

**Helmut Kromrey**

**Wie erkennt man "gute Lehre"?**

**Was studentische Vorlesungsbefragungen (nicht) aussagen**

|  |
|--|
| veröff. in: Empirische Pädagogik, 1994, Jg. 8, H. 2, 153-168 |
|--|

#### VORBEMERKUNGEN

Seit einiger Zeit geistert ein Phänomen durch die hochschulpolitische Diskussion: die "gute Lehre". Ihr wird entscheidende Bedeutung zugeschrieben im Zusammenhang mit aktuellen Problemen des Studierens an den Hochschulen: lange (überlange) Studienzeiten, hohe Fachwechsler- und Abbrecherquoten, überfüllte Hörsäle zu Semesterbeginn, aber nur noch teilweise gefüllte Hörsäle in der zweiten Semesterhälfte, Desorientierung von Studierenden im Hinblick auf Leistungsanforderungen und effektive Studiengestaltung, unzureichende (oder als unzureichend empfundene) Studien- und Prüfungsleistungen u.v.m. Durch eine einschneidende Verbesserung der "Qualität der Lehre" ließen sich - so die verbreitete Überzeugung - manche dieser Probleme spürbar lindern, wenn nicht gar beseitigen. Denn daß es mit der Lehre weithin nicht zum besten stehe, daß sie geradezu "bodenlos schlecht" sei, das scheint in der öffentlichen Meinung nicht ernsthaft umstritten zu sein. So liest man etwa im SPIEGEL 16/1993 (S. 86): "Ein Großteil der Hochschullehrer in Deutschland ist unfähig, gut zu lehren." Und: "Lehrflaschen demotivieren ungezählte junge Menschen und richten damit einen kaum wieder gutzumachenden Schaden an." Was läge da näher als durch geeignete Anreiz- und Kontrollmechanismen die Lehrenden zu einer besseren Erfüllung ihrer Aufgaben zu motivieren und ihnen durch geeignete Informationen aufzuzeigen, an welchen Stellen und auf welche Weise Verbesserungen möglich sind.

Erste Voraussetzung hierfür wäre allerdings, "gute Lehre" überhaupt als solche zu erkennen und sie von "mittelmäßiger" oder gar "schlechter" Lehre abgrenzen zu können. Hier nun ist die empirische Sozialwissenschaft (nicht nur die Psychologie) gefordert: Sie soll durch "Evaluation der Lehre" diese Informationen - in möglichst "objektiver" Form - bereitstellen. Sie soll dies zudem mittels eines Verfahrens leisten, das flächendeckend mit geringem Aufwand und mit geringen methodischen Vorkenntnissen einsetzbar ist und dessen Ergebnisse möglichst schnell vorliegen.

Hier aber liegt das Problem. Selbstverständlich verfügt die Sozialforschung über ein geeignetes Design zur empirischen Bewertung (auch von Lehre), nämlich das der Programm-Evaluation. Es ist allerdings methodisch aufwendig, verlangt vom Durchführenden Forschungserfahrung, kostet viel Zeit, setzt u.a. relativ genaue Kenntnisse über Ziele, Instrumente und Zielgruppen des Programms sowie den Konsens über geeignete Beurteilungskriterien voraus, muß selbstverständlich auf die Inhalte des zu bewertenden Programms (hier: der Lehre in der jeweiligen Veranstaltung) zugeschnitten werden und wäre flächendeckend kaum ohne Störung des Lehrbetriebs anwendbar. ♣ Ein solcherart differenzierender Forschungsansatz ist daher verständlicherweise nicht gemeint, wenn in Aktionsprogrammen zur "Qualität der Lehre" auch deren "Evaluation" vorgesehen ist. Gedacht wird vielmehr an möglichst generelle Kriterien für "Qualität", an Kriterien, die gerade stoff- und fach- und lernzielunabhängig anwendbar sein sollen. Die Lösung dieser an sich unlösbaren Aufgabe liegt für die Apologeten von Lehrevaluation auf der Hand: Man befrage die Studierenden in den jeweiligen Lehrveranstaltungen und erhebe deren Bewertungen; dies kann auch anhand kurzer Fragebögen mit relativ globalen (also:

inhaltsunabhängigen) Beurteilungen geschehen. Die Studierenden - so kann argumentiert werden - sind als die von der Lehre ganz konkret "Betroffenen" in der Lage, aus eigener Erfahrung auch deren Qualität zuverlässig zu beurteilen.

An die Stelle von Evaluation als methodisch kontrolliertem empirischem Verfahren tritt so die Erhebung von Akzeptanzaussagen per standardisierter Befragung. Die damit verbundene Erwartung ist, daß in Veranstaltungen mit "schlechter" Lehre auch die studentischen Akzeptanzaussagen in der Tendenz negativ, in Veranstaltungen mit "guter" Lehre in der Tendenz positiv ausfallen. Berechnet man statistische Kennziffern pro Veranstaltung (etwa Mittelwerte oder Prozentanteile), dann kommen individuelle Abweichungen der einzelnen Urteilenden darin nicht mehr zur Geltung; man erhält verlässliche Qualitätsurteile.♣

Leider erweisen sich solche Vorstellungen als empirisch falsch, insbesondere wenn global zusammenfassende Urteile vom Typ "Zufriedenheit" erfragt werden; darauf wird kurz im ersten Teil des folgenden Textes eingegangen. Danach werden Überlegungen vorgestellt, ob ggf. andere per Befragung erhebbare Indikatoren geeignet sein könnten, als Qualitätsmaß zu fungieren. Den Abschluß bilden resümierende Ausführungen zur Aussagekraft studentischer Akzeptanzerhebungen sowie Empfehlungen für ein alternatives Vorgehen.

## DURCHSCHNITTSWERTE STUDENTISCHER URTEILE ALS "QUALITÄTSINDIKATOR" VON EINGESCHRÄNKTER VALIDITÄT

Differenzierte Auswertungen studentischer Teilnehmerbefragungen in Vorlesungen♣ erweisen einerseits den hohen Informationswert solcher Erhebungen als Rückmelde-Instrument an die jeweiligen Lehrenden, legen andererseits aber auch die große Gefahr von Fehlschlüssen offen, falls die studentischen Aussagen ungeprüft als Qualitäts-Indikatoren interpretiert werden. Insbesondere drei Fehlerquellen sind es, deren Nichtberücksichtigung nahezu zwangsläufig zu Interpretations-Artefakten führen muß: die Gleichsetzung jeder bewertenden Aussage mit "Evaluation", die schematische Berechnung von Mittelwerten pro Veranstaltung und die Unterstellung der Unabhängigkeit von Urteilen auf unterschiedlichen Bewertungsdimensionen.

Die Erhebung bewertender Aussagen ist noch keine Evaluation

Das zentrale Mißverständnis besteht in der häufig zu beobachtenden unreflektierten Gleichsetzung der Antworten auf bewertende (also evaluative) Fragen mit "Evaluation" als methodisch kontrollierter, empirischer Qualitätsbewertung. Ermittelt wird hier lediglich - wie einleitend bereits ausgeführt - die "Akzeptanz" (oder Nicht-Akzeptanz), auf die der beurteilte Sachverhalt bei den Befragten stößt. Dies ist keine unwesentliche Information, insbesondere nicht für die Durchführenden von Lehrveranstaltungen; denn Akzeptanz (oder ihr Fehlen) hat selbstverständlich Auswirkungen auf Lernprozeß und Lernresultat. Aber sie ist nicht (bzw. allenfalls in sehr spezifischen Konstellationen) gleichbedeutend mit Evaluation. Der Grad der Akzeptanz kann von den verschiedensten Bedingungen abhängen, nicht nur von der Qualität des in Frage stehenden Sachverhalts. Bei der Teilnehmerbeurteilung von Lehrveranstaltungen z.B. - so zeigt die Datenanalyse - hängen die Antworten in erster Linie von der generellen Beliebtheit (bzw. Unbeliebtheit) des behandelten Stoffs ab. Das heißt: Werden Durchschnittsurteile (insbesondere von global zusammenfassenden Urteilen,

von sog. Zufriedenheitsfragen) zwischen Veranstaltungen mit unterschiedlichem Stoff miteinander verglichen, geben die Differenzen an erster Stelle Unterschiede in der Akzeptanz des Lehrstoffs wieder, nicht Unterschiede in der Lehrqualität. An zweiter Stelle folgen individuelle Teilnahmegründe: Der Zwang einer "drohenden" Klausur, die Auseinandersetzung mit nicht abwählbarem Prüfungsstoff führen zu negativen, das vorab vorhandene Eigeninteresse am Stoff, die Existenz zumindest von Wahlmöglichkeiten zwischen Alternativen innerhalb eines Pflichtbereichs, die erkennbare Verwertbarkeit für eigene Studienziele führen zu positiven Urteilen. Das heißt: Werden die Durchschnittswerte zwischen Veranstaltungen verglichen, in denen der behandelte Stoff den gleichen Grad an genereller Beliebtheit (bzw. Unbeliebtheit) genießt, dann hängen diese vor allem von der Zusammensetzung der Teilnehmerschaft ab - immer noch nicht von der Lehrqualität. Sind die Teilnehmer überwiegend unfreiwillig, ohne eigenes Interesse in der Veranstaltung? Hatten sie keine Möglichkeit, das in ihren Augen "kleinere Übel" zu wählen? Oder sind sie in der Mehrheit intrinsisch motiviert, können sie den Lehrstoff für ihre eigenen Studieninteressen nutzen? Aus den Antworten auf diese Fragen folgen drastische Unterschiede in den mittleren Akzeptanzwerten. Weitere Einflüsse auf das Antwortverhalten haben das individuelle Studierverhalten (regelmäßige Teilnahme, Vor-/Nacharbeit) und die Studienerfahrung (Erstsemester vs. Hauptstudium). Erst wenn auch in dieser Hinsicht die Teilnehmerstrukturen in den zu vergleichenden Veranstaltungen näherungsweise übereinstimmen, erst dann darf Akzeptanz im Sinne von Qualitätsevaluation interpretiert werden. ♣

Die Berechnung von Mittelwerten setzt geeignete Daten voraus

Bei dem in Lehrevaluations-Studien verbreitet anzutreffenden schematischen Berechnen des arithmetischen Mittels pro Beurteilungs-Item und Lehrveranstaltung werden häufig einfachste methodische Voraussetzungen nicht beachtet (selbst wenn man die Antworten auf den Beurteilungs-Skalen als metrisch skaliert unterstellt). Bekanntlich ist ein Mittelwert nur dann sinnvoll interpretierbar, wenn sich die einzelnen Beobachtungswerte mehr oder weniger stark in einem "typischen" Wertebereich konzentrieren. Ist diese Voraussetzung gegeben, sinkt der Informationswert allerdings mit zunehmender Streuung. Die Vorlesungsbefragungen zeigen nun, daß im Normalfall (und zwar in fast jeder Veranstaltung) die Einschätzungen der Befragten im Hinblick auf fast jedes Beurteilungs-Item sehr stark variieren: Der Informationswert des arithmetischen Mittels ist also gering. Darüber hinaus ist relativ oft die grundlegendere Voraussetzung der Konzentration der Antworten in einem "typischen Bereich" nicht erfüllt; in manchen Lehrveranstaltungen kommt es zu einer ausgesprochenen Polarisierung der Teilnehmerurteile: Der eine Teil der Befragten äußert sich hinsichtlich des erfragten Aspekts positiv ("gut" bis "sehr gut"), der andere Teil negativ ("ausreichend" bis "mangelhaft"). Ein aus diesen beiden "Lagern" gebildetes Durchschnittsurteil "mittelmäßig" oder "befriedigend" hätte in diesem Fall nicht lediglich geringen Informationswert, sondern wäre schlichtweg irreführend und in keiner Weise gleichbedeutend mit einer Veranstaltung, in der die Mehrheit der Teilnehmer mittelmäßige Akzeptanzurteile angekreuzt hat.

Teil-Urteile fügen sich nicht zu einem widerspruchsfreien Gesamt-Urteil zusammen

Vorschläge für ein leicht handhabbares und schnell auswertbares Befragungs- "Evaluationsverfahren" setzen nicht nur auf die Berechnung "einfacher" Mittelwerte, sondern ebenso auf kurze Fragebögen mit möglichst wenigen global zusammenfassenden Urteilen (etwa: Gesamtbewertung der Didaktik oder der Aufbereitung des Lernstoffs oder des persönlichen Auftretens des/der Lehrenden oder gar Zufriedenheit mit der Lehrveranstaltung insgesamt). Ähnlich sind Vorschläge einzuschätzen, zunächst

differenzierter erhobene Einzelurteile im Zuge der Auswertung zu einem Index zusammenzufassen (etwa Addition aller auf die Didaktik bezogenen Einzelitems, dividiert durch die Anzahl dieser Items). Die Mittelwerte pro Veranstaltung bei Erhebung von Globalurteilen sind nahezu identisch mit den Mittelwerten entsprechender Indices, was für die vereinfachte Erhebung von Globalurteilen zu sprechen scheint. Differenziertere Analysen zeigen jedoch, daß auch deren Gültigkeit außerordentlich zweifelhaft ist.

Für einen Teil der Befragten (nahezu die Hälfte) gilt: Die Antworten auf die einzelnen Items werden nicht unabhängig voneinander getroffen; es existiert ein beträchtlicher "Ausstrahlungseffekt". Mit anderen Worten: Die Antworten auf die Detailfragen sind Indikatoren für ein dahinterstehendes "Urteilssyndrom"; dieses kann von "in jeder Hinsicht zufrieden" bis "in jeder Hinsicht unzufrieden" reichen. Ist man generell eher zufrieden (bzw. unzufrieden), werden Qualitätsabweichungen bei dem einen oder anderen Teilaspekt weder kritisiert noch honoriert. Für einen insgesamt eher Unzufriedenen ist - beispielsweise - auch der objektiv vorbildliche Medieneinsatz kein positives Urteil wert; für einen insgesamt Zufriedenen dagegen ist auch ein objektiv völlig unzureichender Medieneinsatz immer noch ein (moderates) Detail-Lob wert. Dies erweist sowohl das Verfahren der Clusteranalyse als auch die Auswertung mittels Latent-class-Modellen.

Allerdings ist nur ein Teil der Urteilsprofile eindeutig auf einer Positiv-Negativ-Dimension anzuordnen. Etwas mehr als die Hälfte zeichnet sich durch Urteilmuster aus, deren Verläufe sich überschneiden. Zum Beispiel: positive Einschätzungen für Aspekte des sprachlichen Vortrags (Verständlichkeit, Tempo, Themenzentrierung) und Stoffaufbereitung (Schwierigkeitsgrad, Menge), aber negative Urteile über die Strukturierung der Veranstaltung (erkennbare Gliederung und Struktur, Klarheit der Lernanforderungen), das persönliche Engagement des/der Lehrenden und die Aufgeschlossenheit für Belange der Studierenden. Oder auch umgekehrt: negative Einschätzungen für sprachlichen Vortrag und Stoffaufbereitung, positive Urteile für persönliches Engagement und Aufgeschlossenheit der Dozenten. Im Durchschnitt über alle Einzeleinschätzungen wären in diesem Fall beide Urteilsprofile - obwohl in jeder Hinsicht aus entgegengesetzten Detailurteilen bestehend - identisch, nämlich "mittelmäßig".

Bei der ersten Gruppe von Befragten "homogenisiert" also der individuelle Gesamteindruck der Veranstaltung die Detailschätzungen; diese haben nur noch einen geringen eigenen Informationswert. Dagegen urteilen die Befragten der anderen Gruppe differenziert. Für sie ist dann jedoch ein die Details einbrennendes Gesamturteil nicht mehr angemessen; es wird weitgehend informationsleer. Erhebt man lediglich zusammenfassende Globalurteile, kann den Daten die jeweils unterschiedliche Informationsqualität nicht mehr angesehen werden. Erst eine sowohl detaillierte Akzeptanzerhebung als auch differenzierte Auswertung liefert interpretierbare Daten.

Die hier nur kurz skizzierte Heterogenität der Einzelurteile und der Urteilsprofile (ausführlicher s. KROMREY 1993a/b, 1994) gilt im übrigen nicht nur zwischen den verschiedenen Lehrveranstaltungen, sondern sie gilt innerhalb fast jeder Lehrveranstaltung. Relativ homogene Akzeptanzurteile in einer Veranstaltung sind die ganz seltene Ausnahme. Die Heterogenität wird umso größer, je unterschiedlicher sich die Teilnehmerstruktur darstellt. Je heterogener jedoch die Urteile sind, desto geringer sind die Chancen von Lehrenden, in ihrer Veranstaltung durch Veränderungen für die Mehrheit der Teilnehmer subjektive Verbesserungen herbeizuführen; jede Veränderung zugunsten des kritischen Teils der Hörer ist zugleich mit dem Risiko behaftet, für den bisher zufriedenen Teil

subjektiv etwas zu verschlechtern - ein Dilemma, zu dessen Lösung Patentrezepte nicht existieren.

## MOTIVIERUNG DER STUDIERENDEN ALS GEEIGNETERER INDIKATOR FÜR "GUTE LEHRE"?

Die Berechtigung einer Gleichsetzung von "Zufriedenheit der Teilnehmer in Lehrveranstaltungen" (bzw. generell von Akzeptanz) mit "Qualität der Lehre" darf jedoch über die bisher dargelegten Gründe hinaus grundsätzlich in Frage gestellt werden. Natürlich ist die Akzeptanz eines Lehrangebotes nicht unbedeutend; eher jedoch dürfte gute Lehre am Lernerfolg der Studierenden zu messen sein. Gute Lehre müßte sich dadurch auszeichnen, daß auch Unbeliebtes und damit generell auf geringere Akzeptanz Stoßendes (aber für das Studium Notwendiges) gelernt wird.

Leider kann Lehre den Lernerfolg nicht direkt herbeiführen; der von manchen ersehnte "Nürnberger Trichter" existiert nicht. Lernerfolg kann sich vielmehr - das dürfte unbestritten sein - nur einstellen, wenn zuvor von den Studierenden aktiv gelernt wird. In diesem Zusammenhang wird eine weitere der (wenigen) Erkenntnisse relevant, über die Einigkeit bestehen dürfte, nämlich daß der Erfolg von Lernbemühungen dann größer ist, wenn diese Lernbemühungen gern unternommen werden, wenn also das zu Lernende "interessant" erscheint. Es ist daher verständlich, wenn von Lehrenden verlangt wird, sie sollten bei den Lernenden Interesse (oder auch: Motivation) "wecken". Was aber kann diese Forderung bedeuten? Da in der Hochschule die Lehrenden immer mit mehr oder weniger großen Gruppen von Studierenden gleichzeitig konfrontiert sind, kann sie sich sinnvollerweise nicht auf die dispositionale Eigenschaft "individuelle Interesse" oder "Interessiertheit" der jeweiligen Lehrveranstaltungs-Teilnehmer beziehen. Gemeint sein kann nur das auf die Lernsituation bezogene Pendant "Interessantheit" der Lehre (oder: "situationales Interesse"): Der Stoff ist so zu präsentieren, daß er "interessant" wirkt, daß die Lernsituation potentiell beim Lernenden Interesse für den zu lernenden Stoff oder für die zu erlernenden Fertigkeiten hervorruft. Der Anteil derjenigen Veranstaltungsteilnehmer, bei denen dies gelungen ist, wäre dann - so könnte man argumentieren - ein geeigneter, durch Befragung ermittelbarer Indikator für gute Lehre.

Welche Chancen hat nun ein Lehrender, durch die Lehre "Interesse zu wecken"? Bekannt ist, daß situationales Interesse bei denjenigen Studierenden relativ leicht herzustellen ist, die bereits "interessiert" sind, die also mit bereits vorher vorhandenem "individuellem Interesse" die Lehrveranstaltung besuchen. Eine erheblich schwerere Aufgabe ist es, situationales Interesse bei solchen Studenten zu wecken, die am Lerngegenstand (zunächst) nicht individuell interessiert sind, die nur deshalb die Lehrveranstaltung besuchen, weil sie dies müssen, weil z.B. eine Klausur oder andere Prüfung droht. Auf eine detaillierte Darstellung kann hier verzichtet werden; vgl. dazu die illustrative Zusammenfassung des Forschungsstands bei KRAPP (1992). Wichtig ist im gegebenen Zusammenhang lediglich, daß sich die (vorab vorhandene) Motivation zur Teilnahme an der Lehrveranstaltung erheblich auf das individuelle Erleben des Grades an "Interessantheit" (situationales Interesse) und damit auf die Einschätzung der Befragten auswirkt, ob die Lehre bei ihnen "Interesse geweckt" (oder auch: sie zur weiteren Beschäftigung mit dem Gegenstand der Lehre motiviert) habe. Die folgende Tabelle belegt den statistischen Zusammenhang zwischen einem Index "Teilnahmemotivation"♣ und den Antworten auf eine Einschätzungsfrage "Förderung von Interesse".♣





gezwungenermaßen (weil es die Studien- und Prüfungsordnung so vorschreibt) die Veranstaltung besuchen. Bei jedem Vierten von ihnen (27,4%) haben die Lehrenden dennoch "Interesse geweckt". "Gute Lehre" hätte man also in einer Veranstaltung zu vermuten, in der die Dozentin oder der Dozent einen hohen Anteil der zunächst Desinteressierten "konvertiert", d.h. für den Lerngegenstand interessiert hat.

Allerdings hat Tab.2 gezeigt, daß das durch die studentische Teilnehmerstruktur in der Veranstaltung geschaffene "Motivationsklima" einen erheblichen Einfluß auf den potentiellen Erfolg der Lehre zu haben scheint. Leider bestätigt sich dieser Eindruck auch bei Kontrolle der individuellen Teilnahmemotivation der Befragten. Ist es schon generell schwer genug, zunächst für den Lerngegenstand nicht interessierte Studierende durch Lehrbemühungen zu motivieren, so scheint dies bei weitgehend negativ eingestellter sozialer (hier: studentischer) Lernumwelt fast gar nicht mehr möglich zu sein. Tab.3 illustriert dies im Detail.

Tab. 3: Förderung von Interesse und vorherrschende Teilnahmemotivation in der Vorlesung (nur Befragte mit "rein extrinsischer" Teilnahmemotivation)

|                 |  | *VORHERRSCHENDE TEILNEHMERMOTIVATION |          |          |          |         |  |
|-----------------|--|--------------------------------------|----------|----------|----------|---------|--|
|                 |  | * > 90 %                             | > 80 %   | gemischt | > 70 %   |         |  |
|                 |  | * extrins.                           | extrins. |          | intrins. | Zeile   |  |
| INTERESSE       |  | * -2                                 | * -1     | * 0      | * 1      | * Total |  |
| GEFÖRDERT?      | ))))))3))))))3))))))3))))))3))))))3))))))1 |                                      |          |          |          |         |  |
|                 | -2   | * 367                                | * 650    | * 399    | * 43     | * 1459  |  |
| überhaupt nicht | *  | 43,5%                                | * 32,1%  | * 15,4%  | * 12,6%  | * 25,1% |  |
|                 | /))))))3))))))3))))))3))))))3))))))1       |                                      |          |          |          |         |  |
|                 | -1   | * 213                                | * 441    | * 500    | * 59     | * 1213  |  |
|                 | *  | 25,2%                                | * 21,8%  | * 19,3%  | * 17,4%  | * 20,9% |  |
|                 | /))))))3))))))3))))))3))))))3))))))1       |                                      |          |          |          |         |  |
|                 | 0  | * 170                                | * 472    | * 794    | * 101    | * 1537  |  |
| teils/teils     | *  | 20,1%                                | * 23,3%  | * 30,6%  | * 29,7%  | * 26,5% |  |
|                 | /))))))3))))))3))))))3))))))3))))))1       |                                      |          |          |          |         |  |
|                 | +1   | * 90                                 | * 427    | * 828    | * 119    | * 1464  |  |
|                 | *  | 10,7%                                | * 21,1%  | * 31,9%  | * 35,0%  | * 25,2% |  |
|                 | /))))))3))))))3))))))3))))))3))))))1       |                                      |          |          |          |         |  |
|                 | +2   | * 4                                  | * 36     | * 72     | * 18     | * 130   |  |
| ja, sehr        | *  | 0,5%                                 | * 1,8%   | * 2,8%   | * 5,3%   | * 2,2%  |  |
|                 | .))))))2))))))2))))))2))))))-              |                                      |          |          |          |         |  |
| Spalte          |  | 844                                  | 2026     | 2593     | 340      | 5803    |  |
| Total           |  | 14,5%                                | 34,9%    | 44,7%    | 5,9%     | 100 %   |  |

Von den Studierenden, die ohne vorab vorhandenes Eigeninteresse eine Vorlesung besuchen und die sich in ihrer Einstellung durch das eindeutig vorherrschende Desinteresse auch der Kommiliton/innen bestätigt sehen (mehr als 90% "rein extrinsische" Hörer), können durch die Lehrbemühungen nur etwas mehr als 10% motiviert werden (darunter praktisch niemand - 0,5% - "sehr"); demgegenüber bleiben nahezu 70% bei ihrer explizit kritischen Haltung: "nicht interessiert" (43,5% "überhaupt nicht"). Erst bei gemischter Teilnahmemotivation werden Lehrbemühungen in deutlicherem Maße erfolgreich: Jeweils etwa ein Drittel der zunächst nicht Interessierten urteilt eindeutig positiv bzw. bleibt bei negativer Einschätzung bzw. antwortet "teils/teils". Bei mehrheitlich intrinsisch motivierter Hörerschaft überwiegt dann auch unter den zunächst Desinteressierten das positive gegenüber dem negativen Urteil (40,3% im Vergleich zu 30,0%).

Dieser bei der Gruppe der Nur-Pflicht-Hörer gewonnene Befund wird erhärtet bei entsprechender Auswertung der Urteile in entgegengesetzter Konstellation: "rein intrinsisch" motivierte Teilnehmer in mehrheitlich desinteressierter Umgebung. ♣ Von ihnen bleiben nur noch gut die Hälfte bei positiven Bewertungen, während jeder Dritte (33,0%) zu einer explizit negativen Einschätzung "konvertiert" wird (Skalenwerte -2 bzw. -1). Zum Vergleich: In gleicher Lernumwelt änderten von den Nur-Pflicht-Hörern lediglich 11,2% ihre Einstellung ins Positive (Tab.3). Noch stärker fallen solche "Konvertierungseffekte" in mehrheitlich positiv eingestellten studentischen Lernumwelten ins Gewicht. Gut 40% der zunächst Desinteressierten finden in solchen Lehrveranstaltungen Interesse am Stoff (Tab.3); in gleicher Lernumwelt äußern sich dagegen nur 4,2% der intrinsisch motivierten Hörer negativ. Fazit: Der "Ansteckungseffekt" der sozialen Lernumwelt überdeckt eindeutig den Einfluß unterschiedlicher Lehrpersonen. Dominant negative ebenso wie dominant positive studentische Lernumwelten scheinen für den Lehrerfolg entscheidender zu sein als die Lehre selbst: "Freiwillige" Lehrveranstaltungsteilnehmer in dominant intrinsisch motivierter Lernumgebung vergeben auf die Frage "Interesse gefördert?" im Durchschnitt einen Skalenwert von +1,3 Punkten (arithmetisches Mittel); das mittlere Urteil von Nur-Pflicht-Hörern in dominant extrinsischer Umgebung beträgt dagegen -1,0 Punkte!

#### NOCH EINMAL: WIE ERKENNT MAN GUTE LEHRE?

Kommen wir zurück zu der Eingangsfrage: "Wie erkennt man gute Lehre?" Es zeigte sich, daß der Versuch, die üblichen Akzeptanzurteile durch einen plausibel erscheinenden, mittels Lehrveranstaltungs-Umfragen erhebbaren alternativen Indikator zur Beurteilung von Lehrqualität zu ersetzen, erfolglos blieb. Bei allem sonstigen Ertrag, den Studentenbefragungen für die Gestaltung von Lehre und Studienbedingungen erbringen: als Maßstab für Lehrqualität scheinen sie wenig geeignet. Es drängt sich die Schlußfolgerung auf: Unter den vorherrschenden Lehr- und Lernverhältnissen an den Universitäten ist gute Lehre per Befragung kaum erkennbar. Überwiegend negativ gestimmte Lernumwelten - und dies ist für viele Fakultäten die Normalsituation - machen gute Lehre unsichtbar, machen sie sogar in ihrer Wirkung zunichte. Überwiegend positiv gestimmte Lernumwelten wiederum machen schlechte Lehre unsichtbar, kompensieren sogar ihre Wirkung. Erst eine in dieser Hinsicht "neutrale" Lernumwelt - die also durch eine Mischung unterschiedlicher Motivkonstellationen der Teilnehmer geprägt ist - läßt sowohl die Sichtbarkeit als auch die Wirkung der Qualität der Lehre zur Geltung kommen.

Dieser Befund sollte durchaus nicht als Empfehlung zur Absage an Lehrbeurteilungen mißverstanden werden. Die vorgetragene Kritik an ihrer Aussagekraft als Evaluation bedeutet selbstverständlich nicht, daß Veranstaltungsumfragen generell einen geringen Informationswert oder generell eine nur zweifelhafte Gültigkeit hätten. Sie liefern im Gegenteil - wie auch andere Daten der Umfrageforschung -, sofern sie auf methodisch angemessenem Niveau erhoben und hinreichend differenziert ausgewertet werden, wertvolle Informationen sowohl für die Hochschule als Institution wie für die Lehrenden wie auch für die Studierenden. Erst eine so gewonnene verlässliche empirische Basis bietet günstige Voraussetzungen für die (notwendige) sachliche Diskussion zwischen allen Beteiligten (sowohl innerhalb der einzelnen Lehrveranstaltungen als auch auf anderen Ebenen). Verlässlich und differenziert gewonnene und ausgewertete Akzeptanzdaten liefern wertvolle Hinweise auf Ansatzpunkte für Bemühungen um Verbesserungen von Lehre und Studium - sie sind aber nicht Evaluationen im Sinne von Qualitätsurteilen.

Die vorgestellten Befunde sollten über diese Einsicht hinaus allerdings auch den Blick dafür schärfen, daß die Chancen der einzelnen Lehrenden begrenzt sind, durch individuelle Bemühungen um Verbesserung ihrer eigenen Lehre Grundsätzliches zur Verbesserung der Qualität des Studiums beizutragen. Die Konzentration der Diskussion allein auf die Evaluation der Lehre in einzelnen Lehrveranstaltungen bedeutet eine unangemessene Verengung der Perspektive, die wieder auszuweiten ist auf Lehre und Studium inclusive der Rahmenbedingungen von beidem. Hierzu tragen veranstaltungsübergreifende Befragungen der Studierenden über das Lehr- und Lernprogramm einer Fakultät und dessen Realisierung, über Studiengründe und -ziele, den Informationsstand und die Informationsweisen der Studierenden, ihren Lern- und Arbeitsstil, aber auch über ihre individuelle Studien- und Lebenssituation erheblich mehr bei als einzelveranstaltungsbezogene Umfragen. ♣

Als Fazit der vorgestellten Befunde drängen sich einige Empfehlungen auf: Man schaue nicht nur fasziniert auf die Lehre innerhalb der einzelnen Veranstaltungen, sondern bemühe sich zuerst um die Gestaltung des gesamten Programms eines Studiengangs. Wenn mehr als die Hälfte der Hörer keinerlei inhaltliches Interesse an den Lehrinhalten aufbringt und Veranstaltungen einzig aus dem Grunde besucht, weil diese von der Studien- und Prüfungsordnung bindend vorgeschrieben sind, dann ist dieses eine alarmierende Situation. Man erhöhe - wo immer möglich - die Wahlmöglichkeiten innerhalb von Lernstoffbereichen; die Analyse der Befragungsdaten zeigt, daß selbst die Möglichkeit der Wahl des (subjektiv) "kleineren Übels" die Lernmotivation verbessert. Man verbessere nicht zuletzt auch das "Image" des Lehrangebots. Es wird natürlich immer beliebte und weniger beliebte Studieninhalte in jedem Fach geben. Bei den weniger beliebten Teilen kann jedoch zumindest die Einsicht in deren Notwendigkeit für das Studium und für den Beruf gefördert werden. Die o.g. lehrveranstaltungsübergreifenden Befragungen zur Studiensituation zeigen, daß in dieser Hinsicht ein erhebliches Defizit existiert. Natürlich muß auch überzeugend nachweisbar sein, daß die jeweiligen Lerngegenstände tatsächlich notwendig und sinnvoll sind. Man versuche schließlich, den "good will" der Adressaten der "Dienstleistung Lehre" gegenüber dem Angebot zu fördern. Marketing-Strategen wissen, daß für die Akzeptanz eines Angebots dessen Qualität allein noch nicht hinreichend ist. Innerhalb eines solchen transparenten und in sich schlüssigen Studienprogramm müßte dann auch Platz sein für unterschiedliche Formen und Qualitäten von Lehre und Didaktik: für den "spröde" Vortragenden (der aber neue Erkenntnisse und Forschungsergebnisse zu präsentieren hat) ebenso wie für den mitreißenden Redner, der in seiner Art auch eher "trockenen" Standardstoff interessant werden läßt.

## LITERATUR

- ENDRUWEIT, G. (1992). Programmevaluation als Laienspiel. In *Soziologie*, H. 2, 107-115.
- KRAPP, A. (1992). Interesse, Lernen und Leistung - Neue Forschungsansätze in der Pädagogischen Psychologie. In *Zeitschrift für Pädagogik*, 38, 5, 747-770.
- KROMREY, H. (1988). Akzeptanz- und Begleitforschung. Methodische Ansätze, Möglichkeiten und Grenzen. In *Massacommunicatie*, 16, 3, 221-242.
- KROMREY, H. (1993a). Studentische Vorlesungskritik. Empirische Daten und Konsequenzen für die Lehre. In: *Soziologie*, H. 1, 39-56.
- KROMREY, H. (1993b). Studentische "Lehreevaluation" oder (nur) "Teilnehmerbefragung" in Lehrveranstaltungen? In H. O. Gralki, D. Grünh & H. Hecht (Hg.): *Evaluation schafft Autonomie*.

Perspektiven der Lehrbewertung an Hochschulen, Berlin: Dokumentationsreihe der FU Berlin, Heft 20, 43-51.

KROMREY, H. (1994). Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In P. Mohler (Hg.): Hochschullehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung, Münster, New York: Waxmann.

SCHMIDT, J. (1980). Evaluation I, Dozentenkurs, Essen: HDZ

WOTTAWA, H. & THIERAU, H. (1990). Lehrbuch Evaluation, Bern, Stuttgart, Toronto: Huber.

♣ Vgl. dazu etwa ENDRUWEIT 1992, KROMREY 1988, WOTTAWA/THIERAU 1990.

♣ So argumentiert etwa explizit SCHMIDT, 1980, 51 f.; implizit sind solche Gedankengänge auch in aktuelleren Begründungen enthalten.

♣ Die Ausführungen stützen sich auf Vorlesungsbefragungen im Rahmen des Projekts "Evaluation der Lehre an der Ruhr-Universität Bochum", durchgeführt i. A. des Rektorats in Zusammenarbeit mit der Universitätskommission für Lehre, Studium und Studienreform; Erhebungen vom WS 1991/92 bis WS 1992/93 in den Fakultäten Wirtschaftswissenschaft, Sozialwissenschaft, Psychologie, Bauingenieurwesen, Physik/Astronomie und Medizin; 10.374 auswertbare Fragebögen.

♣ Als Ansatz in dieser Richtung kann das Projekt "Evaluation der Lehre" an der Universität Mannheim gelten; vgl. den Beitrag von H.-D. DANIEL in diesem Heft.

♣ Index der Motivation zur Teilnahme an der Vorlesung mit den Ausprägungen "extrinsisch" (= Teilnahme ausschließlich, weil es sich um eine Prüfungs- bzw. Klausurpflichtveranstaltung handelt), "intrinsisch ohne Pflicht" (= Teilnahme ohne Prüfungs- oder Klausurpflicht, z.B. aus persönlichem Interesse, Wahlveranstaltung, zur Auffrischung von Kenntnissen etc.) sowie "intrinsisch + Pflicht" (= Teilnahme wg. Prüfungs- bzw. Klausurpflicht und aus anderen Gründen).

♣ Fragewortlaut: "Hat die Vorlesung Ihr Interesse am Fach insgesamt gefördert?"; Einschätzung anhand einer 5er-Skala mit Werten von -2 (nein, überhaupt nicht) bis +2 (ja, sehr).

♣ Index zur Beschreibung der vorherrschenden Motivation der Hörschaft, gebildet als der Anteil der extrinsisch bzw. intrinsisch motivierten Teilnehmer in der jeweiligen Vorlesung: -2: mehr als 90% Nur-Pflicht-Hörer, -1: mehr als 80% Nur-Pflicht-Hörer, +2: mehr als 75% "rein intrinsisch" Motivierte (ohne Prüfungs-/Klausurpflicht), +1: mehr als 70% intrinsisch Motivierte (mit oder ohne zusätzliche Prüfungspflicht), 0: übrige Veranstaltungen).

♣ Auf weitere Tabellen wird aus Platzgründen verzichtet.

♣ Bisher gewonnene Erfahrungen mit fakultätsbezogenen Studiensituationsbefragungen an der Ruhr-Universität Bochum bestätigen deren hohen Informationswert. Erste Auswertungsergebnisse können beim Verfasser angefordert werden.