – Schwachstellenanalyse betreibt und erkannte Schwachstellen beseitigt. Aber auch in einem solchen Fall ist die Evaluation nicht Selbstzweck, sondern offenbar ein Instrument der mit möglichen Konsequenzen verbundenen Kontrolle – andernfalls sähe

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: *Wer ist Träger des Qualitätsentwicklungs-Projekts?* Wer ist dafür verantwortlich, dass die gelieferten Evaluationsinformationen 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.

6. Qualität entwickeln, ohne Qualität zu „messen“ – einige Beispiele

Zu Beginn (Abschnitt 2.1) war resümiert worden, im Kontext Studium und Lehre an der Hochschule sei das speziell für die Evaluierung entwickelte methodische Design – das Konzept der (experimentellen oder quasi-experimentellen) Programmforschung – nicht einsetzbar. Im Abschnitt 3 wurde des weiteren festgestellt, auch das „Messen“ von Qualität als Aufgabe von Evaluation sei nicht einlösbar. Es wäre jedoch falsch, daraus die Empfehlung herzuleiten, in der Hochschule von dem Vorhaben generell Abstand zu nehmen. Angezeigt ist lediglich der Verzicht auf diese Formen von Evaluation. Es wurden ja auch bereits einige Beispiele skizziert, wie dennoch sinnvoll und erfolgversprechend auch in der Hochschule evaluiert werden kann, allerdings *nicht* verstanden als *spezifisches methodologisches Konzept*, sondern als empirische *Sozialforschung in einem spezifischen Verwertungskontext* – als methodisch kontrollierte, verwertungs- und bewertungsorientierte Form des Sammelns und Auswertens von Informationen. Damit ist der Evaluation explizit die Aufgabe zugewiesen, zur Entwicklung und Verbesserung von Qualität durch Bereitstellung einer differenzierten und qualitätsrelevanten Beratungs- und Entscheidungsbasis beizutragen. Evaluation ohne „follow up“ ist nutzlos.

So klagten Künzel/Nickel/Zechlin, die über Erfahrungen aus einem Organisationsentwicklungsprojekt berichten: „Evaluieren ist schon schwierig genug, aber noch eine viel größere Herausforderung stellt die Veränderung der Realität an Hochschulen dar. Solange man nur Datenmaterial zur Qualität von Lehre und Forschung auf geduldigem Papier zusammenstellt und darüber in den Gremien redet, bleibt die Evaluation folgenlos und damit harmlos. Unbequem wird es erst, wenn aus dem vermeintlichen Datenfriedhof

Konsequenzen auf der Handlungsebene gezogen werden. (...) Ob und was auf der Handlungsebene dann tatsächlich passiert, ist die Messlatte für den Erfolg einer Evaluation.“ (1999, S.105).

In ihrem Beitrag schildern die Autoren konkret den Ablauf eines solchen Entwicklungsprojekts – von der „Stärken-Schwächen-Analyse als Basis von Veränderungsprozessen“ über die „Erstellung eines Handlungskatalogs“, über „Zielplanung und Zielvereinbarung“ bis zu Problemen in der „Leistungs- und Entscheidungsstruktur“ (S.107-113). Mit Verweis darauf wird an dieser Stelle auf ein diesbezügliches Beispiel verzichtet. Stattdessen wird im folgenden etwas ausführlicher ein Anwendungsbeispiel auf der Mikroebene – die Aufgabe „Qualitätsentwicklung in Lehrveranstaltungen“ – illustriert.

6.1 Beispiel: Zielgruppenorientierte Lehre

Im Abschnitt 4 wurde dargestellt, dass Lehrqualität sinnvoll nur relational – als Angemessenheit des Angebots (der Lehrenden) für definierte „Kunden“ (Studierende) – entwickelbar ist. *Lehre* kann – wie sehr sie auch einer „best practice“ didaktischer Kunst folgen mag – immer nur in begrenztem Ausmaß *Lernen* bewirken. Ob sie den Lernprozess und das Lernergebnis positiv beeinflusst oder erfolglos bleibt, hängt nicht in erster Linie von der didaktischen Qualität der Darbietung ab, sondern ist das Resultat der gesamten Lehr-Lern-Situation. Damit sind vielfältige – in wechselseitiger Beziehung stehende – Dimensionen angesprochen, u.a.

- das Lehr- bzw. Ausbildungsprogramm / der Studiengang
(zu vermittelnde „Inhalte“ – also Kenntnisse, Fertigkeiten und Fähigkeiten – und deren Stellenwert im Gesamtcurriculum)
- die „Lehrmittel“
(Lehrpersonal, Medien, Aufgaben, Betreuung/Beratung)
- die „Lerner“
(übergeordnete Studienziele, Lernziele in der Veranstaltung, Teilnahmegründe, Interesse/Motivation, Vorkenntnisse, Lern-Erfahrungen, Lern- und Arbeitsstile, Stellenwert des Studierens gegenüber anderen Tätigkeiten und Verpflichtungen, Zeitbudget zum Lernen für diese Veranstaltung)
- die materielle Lernumgebung
(Lernort, Ausstattung, zeitliche/räumliche Flexibilität, Teilnehmerzahl)
- die soziale Lernumgebung
(Einzel-/Gruppenlernen, Lern-„Klima“, Stellenwert des Studiums im sonstigen Lebenskontext, Kommiliton(inn)en, Familie/Bekanntkreis).

Nur ein kleiner Teil der genannten Dimensionen ist vom Lehrenden *gestaltbar*. Alle aber müssen für einen gelingenden Lernprozess *bekannt sein* und bei der Lehr-/Lernplanung *berücksichtigt werden*. D.h. die gestaltbaren Faktoren der Lernumwelt sind so auf die nicht-veränderbaren Faktoren abzustimmen, dass ein in sich stimmiges, situationsangepasstes Lernarrangement zustande kommt. Die Konsequenz aus dieser Forderung ist, dass jede Lehr-/Lernplanung immer zumindest ein Minimum an Vorab-Informationen verlangt.

Von diesen Überlegungen ausgehend, wurde (und wird) am Institut für Soziologie (IfS) der Freien Universität Berlin ein Projekt mit dem Ziel der Lehrqualitäts-Entwicklung für *interessierte Studierende* (als explizite Zielgruppe⁷) durchgeführt. Für wiederkehrende

⁷ Jede/r Lehrende wird die Erfahrung gemacht haben, dass ein nicht unerheblicher Teil der Immatrikulierten nicht aus intrinsischem Interesse am Fach studiert und daher auch in manchen Lehrveranstaltungen nur anwesend ist, um den formalen Anforderungen der Studienordnung zu genügen. Studierendenbefragungen

Lehrveranstaltungen, die zum Pflichtkanon des teilweise neu zu konzipierenden Diplomstudiengangs gehören, sollen für diese Zielgruppe in den jeweiligen Veranstaltungen "Lösungen nach Maß" gefunden werden. Dies erfordert hinreichende Informationen über die Teilnehmer; in Teilnehmerbefragungen können sie ermittelt werden.

Allerdings ist Qualitätsentwicklung und Qualitätssicherung nicht auf der Basis grober, globaler, vereinfachender Informationen möglich. Benötigt werden detaillierte Daten über die „Kunden“, für die die Dienstleistung entwickelt werden soll; hier: Lernziele und Lernvoraussetzungen der Studierenden, Erwartungen, Ansprüche, natürlich deren Einschätzungen und Urteile. Auf all dies muss eine Lehrperson eingehen können, wenn sie eine Lehrveranstaltung zielgruppenorientiert und adressatengerecht konzipieren und durchführen will. Dabei kann es sich herausstellen, dass die Erwartungen der Teilnehmer zueinander in Widerspruch stehen, nicht "unter einen Hut" zu bekommen sind, oder dass die Teilnehmererwartungen mit den Absichten und Zielen der Lehrperson oder der Lehrinstitution in Widerspruch stehen. In solchen Fällen kann (und soll) das Ergebnis in der Lehrsituation thematisiert werden, muss entschieden werden, welchen Erwartungen entsprochen werden kann und welchen nicht; dann haben auch die Teilnehmer die Chance, sich zu entscheiden, ob sie dennoch weiter teilnehmen wollen oder ob sie lieber wegbleiben.

Qualitätsentwicklung (und Qualitätssicherung) ist allerdings kein punktuelles, lediglich einmal stattfindendes Vorhaben, sondern zwangsläufig ein Prozess, der einige Zeit braucht und einige Durchläufe benötigt. Von daher ergibt sich: Es ist ein Konzept, das *nicht in jedem Semester in jeder Veranstaltung* verfolgt werden kann. Es eignet sich eher für ein auf Wiederholung angelegtes Lehrprogramm: Einführungsveranstaltungen im Grundstudium, regelmäßig wiederkehrende Bestandteile des Hauptstudiums-Curriculums. In anderen Situationen ist der Lehrprozess im Idealfall im Diskurs zwischen Lehrenden und Lernenden „auszuhandeln“.

Bestandteile des am IfS verfolgten Konzepts sind im Kern zwei schriftliche Befragungen (eine zu Beginn, eine abschließende gegen Ende des Semesters):

Die Anfangsbefragung der Teilnehmer (in der zweiten Semesterwoche) erhebt für jede Lehrveranstaltung folgende Schwerpunkte:

- Gründe für die Teilnahme an dieser Veranstaltung; Informationsquellen, aufgrund derer die Wahl getroffen wurde;
- Beschreibung der Studiensituation der Befragten: zeitliche Belastung durch Jobs und andere Verpflichtungen außerhalb des Studiums (etwa Familie, Ehrenämter), geplante zeitliche Investition in diese Veranstaltung, beabsichtigte Lernform (Einzelarbeit, Gruppe, Besuch ergänzender Veranstaltungen, Selbststudium);
- persönliche Lernziele, an die Veranstaltung geknüpfte Erwartungen und Befürchtungen;
- die persönliche "Didaktik-Theorie" der Teilnehmer (Einschätzung der Wichtigkeit bestimmter didaktischer Aspekte wie Diskussion, Skript, Medieneinsatz, Wiederholungen, Lernfortschrittstests etc.).

Schwerpunkte der Abschlussbefragung der Teilnehmer (in der vorletzten Veranstaltungswoche) sind:

bestätigen diesen Eindruck; von über 10.000 befragten Hörern in Vorlesungen mehrerer Fakultäten der Ruhr-Universität Bochum waren 55,6 % (!) ausschließlich deshalb in der Veranstaltung, weil es die Studien- oder Prüfungsordnung zwingend vorschrieb (s. auch weiter unten: Abschnitt 6.3). Es wäre wenig sinnvoll, wollte man „Lehrqualität“ für diese Personengruppe optimieren.

- Wiederaufgreifen der Themenbereiche "Erwartungen / Befürchtungen" und abschließende (rückblickende) Beurteilung, ob sie sich bestätigt haben bzw. ob es besser oder schlechter gelaufen ist;
- Wiederaufgreifen des Themas "Lernformen": tatsächlich in die Veranstaltung investierter Arbeitsaufwand; Beeinträchtigung des Studiums durch außeruniversitäre Verpflichtungen;
- Wiederaufgreifen der didaktischen Aspekte und deren rückblickende Beurteilung für die abgelaufene Veranstaltung; Beurteilung der Person und des Auftretens der/des Lehrenden sowie des wahrgenommenen eigenen Lernerfolgs;
- Informationen über Teilnehmerfluktuation (eigener Wechsel der Lehrveranstaltung und dessen Gründe; Berichte über Gründe von Bekannten, die aus der Veranstaltung weggeblieben sind).

Die Anfangsbefragung wird bis zur vierten Semesterwoche ausgewertet und in die Veranstaltung rückgekoppelt (Diskussion mit den Teilnehmern). Auf der Basis der gewonnenen Informationen kann noch im laufenden Semester das Angebot und seine Darbietung an die Zielgruppe angepasst werden (bzw. bei nicht beeinflussbaren Rahmenbedingungen kann dies mit den Teilnehmern erörtert werden). Die Endbefragung (einschließlich der Evaluationen durch die Teilnehmer) ergibt Informationen für die längerfristige Planung des Veranstaltungstyps.

Geplant waren in diesem Projekt zusätzlich Gruppendiskussionen mit den Teilnehmern im Folgesemester, sobald die Gesamtauswertungen vorlagen. Diese kamen jedoch nicht zustande. Das Interesse der Studierenden, sich mit der Lehre auseinanderzusetzen, erweist sich schon bei der konkreten Rückmeldung in die laufende Veranstaltung als enttäuschend gering. Sich im Nachhinein noch einmal mit einer bereits absolvierten Veranstaltung abstrakt (mit dem Blick auf die künftige Gestaltung des Veranstaltungstyps für spätere Adressaten) auseinanderzusetzen – dazu fehlt offenbar die Motivation. Dagegen ist die Akzeptanz der nur punktuell auszufüllenden Rückmelde-Fragebögen bei den Studierenden relativ hoch: Mehr als 75 % plädieren für regelmäßige Wiederholungen solcher Befragungen. Allerdings wurde schon ab dem zweiten Semester auch vermehrt Unmut über die „lästige Befragerei“ geäußert. Doch selbst wenn sich auf die Dauer - im Zuge einer „Veralltäglichung“ des Ansatzes - die Beteiligung auf einem niedrigeren Niveau einpendelt, sollte das kein Grund zur Resignation sein. Wenn mit Bemühungen um eine qualitative Verbesserung des Studienservice Lehre nicht ein „Marktanteil“ von 100 % erreichbar ist, reicht dies nicht als Argument für eine nur mittelmäßige Qualität der Dienstleistung Lehre aus. Für die Zielgruppe der Studieninteressierten lohnt sich die Mühe.

Das gleiche Ziel – Qualitätsverbesserung der Lehre in Veranstaltungen durch einen formalisiert geförderten Dialog zwischen Lehrenden und Lernenden – verfolgt ein Vorhaben in einem Reformprojekt „Intensivstudium Psychologie“ am Fachbereich Erziehungswissenschaft und Psychologie der Freien Universität Berlin. In diesem Konzept – „dialogische Evaluation“ genannt – wird der Kommunikation zwischen den am Lernprozess Beteiligten ein noch höheres Gewicht beigemessen als im oben geschilderten Verfahren. Die Erfahrungen werden in einem jetzt vorliegenden Werkstattbericht dokumentiert (Knäuper, Kroeger u.a. 1999).

6.2 Beispiele evaluationsverwertbarer Informationen aus Befragungen in Lehrveranstaltungen

Erhebt man studentische Urteile in Lehrveranstaltungen und möchte die Bewertungen als Qualitätsindikatoren interpretieren (z.B. durch Berechnung von Mittelwerten pro Veranstaltung, s. oben, Abschnitt 1), dann wird man in der Regel mit einem „störenden“ Resultat konfrontiert: Die Studierenden sind sich nicht einig. Was einer Gruppe von Teilnehmern als überzeugende Lehre erscheint, kritisieren andere als absolut untauglich; wieder andere urteilen „teils / teils“ – und alles in derselben Veranstaltung, die doch „objektiv“ für alle identisch ist (s. oben, Abschnitt 2.2 und im Detail Kromrey 1994/1995b). Analysiert man die Fragebögen im Hinblick auf die Teilnehmerstruktur in den Veranstaltungen, dann findet man parallel zur *Heterogenität der Urteile* eine entsprechende Vielfalt von Interessenlagen und Teilnahmegründen, von Erwartungen und Befürchtungen, von Zugehörigkeiten zu Studienphasen und Studiengängen, manchmal sogar zu Fachbereichen/Fakultäten. Die feststellbare Heterogenität wird nicht nur von manchen Lehrenden erheblich unterschätzt (sie haben schließlich ihre Veranstaltung für eine ganz bestimmte Zielgruppe eines bestimmten Studiengangs angekündigt). Auch studentische Interessenvertreter verfallen in den gleichen Fehler, wenn sie sich für *die* (vermeintlichen) Interessen *der* Studierenden einsetzen.

Möchte man die in solchen Kontexten erzielten Befragungsergebnisse *als* Evaluation nutzen, so ist diese Heterogenität subjektiver Bestimmungsgründe der Urteilsfindung (und natürlich der Urteile selbst) ein Indiz für die äußerst zweifelhafte Gültigkeit der Datenbasis. Dies ist jedoch nicht gleichbedeutend damit, dass die Informationen selbst von geringem Wert seien. Im Gegenteil: Sie informieren über wesentliche Randbedingungen für die Lehre und können für die inhaltliche und formale Gestaltung außerordentlich wichtige Anregungen geben, z.B.: Orientierung an vielfältigen Beispielen, bei Übungen Aufteilung in homogene Teilgruppen, bei Gruppenarbeit bewusste Diskussion zwischen unterschiedlichen (jeweils homogenen) Teilgruppen. Heterogenität macht die Lehre schwerer, kann aber durchaus auch positiv als bewusst eingesetztes didaktisches Prinzip genutzt werden. Auch für die Lehrangebots-*organisation* ist eine konkrete Schlussfolgerung naheliegend: Parallelangebote der betreffenden Veranstaltung mit jeweils unterschiedlicher Didaktik; etwa: eine Veranstaltung mit stärkerer Betonung der Eigenaktivität der Studierenden („betreutes Selbstlernen“), eine andere im „herkömmlichen“ Stil für rezeptives Lernen (Vorlesung mit Übung). Wo die personelle Kapazität nicht zur Duplizierung des Angebots ausreicht, kann eine solche Variation über die Semester verteilt stattfinden.

Ein weiterer Auswertungsbefund aus Lehrveranstaltungsbefragungen ist frappierend und wird selten berücksichtigt: der Einfluss der studentischen „peers“ auf die Wahrnehmung und Beurteilung der Lehre, insbesondere derjenigen im selben Hörsaal. Wird eine Veranstaltung überwiegend von desinteressierten Studierenden besucht, werden *alle* Teilnehmer negativ beeinflusst. Im Gegenzug findet sich ein deutlicher „positiver Ansteckungseffekt“, sofern die Mehrheit der Teilnehmer Interesse zeigt. Dies gilt für Urteile über die Lehrdarbietung ebenso wie für die Selbsteinschätzung des eigenen Lernerfolgs oder die Bereitschaft, sich mit dem behandelten Stoff intensiver selbstständig zu beschäftigen (dazu im Detail Kromrey 1994). In Veranstaltungen mit negativer Grundstimmung hat die Lehrperson – wie Befragungen belegen – kaum Chancen, durch eigene Bemühungen „Interesse zu wecken“. Eine genauere vergleichende Analyse über eine Vielzahl von Veranstaltungen hinweg zeigt jedoch: Eine ähnliche Wirkung wie das von Studierenden mitgebrachte eigene *Interesse* am Stoff hat die Einsicht in den *Nutzen* für das weitere Studium oder für den späteren Beruf. Und diese wiederum lässt sich im Kontext des Gesamtfachs durch ein in sich stimmiges und für die Studierenden in seinem Sinn nachvollziehbares Curriculum ebenso fördern wie durch Studienberatung und Orientierungsangebote über die Berufspraxis. Auch dies ein Beispiel dafür, dass die Feststellung einer vermeintlichen „Störgröße“ zu einer wichtigen Information für die Planung werden kann.

6.3 Beispiele evaluationsverwertbarer Informationen aus Befragungen zur Studiensituation in Fachbereichen/Fakultäten

Wie in einzelnen Lehrveranstaltungen, so können Studierende auch in Fachbereichen / Fakultäten oder Studienfächern befragt werden: über ihre soziale Herkunft, ihre Gründe der Wahl von Studienfach und -ort, über Studienmotive, Arbeitsbelastung, Zeitbudget, Berufspläne, die Transparenz von Studienanforderungen, über Einschätzungen des Lehrangebots sowie der Betreuungs- und Prüfungssituation, über die wahrgenommene Abstimmung zwischen Lehr- und Prüfungsinhalten u.ä.m. Und wie in einzelnen Lehrveranstaltungen, ist auch in den meisten Fächern die große Heterogenität ein ins Auge fallendes Charakteristikum der Resultate.

So ergab etwa an der Fakultät für Sozialwissenschaft der Ruhr-Universität Bochum die Vollerhebung aller anwesenden Studierenden in den Lehrveranstaltungen einer Stichtagswoche die folgende Verteilung der Studienziele:

- 44 % mit Studienziel Diplom-Sozialwissenschaft,
- 14 % Lehramt mit sozialwissenschaftlichem Fach,
- 22 % sozialwissenschaftliches Nebenfach in Studiengängen anderer Fakultäten,
- 20 % Studierende anderer Fakultäten ohne sozialwissenschaftliches Fach (in manchen Lehrveranstaltungen machten die Nicht-Sozialwissenschaftler sogar die Mehrheit aus).

Die offizielle Studienplanung orientiert(e) sich – natürlich – weitestgehend am Diplomstudiengang Sozialwissenschaft, also an der Minderheit. Die Mehrheit der anwesenden Studierenden dagegen orientierte ihr Nachfrageverhalten an anderen Motiven und verfolgte anders gelagerte Ziele als das für die Angebotsplanung maßgebliche Leitbild.

Um bei derselben Fakultät zu bleiben: Selbst deren eigentliche Zielgruppe – also die Diplom-Studierenden – hatte alles andere als homogene Studienmotive.

Schon die Information über die Gründe für die Wahl des Fachs überraschte: Nur für eine Minderheit galt – wie sich herausstellte – dieses Fach als „erste Wahl“; für die Mehrheit war es eher eine Übergangs- oder eine Notlösung:

- Rund 30 % woll(t)en eigentlich ein anderes Fach studieren; 22 % konnten ihr Wunschstudium – vor allem Publizistik und Psychologie – wegen Zulassungsbeschränkungen nicht (sofort) realisieren und benutzten die Sozialwissenschaft als „Parkfach“; 8 % hatten andere Gründe.
- Weitere rund 30 % hatten zunächst ein anderes Fachstudium begonnen und es dann abgebrochen (Hauptgründe: zu schwer, zu arbeitsintensiv; Sozialwissenschaft bot sich in räumlicher Nähe als Unterschlupf an, um nicht die Universität ganz verlassen zu müssen).
- Nur rund 40 % hatten die Sozialwissenschaft von vornherein und mit der Absicht des Fachstudiums gewählt!

So kann es nicht verwundern, dass sich in der Gesamtheit der Studierenden die unterschiedlichsten Motivations-Profile repräsentiert fanden. Eine Clusteranalyse der Studiengründe, Abschlussziele und Berufsvorstellungen ergab unter den Befragten neun gut unterscheidbare Gruppen:⁸

⁸ Ganz ähnlich war das Resultat in den anderen 12 Fakultäten, in denen entsprechende Erhebungen durchgeführt wurden. Ausnahmen bildeten lediglich stark „verschulte“ Fächer (hier z.B. Bauingenieurwesen) oder solche mit einer „verbindlicheren“ Fachidentität (hier z.B. Theologie).

Gruppe 1: Breites Studieninteresse (Wissenschaft + Beruf)
(22,0% der Sowi-Diplom-Studierenden, 20,1% der Studierenden anderer Fächer und Studiengänge)

- gewünscht: berufsrelevante *und* fachwissenschaftliche Ausbildung; Wahl von Inhalten nach persönlichen Interessen
- erhofft: Basis für gute berufliche Existenz; noch keine spezifischen Berufsvorstellungen vorhanden

Gruppen 2/3: strikt berufsorientiertes Studieninteresse

2: Beruf hat unbedingten Vorrang (Sowi 9,6%; andere 11,5%)

3: individuelle Lebensgestaltung neben Studium (Sowi 5,0%; andere 7,9%)

- Wahl von Studium und Inhalten nach beruflicher Relevanz (persönliche Interessen ggf. zurückstellen)
- gewünscht: umfassende Kenntnisse über berufliche Anforderungen
- 2: auch Teilhabe an studentischem Leben wird zurückgestellt, individuelle Lebensgestaltung wird dem Berufsziel untergeordnet
- 3: Kontakte mit anderen Studierenden und individuelle studentische Lebensgestaltung sind auch wichtig

Gruppe 4: sozialwissenschaftlich orientierte (zugleich beruflich verwertbare) Bildung
(Sowi-Diplom: 14,7%, andere: 11,9%)

- erhofft: gute Basis für berufliche Existenz; gesicherter Status, Ansehen
- gewünscht: fundierte wissenschaftliche Ausbildung, sozialwissenschaftlich interessierende Inhalte; noch keine Berufsorientierung oder Spezialisierung
- gewünscht: Möglichkeit, eigenen Interessen und Neigungen zu folgen, Leben individuell zu gestalten, studentische Vorzüge zu nutzen

Gruppen 5/6: Wissenschaft bzw. Bildung ohne Berufsorientierung

5: wissenschaftliche Bildung (Sowi 9,6%; andere 16,5%)

6: persönliche Bildung (Sowi 13,8%; andere 7,9%)

- kein Bezug zu späterer beruflicher Existenz; nicht gewünscht: Informationen über berufliche Anforderungen
- Interesse an persönlicher Bildung und an sozialen Problemen
- 5: gewünscht: fundierte *wissenschaftliche* Ausbildung
- 6: Ausbildungs-Interesse ist eher gering (wissenschaftliches noch geringer als berufliches)

Bei aller Unterschiedlichkeit der Qualifikationsziele, die mit dem Studium erreicht werden sollen, ist man sich in den Gruppen 1 bis 6 einig darin, das Studium bis zu einem Examen zu Ende führen zu wollen. Es besteht keine Bereitschaft zum Studienabbruch, falls sich eine gute berufliche Chance bietet, für die ein Studienabschluss nicht erforderlich ist.

Gruppen 7-9: kein Interesse an Studienabschluss ("Aufbewahrungsfunktion" der Hochschule)

7: Studium als Lebensstil (Sowi 9,2%; andere 8,3%)

8: "Null Bock" (Sowi 8,3%; andere 7,6%)

9: Studium als Überbrückung zur Berufsfindung (Sowi 7,8%; andere 8,3%)

- wichtig für alle drei Gruppen: Status als immatrikulierte Person; Nutzung der Vorteile des Studentenausweises
- nicht gewünscht: Teilhabe an studentischem Leben, Kontakte mit anderen Studierenden
- 7: Vorzug: individuelle Lebensgestaltung; es wird z.Z. keine Alternative zum Student-Sein gesehen; keine Bereitschaft zum Studienabbruch bei Berufschance
- 8: keinerlei Interesse am Studienfach, kein Interesse an wissenschaftlicher Ausbildung; dennoch: Hoffnung, dass das Student-Sein berufliche Chancen fördert; Bereitschaft zu Studienabbruch, falls sich eine Berufschance bietet
- 9: geringes Ausbildungsinteresse; zwar berufs-, aber nicht abschlussorientiert; Beschäftigung mit ausgewählten, interessierenden Inhalten des Fachs; sofortiger Studienabbruch bei Berufschance

Es dürfte unmittelbar erkennbar sein, dass eine Lehrangebotsplanung, die allen Studierenden-Gruppen gleichzeitig gerecht werden könnte, schlechterdings unmöglich ist. Selbst wenn die nicht im eigentlichen Sinne studierwilligen Immatrikulierten (Gruppen 8-10)⁹ außer Betracht bleiben, existieren unterschiedliche bis gegensätzliche Anforderungsprofile: von strikter Berufsorientierung (Gruppen 2 und 3) bis zu starker Bildungsorientierung (Gruppen 5 und 6), von eng sozialwissenschaftlicher Orientierung (Gruppe 4) bis zu breitem Interesse an Bildung und Beruf (Gruppe 1).

Will eine Fakultät / ein Fachbereich die von hochschulpolitischer Seite erhobene Forderung einlösen, Studiengänge zu konzipieren und zu realisieren, die innerhalb einer angebbaren Zahl von Semestern (Regelstudienzeit) „studierbar“ sind, dann wird sie unter der gegenwärtigen Bedingung knapper Kassen nicht das gesamte Spektrum von Anforderungsprofilen der *Studierenden* (wobei dieser Begriff hier im wörtlichen Sinne gemeint ist, also die eher Studierunwilligen bewusst ausklammert) bedienen können. Das im vorigen Abschnitt (6.2) formulierte Postulat, dass die Entwicklung von Lehrqualität in einzelnen Lehrveranstaltungen jeweils nur zielgruppenorientiert erfolgen könne, gilt in gleicher Weise für das Curriculum und das Studienangebot eines ganzen Faches. Dies geschieht natürlich seit jeher, und zwar durch Orientierung an einem (expliziten oder zumindest impliziten) idealtypischen Leitbild: breites Studieninteresse auf Seiten der Nachfrager, Berufsqualifizierung durch Wissenschaft auf Seiten des Angebots. Befragungen wie die hier als Beispiel angeführten liefern allerdings empirisch fundiertes Wissen über die „Nachfrageseite“ und schützen vor Fehlentscheidungen bei Planungen, die ausschließlich nach bestem Gutdünken „am grünen Tisch“ erfolgen.

⁹ Dieser Personenkreis macht unter den Befragten immerhin gut 25% aus. Unter den Immatrikulierten insgesamt dürfte er allerdings erheblich höher liegen, da die Befragung im Semester in einem Querschnitt von Lehrveranstaltungen durchgeführt wurde und daher prinzipiell nur diejenigen erfassen konnte, die zumindest in der Universität anwesend waren.

7. Literatur

- 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, S. 41-49
- Donabedian, A., 1980: Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment, Ann Arbor, MI
- Frey, Siegfried; Frenz, Hans-G., 1982: Experiment und Quasi-Experiment im Feld. In: Patry, J.-L. (Hg.): Feldforschung, Bern, Stuttgart, S. 229-258
- HRK Hochschulrektorenkonferenz (Hg.), 1998: Evaluation. Sachstandsbericht zur Qualitätsbewertung und Qualitätsentwicklung in deutschen Hochschulen. Dokumente & Informationen 1/1998, Bonn: HRK
- 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>).
- Kromrey, Helmut, 1994: Wie erkennt man „gute Lehre“? Was studentische Vorlesungsbefragungen (nicht) aussagen. In: Empirische Pädagogik, Jg. 8, H. 2, S. 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, S. 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., S. 105-127
- Kromrey, Helmut, 1999: Von den Problemen anwendungsorientierter Sozialforschung und den Gefahren methodischer Halbbildung. In: SuB Sozialwissenschaften und Berufspraxis, Jg. 22, H. 1, S. 58-77
- 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, S.105-119
- Landfried, Klaus, 1999: Qualitätssicherung als Aufgabe wettbewerbsfähiger Hochschulen. In: HRK (Hg.): Ein Schritt in die Zukunft. Qualitätssicherung im Hochschulbereich. Beiträge zur Hochschulpolitik 3/1999, Bonn: HRK, S. 7-13
- 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, S. 21-46
- Schmidt, Jörn, 1980: Evaluation als Diagnose. HDZ-Dozentenkurs, Essen
- 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, S. 243-249

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, S. 45-54

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