



**bmb+f**

Bundesministerium für  
Bildung, Wissenschaft,  
Forschung und Technologie

---

# Lehrevaluation und studentische Veranstaltungskritik

Projekte, Instrumente  
und Grundlagen





Natalija el Hage

# **Lehrevaluation und studentische Veranstaltungskritik**

Projekte, Instrumente und Grundlagen

Bonn 1996

Herausgegeben vom

Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie

Die Untersuchung entstand im Rahmen des BMBF-Forschungsprojekts „Studien-situation und studentische Orientierungen“ der Arbeitsgruppe Hochschul-forschung an der Universität Konstanz.

Die Autorin trägt die Verantwortung für den Inhalt.

Herausgeber:

Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie

Referat Öffentlichkeitsarbeit

53170 Bonn

Telefax: (02 28) 57-39 17

E-Mail: [information@bmbf.bund400.de](mailto:information@bmbf.bund400.de)

Druck:

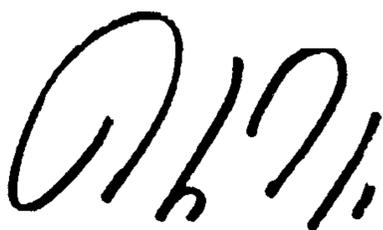
Richard Thierbach, Buch- und Offsetdruckerei, Mülheim/Ruhr (August 1996)

Gedruckt auf Recyclingpapier

## Vorwort

Die Frage nach der Qualität der Lehre an den Hochschulen ist ein wesentliches Element der bildungspolitischen Reformdebatte. Die Evaluation der Lehre ist als Instrument der Qualitätssicherung mittlerweile fester Bestandteil der Studienreform an deutschen Hochschulen. Die wissenschaftliche Kontroverse um die verschiedenen Evaluationsmodelle, die Angemessenheit der Untersuchungsinstrumente sowie die Validität und Reichweite der Evaluationsergebnisse ist noch keineswegs verstummt. Dennoch sind Lehrevaluation und studentische Veranstaltungskritik wichtige Elemente einer Hochschulreform, die wir im Interesse der Studierenden und der Lehrenden jetzt zügig angehen werden. Im Rahmen des neuen Hochschulsonderprogramms III, über das ich mit den Ländern kürzlich eine Verständigung erzielt habe, werden auch Mittel für die interne und externe Evaluation bereitgestellt werden.

Es ist das Verdienst der vorliegenden Studie, neben der Darstellung und Untersuchung der rechtlichen und methodischen Grundlagen der Lehrevaluation eine wissenschaftlich fundierte Auseinandersetzung auf der Grundlage empirischer Untersuchungen zur Frage der Gültigkeit und Genauigkeit studentischer Befragungen zu führen. Dabei werden exemplarisch bereits erprobte Evaluationsmodelle und etablierte Instrumente betrachtet. Die Autorin setzt sich ferner mit elf typischen – in der öffentlichen Diskussion immer wieder anzutreffenden – Vorbehalten gegenüber der studentischen Veranstaltungskritik auseinander. Sie leitet schließlich Kriterien und Empfehlungen für eine anforderungsspezifische Datenerhebung ab und plädiert im Interesse der Wirksamkeit der Lehrevaluation für die Etablierung eines verantwortlichen Qualitätsmanagements auf der Ebene der Fachbereiche.



Dr. Jürgen Rüttgers

Bundesminister für Bildung, Wissenschaft, Forschung und Technologie



# Inhalt

	Seite
<b>1 Aufgaben und Probleme der Lehrevaluation</b> .....	1
1.1 Ebenen und Bereiche der Evaluation an Hochschulen .....	2
1.2 Lehrevaluation im Meinungsstreit .....	4
1.2.1 Lehrfreiheit und Datenschutz.....	4
1.2.2 Kompetenz und Urteilsfähigkeit der Studierenden .....	7
1.2.3 Art und Verwendung von Lehrevaluationen.....	8
1.3 Aufgaben des Berichtes: Dokumentation, Klärungen und Folgerungen.....	9
<b>2 Ansätze und Projekte der Lehrevaluation in Deutschland</b> .....	12
2.1 Projekte des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF, früherer BMBW).....	13
2.2 Projekte der Bundesländer .....	15
2.3 Hochschulrektorenkonferenz, Kultusministerkonferenz, Wissenschaftsrat.....	19
2.4 Kooperationen und Verbünde .....	22
2.5 Lehrevaluation mit externen Institutionen .....	23
2.6 Studentische Evaluationsprojekte .....	26
<b>3 Grundlagen der Lehrevaluation in den Bundesländern</b> .....	27
3.1 Verpflichtungen zur Lehrberichterstattung .....	27
3.2 Überblick zu den gesetzlichen Bestimmungen .....	29
<b>4 Zur Validität studentischer Veranstaltungskritik</b> .....	36
4.1 Beurteilung der Lehre: Kollegen und Studierende im Vergleich.....	37
4.2 Vergleich zwischen Studierenden und anderen Beobachtern/Beobachterinnen .....	40
4.3 Selbstbeurteilungen von Lehrenden und studentische Veranstaltungskritik .....	41
4.4 Qualität der Lehre aus der Sicht junger Berufstätiger und Ehemaliger .....	42

4.5	Lernzuwachs als Maß der Lehrqualität und -effektivität .....	43
4.6	Verbesserung der Lehrbewertung nach Veränderung kritischer Aspekte.....	46
4.7	Weiterempfehlung bzw. Besuch eines weiteren Kurses .....	46
4.8	Reliabilität studentischer Veranstaltungskritik .....	47
<b>5</b>	<b>Einflüsse auf studentische Urteile zur Lehre .....</b>	<b>49</b>
5.1	Erstes Argument: Der Unterhaltungswert dominiert den Inhaltswert .....	52
5.2	Zweites Argument: Beliebtheitsurteil statt Leistungsurteil.....	56
5.3	Drittes Argument: Unreife und Unerfahrenheit der Studierenden.....	57
5.4	Viertes Argument: Nicht die Lehrenden, sondern der Kurs bestimmt die Lehrqualität .....	59
5.5	Fünftes Argument: Art und Form des Kurses beeinflussen die Bewertung .....	61
5.6	Sechstes Argument: Der Status der Lehrenden verzerrt die Lehrbewertung .....	67
5.7	Siebttes Argument: Eigenschaften von Lehrenden verzerren die Bewertung .....	68
5.8	Achstes Argument: Eigenschaften von Studierenden beeinflussen die Bewertung .....	69
5.9	Neuntes Argument: Übereinstimmung von Lehrenden und Studierenden.....	73
5.10	Zehntes Argument: Das Geschlecht beeinflusst die Bewertung .....	74
5.11	Elftes Argument: Noten (erwartete oder erhaltene) beeinflussen die Bewertung .....	77
5.12	Zusammenfassung und Folgerungen.....	82
<b>6</b>	<b>Prinzipien der Entwicklung von Instrumenten zur Lehrevaluation .....</b>	<b>85</b>
6.1	Zwecke und unterschiedliche Anforderungen der Lehrevaluation .....	85
6.2	Zur Notwendigkeit multidimensionaler Lehrbewertungen.....	87
6.3	Auswahl und Abgrenzung verschiedener Dimensionen .....	90
6.4	Vergleichsmaßstab der Beurteilenden.....	97

<b>7</b>	<b>Kriterien und Empfehlungen zur Gestaltung von Fragebogen</b> .....	101
7.1	Formulierung der Fragen und Items.....	101
7.2	Zur Gestaltung des Gesamtbogens.....	105
7.3	Vergleichstabellen nach Veranstaltungsart und Fächern.....	109
7.4	Ergebnisdarstellung und ihre Vorbedingungen.....	111
7.5	Zeitpunkt und Frequenz der Beurteilung.....	114
7.6	Zusammenfassende Übersicht: Empfehlungen.....	116
<b>8</b>	<b>Instrumente zur Lehrevaluation in Deutschland: Beispiele</b> .....	120
8.1	Fragebogen zur Beurteilung von Vorlesungen (VBVOR) von J. M. Diehl.....	122
8.2	Fragebogen zur Beurteilung von Seminaren mit Referaten (VBREF) von J. M. Diehl.....	124
8.3	Evaluation der Lehre an der Ruhr-Universität Bochum: Fragebogen für Vorlesungen (I. Studierende, II. Dozent(inn)en) von H. Kromrey.....	126
8.4	Fragebogen zur studentischen Rückmeldung in Lehrver- anstaltungen von H. Kromrey.....	128
8.5	Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE) von H. Rindermann und M. Amelang.....	131
8.6	Bielefelder Fragebogen zur Bewertung von Lehrveranstaltungen von W. D. Webler, IZHD/Arbeitsgruppe Studentische Veran- staltungskritik.....	136
8.7	Fragebogen zur „Evaluation der Lehre“ an der Universität Mannheim.....	137
<b>9</b>	<b>Möglichkeiten der Datenverwendung: Reichweite und Wirksamkeit</b> .....	140
9.1	Die Wirkungslosigkeit reiner Datenerhebung.....	140
9.2	Verwendungsmöglichkeiten von Beurteilungsdaten.....	143
9.3	Gesamtkonzept und Strategien der Verwendung.....	154



## **1 Aufgaben und Probleme der Lehrevaluation**

Seit einigen Jahren wird das Thema Evaluation der Lehre in Deutschland heftig diskutiert. Was am Anfang als populistische Schlagzeile erschien, als Modethema ohne Folgewirkung, hat sich sowohl in der wissenschaftlichen Auseinandersetzung als auch im politischen Diskurs etabliert, sogar in die Gesetzgebung der Länder Eingang gefunden. In Deutschland gibt es zur Zeit keine nennenswerte hochschulpolitische Institution, die sich nicht auf die eine oder andere Art mit der Lehrevaluation als Beitrag zur Stärkung der Lehre auseinandersetzen würde und dazu Überlegungen und Empfehlungen verabschiedet (vgl. zuletzt Wissenschaftsrat 1996).

An vielen Hochschulen und Fachbereichen werden mittlerweile Evaluationen der Lehre vorgenommen. Ein zentrales Element der Lehrevaluation bildet die studentische Veranstaltungskritik. Trotz mancher Modellvorhaben und unzähliger Projekte dazu und trotz zunehmender gesetzlicher Regelungen über Lehrberichte der Hochschulen in den Bundesländern sind studentische Veranstaltungskritiken nach wie vor umstritten: hinsichtlich ihrer grundsätzlichen Gültigkeit, hinsichtlich zuverlässiger Instrumente und schließlich auch hinsichtlich ihrer Verwendung.

Im Mittelpunkt dieses Berichtes steht die studentische Veranstaltungskritik als gewichtiger Teil von Lehrevaluation und Lehrberichterstattung. Zum einen hat der Bericht eine dokumentarische Funktion, indem er

- die Ansätze und Projekte der Lehrevaluation in Deutschland von Seiten des Bundes, der Länder und anderer Instanzen zusammenstellt;
- einen Überblick zur Gesetzgebung der Länder zum Thema Lehrberichte, Lehrevaluation und studentische Veranstaltungskritik liefert;
- Befragungsinstrumente zur Beurteilung von Lehrveranstaltungen und von Lehrenden durch Studierende vorstellt.

Zum anderen soll dieser Bericht zur Klärung strittiger Fragen beitragen, die sich auf die Gültigkeit studentischer Stellungnahmen und mögliche, die studentischen Urteile beeinflussende Faktoren beziehen. Im Mittelpunkt steht die Frage, ob Studierende die Lehrleistung ihrer Lehrenden zutreffend beurteilen können.

Außerdem werden Inhalt und Gestaltung entsprechender Erfassungsinstrumente behandelt; vor allem die zentralen Dimensionen der Lehrqualität werden herausgearbeitet. Schließlich werden die Möglichkeiten der Verwendung der gewonnenen Daten und Informationen zur wirksamen Stärkung und Verbesserung der Lehre diskutiert. Denn nur, wenn die Lehrevaluation und Lehrberichterstattung in ein umfassenderes Konzept zur Entwicklung der Lehrqualität eingebettet wird, ist dadurch Wirksamkeit zu erreichen.

Mit seinen verschiedenen Teilen stellt der Bericht eine breite Einführung in die Evaluation der Lehre dar: Er bietet nicht nur eine Übersicht über Ansätze und Erfahrungen, über Untersuchungen und Studien in diesem für die Hochschulentwicklung bedeutsamen Feld, sondern er zieht auch Folgerungen und spricht Empfehlungen aus für die Gestaltung der Instrumente, ihren praktischen Einsatz sowie die wirksame Nutzung der Lehrevaluation.

## **1.1 Ebenen und Bereiche der Evaluation an Hochschulen**

Um das Thema "Evaluation der Lehre" systematisch behandeln zu können, werden drei Ebenen unterschieden, die aufgrund der Organisation der Hochschulen in Deutschland zu beachten sind: (I) die Hochschule in ihrer Gesamtheit; (II) die einzelnen Fachbereiche mit ihrem Kollegium; (III) die einzelnen Lehrveranstaltungen und ihr Dozent bzw. ihre Dozentin.

Bei der Diskussion für und wider die Lehrevaluation und bei Überlegungen ihrer praktischen Verwendung wird häufig zwischen diesen Organisationsebenen der Hochschule nicht unterschieden, was zu Unklarheiten über angemessene Beurteilungsinstrumente und zu Mißverständnissen über die Vergleichsmöglichkeiten führt. Deshalb sei dieser Unterschied kurz erläutert.

### **Ebene I: Hochschule in ihrer Gesamtheit**

Wird die Hochschule in ihrer Gesamtheit als Institution evaluiert, so sind die Leistungen in Forschung, Administration und Lehre getrennt zu erfassen. An dieser Stelle soll nicht weiter auf die Möglichkeiten der Messung und Bewertung von Forschungsleistungen eingegangen werden - dazu gibt es sowohl eingespielte Verfahren als auch eine umfängliche Literatur.

Die Administration einer Hochschule läßt sich am ehesten mit Mitteln und Methoden bewerten, die auch bei der Evaluation anderer Verwaltungen oder Unternehmungen (bzw. ihrer Abteilungen) angewandt werden. In Bezug auf Hochschulen würde es sich dabei um Fragen des effizienten Managements der Studierendenverwaltung, des Beschaffungswesens und der Mittelverwendung oder des Personaleinsatzes handeln.

Der Bewertung der Lehre auf Hochschulebene stehen eine ganze Reihe von Schwierigkeiten entgegen. Es erscheint angebracht, auf dieser Ebene die Qualität der Lehre als einen Teil der Studienbedingungen einer Hochschule insgesamt zu betrachten, wie Öffnungszeiten von Lehrvorrichtungen, Ausstattung mit studentischen Arbeitsplätzen, fachübergreifende Lehr- und Qualifikationsangebote (z.B. Studium generale, Sprachkurse, EDV-Anwendung).

Fraglich bleibt jedoch, inwieweit Bewertungen zur Lehrleistung, wie übrigens auch bei der Forschungsleistung, über eine gesamte Hochschule aussagekräftig und von praktischem Nutzen sein können. Die Untersuchung von Giesen und Jansen (1983) kommt zu dem Ergebnis, daß "... der Varianzanteil, der auf das spezifische Erleben der Fachgruppen zurückgeführt werden kann, stets größer ist als der universitätsspezifische Anteil" (S. 226).

Diese Ergebnisse geben einen wichtigen Hinweis für Evaluationsmaßnahmen zur Lehre: Auf der Hochschulebene sind nur ganz allgemeine - vor allem fachbereichsunspezifische - Studienbedingungen, d.h. Lehr- und Lernbedingungen, zu erfassen. Es erscheint wenig sinnvoll und zudem methodisch außerordentlich bedenklich, Durchschnittswerte über Fachbereiche zu ziehen und einen „Gesamtwert“ für eine Hochschule zu bilden, um diese Werte dann, etwa in Form eines Rankings, zu vergleichen.

## **Ebene II: Fachbereiche als Organisationseinheit**

Bei der Evaluierung eines Fachbereichs zeigt sich die administrativ-organisatorische Seite mit der Lehre eng verwoben. Der Lehrkörper eines Fachbereichs ist sowohl für allgemeine Faktoren der Studienbedingungen, wie z.B. Öffnungszeiten von Institutsbibliothek und Institutssekretariat, als auch für organisatorische Aspekte der Lehre, z.B. das Angebot genügender Pflichtveranstaltungen oder die Vermeidung zeitlicher Überschneidungen von Kursen, verantwortlich. Schwieriger ist zu sagen, inwieweit ein Fachbereich auch für die Lehrqualität der einzelnen Veranstaltungen zuständig ist, da gegenwärtig vor allem die jeweiligen Lehrstuhlinhaber und -inhaberinnen für ihre Kurse verantwortlich sind.

Möglich und sinnvoll sind für den Fachbereich als Organisationseinheit somit Daten, die sich auf die Gesamtsituation und Auswirkungen der Lehre beziehen, wie z.B. Studienkonzepte und Schwerpunkte, Termineinhaltung, Prüfungsergebnisse und Semesterzahlen, Absolventen-/ Absolventinnenzahlen oder der Berufseinstieg und Berufsverlauf der Absolventen und Absolventinnen (vgl. Webler, 1993).

## **Ebene III: Lehrveranstaltungen und Lehrende**

Für die Lehrveranstaltungen zeichnet vor allem der jeweilige Dozent/ die jeweilige Dozentin verantwortlich. Studentische Veranstaltungskritik bezieht sich gewöhnlich fast ausschließlich auf die Ebene einzelner Lehrveranstaltungen und Lehrender. Es wäre jedoch angebracht, auch fachspezifische Studienbedingungen und den organisatorischen Aspekt der Lehre mit einem ähnlichen Instrument zu

erfassen (d.h. zur Evaluation der Ebene des Fachbereichs die Erfahrungen und Urteile der Studierenden heranzuziehen und zu erfragen).

Lehrveranstaltungen und Lehrende durch Studierende beurteilen zu lassen, kann in ganz unterschiedlicher Weise erfolgen: z.B. als unmittelbare Rückmeldung auf eine einzelne Veranstaltung, als Beurteilung einer ganzen Veranstaltungsreihe in einem Semester oder als Bewertung der didaktischen Qualitäten der Lehrenden insgesamt.

Inwieweit dabei schriftliche Fragebogen mit vorgegebenen Antwortmöglichkeiten einsetzbar sind, ist die entscheidende Frage. Das hängt davon ab, ob die Studierenden in der Lage sind, zutreffend ihre Erfahrungen mitzuteilen und relativ unbeeinflusst ihre Urteile abzugeben. Gerade diese Fähigkeit wird vielfach angezweifelt. Deshalb wird bestritten, daß eine Befragung in schriftlicher Form gültige und verlässliche Resultate über die Lehrleistung der Lehrenden liefern könne.

Da die drei Ebenen -Hochschule, Fachbereich, Lehrveranstaltung- weitgehend unabhängig sind (es ist z.B. ein ausgezeichnete Kurs an einer Hochschule mit schlechten Studienbedingungen und chaotischer Organisation des Fachbereichs denkbar), ist die sorgfältige Unterscheidung der verschiedenen Ebenen unabdingbar. Auch die Westdeutsche Rektorenkonferenz (WRK) sprach sich bereits 1986 für "vergleichende Leistungsbeurteilungen zwischen gleichen Fächern", also dem "intradisziplinären Vergleich" aus. Sie warnte zurecht davor, die Hochschule in ihrer Ganzheit gleichzusetzen mit der Summe der Leistungen einzelner Fächer (vgl. WRK, 1986). Dies gilt für die Forschung wie für die Lehre.

## **1.2 Lehrevaluation im Meinungsstreit**

In der politischen und wissenschaftlichen Debatte um die Lehrevaluation der letzten Jahre sind verschiedene Kontroversen zu erkennen, auf die sich Befürwortung und Widerstand gegen mögliche Evaluationsmaßnahmen, wie zum Beispiel die Erstattung von Lehrberichten oder die Befragung der Studierenden, immer wieder beziehen. Die grundsätzlichen Konfliktlinien in diesem Bereich betreffen Fragen der gesetzlichen Grundlage, der Art und Verwendung von Lehrevaluationen sowie der studentischen Kompetenz und Urteilsfähigkeit.

### **1.2.1 Lehrfreiheit und Datenschutz**

Eine entscheidende Debatte bezieht sich auf den rechtlichen Rahmen studentischer Veranstaltungskritiken und die mögliche Veröffentlichung ihrer Ergebnisse. Häufig scheinen aber die Auseinandersetzungen um rechtliche Vorschriften nur die Verlagerung eines anderen Streitpunkts darzustellen: " ... anders als z.B.

in den USA, wo Transparenz und Rechenschaftablegung auch in der Hochschullehre gesellschaftlich eingeklagt werden, kann sich bei uns Hochschullehre gegen gesellschaftliche Einsicht- und Einflußnahme weitgehend abschotten" (Heger, 1991, S. 52).

Noch deutlicher wird in einer Zwischenbilanz zur Lehrqualität des Ministeriums für Bildung und Wissenschaft ausgeführt: "Umgekehrt drängt sich der Verdacht auf, daß die Freiheit der Wissenschaft, die die Freiheit der Lehre umschließt, als Schutzargument mißbraucht wird, wenn sie zur systematischen und rechtlich abgesicherten Abschottung gegen jede Form der Kritik aufgefaßt wird. Dies ist um so erstaunlicher, als das andere große Standbein der Wissenschaft, die Forschung, öffentlich zugänglich verhandelt und auch kritisiert wird" (Bundesministerium für Bildung und Wissenschaft, 1993, S. 3).

Bei der Auseinandersetzung rechtlicher Art werden vor allem zwei Dimensionen thematisiert: zum einen Art. 5 Abs. 3 GG, der die Lehrfreiheit garantiert, und zum anderen Bestimmungen des Datenschutzes. Es wird diskutiert, ob es möglich ist "Dritten", z.B. Fachschaften oder Hochschulgruppen, zu verbieten, studentische Veranstaltungskritik durchzuführen. Zwar ist dies nicht möglich, da die "Freiheit der Meinungsäußerung" jeden zur Veranstaltungskritik berechtigt. Aber es ist zur Zeit umstritten, ob die Regelungen zur "Dienstaufsicht" von Hochschulleitung oder Wissenschaftsministerium sich derart ändern ließen, daß von dieser Seite Lehrbewertung organisiert und in Entscheidungen miteinbezogen werden könnten, und ob dies mit dem Art. 5 Abs. 3 GG vereinbar sei.

Grottian (vgl. 1992, S. 30 ff.) ist der Ansicht, daß die Lehre eine "wenigstens universitätsöffentliche Angelegenheit" sei. Die Bewertung der Lehre vergleicht er mit der der Forschung: "In der wissenschaftlichen Auseinandersetzung ertragen wir Kritik, bis ins Persönliche gehende Verrisse, mit einiger Gelassenheit; dagegen wird in der Lehre öffentliche Kritik zu einer Verletzung der persönlichen Intimsphäre hochstilisiert" (Grottian, 1992, S. 31). Er geht davon aus, daß studentische Veranstaltungskritik zur Dienstpflicht erklärt werden kann und beides mit dem Art. 5 Abs. 3 GG vereinbar ist. Gralki nennt Grottians Ansatz provokativ und kritisiert, daß im Vergleich zur wissenschaftlichen Kritik studentische Lehrbewertung anonym sei (vgl. Gralki, 1992).

Eine häufiger diskutierte Frage ist, ob von Studierenden oder der Verwaltung erhobene Bewertungsdaten zur Lehre veröffentlicht werden dürfen. Grottian vertritt die Auffassung: "Studierende haben als erste Adressaten der Lehre ein demokratisches Recht, die Befragungsergebnisse verantwortlich diskutieren zu können" (Grottian, 1992, S. 31). Erörtert wird, ob dies mit dem aus dem Artikel 2 Abs. 1 GG abgeleiteten "informationellem Selbstbestimmungsrecht" vereinbar sei. Grottian dazu: "Zwar liegt in der Datenerhebung und namentlichen Veröffentlichung ein Eingriff in das Recht auf informationelle Selbstbestimmung, doch un-

terliegt das Recht auf informationelle Selbstbestimmung seinerseits Einschränkungen zugunsten der Allgemeinheit" (Grottian, 1992, S. 32). Informationelle Selbstbestimmung bleibt sicherlich im "unantastbaren Intimbereich" zu wahren, die Lehrtätigkeit eines Hochschullehrers ist jedoch allgemein zugänglich und prinzipiell eine "öffentliche Veranstaltung".

Detmer (1991) erläutert aus einem Urteil des Oberverwaltungsgerichts Berlin (OVG PV Bln 9.91 vom 18. Juli 1991), daß Datenerhebungen und -veröffentlichungen mitbestimmungspflichtig im Sinne des Personalrechts und somit nur mit Einwilligung der Betroffenen möglich sind.

Der frühere hessische Datenschutzbeauftragte Spiros Simitis vertritt eine andere Sicht: Aufbau und Stil von Veranstaltungen seien keine "schützenswerten persönlichen Daten". Wird studentische Veranstaltungskritik allerdings von der Hochschulleitung und u.U. gegen den Willen der Lehrenden organisiert, würde dies in den Grundsatz der Lehrfreiheit eingreifen (vgl. SPIEGEL-SPEZIAL, 1993, S. 124 ff.). Somit würde weniger der Datenschutz als die "garantierte Lehrfreiheit" einer verpflichtenden und öffentlichen Veranstaltungskritik im Wege stehen.

Endgültig kann diese Rechtsunsicherheit vermutlich nur vom Bundesverfassungsgericht entschieden werden. Wie dies entschieden wird, ist schwer einzuschätzen, da verschiedene Ansichten auch zu der Rolle der Studierenden an den Hochschulen mit hineinspielen. So zitiert z.B. Krieger aus einem Verwaltungsgerichtsbeschuß (VG FK (Bln) - C - 6.91) des Verwaltungsgerichts Berlin vom 19. April 1991: "Auch wenn man dieser Rechtsprechung folgt, so setzt §85 Abs. 1 Nr. 13 Pers VG jedenfalls voraus, daß die technisierte Verhaltens- oder Leistungsüberwachung auf Daten beruht, die hierzu überhaupt geeignet sind. Von vornherein ungeeignet dürften neben Statusdaten (Name, Anschrift, Geburtstag usw.) Bewertungen durch Personen sein, die zu einer objektiven Beurteilung generell nicht in der Lage sind. Die Einschätzung von Lehrveranstaltungen durch Studenten, wie sie bei der hier streitigen Fragebogenaktion vorgesehen ist, dürfte zu dieser Kategorie ungeeigneter Daten gehören" (Krieger 1992, S. 59).

Dieser Argumentation schloß sich übrigens die zweite Instanz, das Berliner Oberverwaltungsgericht, nicht an. Es befand, daß es nur darauf ankäme, daß dieser subjektiven Leistungsbewertung Bedeutung für die Lehre beigemessen würde (vgl. Detmer, 1991).

Wie die rechtlichen Debatten ausgehen und von Gerichten entschieden werden, bleibt gegenwärtig weitgehend offen - zu kontrovers fallen die Stellungnahmen der Rechtsexperten aus, zu unterschiedlich urteilen verschiedene gerichtliche Instanzen. Jedoch ist es für den juristischen Streit keineswegs unwichtig, ob die "subjektive Leistungsbewertung" verschiedener Studierender eine "objektive Be-

urteilung" von Lehre und Lehrenden darstellen kann (vgl. auch das Rechtsgutachten für den Deutschen Hochschulverband von Hufen, 1995).

### **1.2.2 Kompetenz und Urteilsfähigkeit der Studierenden**

Der Meinungsstreit, ob studentische Veranstaltungskritik als subjektive Meinungsbefragung (auch im Sinne einer Verbraucherbefragung), wie dies z.B. Anneliese Monika Grüger (1982) darstellt, oder als objektives Beurteilungsinstrument angesehen werden kann, bezieht sich zwar vordergründig auf methodische Aspekte der Art, des Einsatzes und der Auswertung solcher Instrumente. Dahinter verbergen sich aber prinzipielle Bedenken an der Berechtigung solcher Verfahren und an der Fähigkeit der Studierenden zu kompetenten Urteilen.

"Die Argumentationslinie beginnt beim grundsätzlichen Zweifel an der Kompetenz der Studierenden, Lehrleistung überhaupt bewerten zu können" (Preißer, 1991, S. 74). Der Frankfurter Psychologe Süllwold (1992) vertritt die Auffassung, daß Unterschiede in der Beurteilung verschiedener Hochschullehrer und -lehrerinnen einzig und allein auf unterschiedliches Bewertungsverhalten der Studierenden zurückzuführen ist, und nichts über deren tatsächliche Lehrqualität aussagen; er folgert daher: "Erhebungen über die Beurteilung von Hochschullehrern durch Studierende sind in fundamentaler Weise fehlerhaft, wenn die einzelnen Hochschullehrer durch verschiedene Studentengruppen beurteilt werden" (Süllwold, 1992, S. 34). Damit schließt Süllwold letztlich aus, daß objektive Instrumente, wie sie in vielen psychologischen Bereichen eingesetzt werden, für die studentische Veranstaltungskritik möglich sind.

Neben solchen grundsätzlichen Vorbehalten wird häufig auch ins Feld geführt, daß Gegebenheiten der Studierenden oder des Kurses deren Urteile beeinflussen oder verzerren. Insbesondere wird befürchtet, daß schwierigere und anspruchsvollere Kurse eine schlechtere Beurteilung erfahren. Oder daß eher die Freundlichkeit und Beliebtheit eines Hochschullehrers von den Studierenden bewertet wird und weniger seine tatsächliche Lehrleistung.

Solche Vorbehalte sind ernst zu nehmen, zumal wenig transparente und scheinbar verzerrte Ergebnisse von Lehrbewertungen ihren Grund auch in der Qualität der verwendeten Instrumente und weniger in der unzureichenden Urteilsfähigkeit der Studierenden haben können.

Beim Einsatz studentischer Veranstaltungskritiken mittels Fragebogen kommt es entscheidend auf die Art der Verwendung an, welche Ansprüche an deren Standards hinsichtlich Gültigkeit und Zuverlässigkeit zu stellen sind.

Auf der anderen Seite haben amerikanische Studien, in letzter Zeit auch vermehrt deutsche Studien, sich bemüht, aufzuklären, ob studentische Urteile valide sind, was sie beeinflußt und von welcher Güte die Erfassungsinstrumente sein sollten. Allerdings besteht der Eindruck, daß diese Untersuchungen nur wenig rezipiert werden und herkömmliche „Vorurteile“ gegen die Kompetenz und Urteilsfähigkeit der Studierenden zur Lehrbewertung weiter gepflegt werden.

### **1.2.3 Art und Verwendung von Lehrevaluationen**

Die Art und Verwendung der Lehrevaluationen ist ebenfalls umstritten. Vier Konfliktfelder, die in Diskussionen immer wieder heftig thematisiert werden, lassen sich unterscheiden: die Einschätzung des Nutzens, die mögliche staatliche Einflußnahme, die Veröffentlichung von Ergebnissen und die praktischen Konsequenzen von Lehrbewertungen.

Das erste Konfliktfeld bezieht sich auf die grundsätzlichen Fragen nach dem Nutzen solcher Verfahren. Sehen manche Befürworter und Befürworterinnen in Maßnahmen der Lehrbewertung das Allheilmittel für die meisten Schwierigkeiten an den Hochschulen, bezweifeln Gegner und Gegnerinnen deren prinzipielle Brauchbarkeit oder halten den Aufwand im Vergleich zum Nutzen für zu hoch. Als ein Teilbereich dieser Auseinandersetzung kann die Frage angesehen werden, ob es an den deutschen Hochschulen nicht gravierendere Probleme gäbe als die Auseinandersetzung um die Lehrqualität. Damit wird auch die öffentliche Aufmerksamkeit kritisiert, die ihren Fokus in den letzten Jahren ganz besonders auf diesen Aspekt gelenkt hatte.

Der nächste Punkt der Auseinandersetzung in diesem Bereich, von übergeordneten Gremien wie den Landesrektorenkonferenzen oder der HRK häufiger aufgegriffen, ist die Problematik einer möglichen staatlichen Einflußnahme. Es wird diskutiert, inwieweit über die Lehrevaluation die Autonomie der Hochschule bzw. die Freiheit der Lehrenden beeinträchtigt werden könnte. Dabei reichen die Ansichten von der Befürwortung einer Lehrevaluation, die zwar von den Hochschulen durchgeführt, von staatlicher Seite jedoch begleitet wird (wie dies z.B. in den Niederlanden der Fall ist), bis hin zu der immer noch existierenden Ansicht, daß Lehre ein Privatissimum des einzelnen Professors sei und sich nicht einmal die jeweilige Hochschule einmischen dürfe.

Außer diesen beiden grundlegenden Abwägungsbereichen bestehen zudem ganz praktische Überlegungen und Ängste. Zum einen beziehen sie sich auf die mögliche Veröffentlichung von Evaluationsergebnissen. In den Hochschulsensaten ist ein Ringen und Feilschen zu beobachten. Werden Evaluationsdaten veröffentlicht? Sollen sich alle Lehrenden damit einverstanden erklären? Wie detailliert soll/ kann/ darf veröffentlicht werden? Nur kumulierte Ergebnisse? Oder mit Veranstaltungstitel? Gar mit dem eigenen Namen? In dieser Auseinandersetzung

geht es nicht allein darum, professorale Privilegien zu verteidigen. Deutlich wird zugleich die Unsicherheit, in einem Bereich bewertet zu werden, den man nicht eigens gelernt hat und zu dem man keine Beurteilungserfahrungen sammeln konnte.

Zum anderen werden die möglichen Konsequenzen von Lehrbewertungen debattiert, wobei alle bisher genannten rechtlichen und methodischen Argumente einfließen. Die Streitpunkte reichen von der Frage, ob sich solche Bewertungen überhaupt auf Berufungen, Bleibeverhandlungen oder gar das Einkommen auswirken sollten, über die Frage einer darüber möglicherweise verstärkten staatlichen Steuerung (Mittelvergabe, Karriere, Gratifikationen) bis hin zu Fragen nach Art und Ausmaß möglicher persönlicher Konsequenzen.

In Deutschland haben sich, neben den unzähligen Projekten studentischer Veranstaltungskritik, die überwiegend die einzelne Veranstaltung im Blickpunkt haben, vor allem Vorhaben etabliert, die nach dem sogenannten 'Niederländischen Modell' vorgehen (vgl. Richter, 1995). Dabei werden nach internen Datensammlungen Lehrende von Kollegen und Kolleginnen (den sogenannten 'peers') beurteilt und in einem weiteren Schritt konkrete Empfehlungen formuliert. Nach einer gewissen Zeit gibt es ein Bewertungstreffen, bei dem die Umsetzungen der vorgeschlagenen Maßnahmen besprochen werden (siehe Holtkamp & Schnitzer, 1992; Reissert, 1994). Dieses Verfahren, bei dem es immer um die Beurteilung eines Faches oder Fachbereiches geht, wird mittlerweile bei nahezu allen größeren Evaluationsprojekten in Deutschland verwandt (vgl. Kapitel 2).

### **1.3 Aufgaben des Berichtes: Dokumentation, Klärungen und Folgerungen**

Eine große Anzahl an Projekten zur Beurteilung und Verbesserung der Lehre sind in den letzten Jahren auf den Weg gebracht worden. Die unüberschaubare Anzahl von kleineren und größeren Vorhaben der Lehrevaluation, die häufig Daten der studentischen Veranstaltungskritik verwenden, wird von sehr unterschiedlichen Initiatoren und Initiatorinnen betrieben. Von ganzen Hochschulen, die mit externer Hilfe evaluiert werden, bis hin zu studentischen Fachschaften in einzelnen Fächern ist alles vertreten. So unterschiedlich wie die Betreiber und Betreiberinnen der Projekte sind, so unterschiedlich ist die Qualität der verwendeten Fragebogen zur studentischen Veranstaltungskritik.

In Kapitel zwei wird eine Übersicht über Vorhaben und Projekte vorgestellt, die sich mit Lehrevaluation und Lehrberichterstattung in Deutschland befassen. Die entsprechenden Ansprechpartner und -partnerinnen in den einzelnen Institutionen werden im Anhang genannt.

Die meisten Bundesländer haben ihre Hochschulgesetze geändert, um mehr oder weniger detailliert die Berichterstattung über die Situation der Lehre festzulegen. Dabei sind neben statistischen Daten (Studiendauer, Abbruchquote etc.) zum Teil auch Ergebnisse der Befragung von Studierenden vorgesehen. Eine Übersicht zur Gesetzgebung in den Bundesländern ist in Kapitel drei zu finden. Die Unterschiede hinsichtlich der Vorgaben zur Lehrberichterstattung lassen erwarten, daß eine Vergleichbarkeit kaum gegeben ist.

Da viele der in Deutschland durchgeführten Evaluationsprojekte auch Daten aus der umstrittenen studentischen Veranstaltungskritik einbeziehen, ist es nützlich und einer Versachlichung der Diskussion dienlich, sich anhand empirischer Studien darüber zu vergewissern, welche Zweifel gegenüber studentischen Veranstaltungskritiken begründet sind und welche nicht. Eine derartige Aufarbeitung, wie sie in Kapitel 4 vorgelegt wird, kann nicht alle Fragen endgültig beantworten. Sie vermag aber, entsprechend dem Stand der empirischen Forschung, zu konstatieren, inwieweit studentische Urteile über Lehrveranstaltungen und Dozenten als gültige, zuverlässige und praktikable Grundlagen der Lehr-evaluation herangezogen werden können.

Die vorhandenen Untersuchungen stammen fast durchweg aus den USA, in denen studentische Veranstaltungskritik und Lehrevaluation in ihren verschiedenen Formen längst gebräuchlich sind. In letzter Zeit sind verschiedentlich aufschlußreiche deutsche Arbeiten zu diesem Thema erschienen (z.B. Kromrey, 1993 und 1994b; Daniel, 1995b; Rindermann, 1995a; Mohler, 1995); diese werden in diesen Bericht nicht mehr eingearbeitet. Alles in allem bestätigen sie aber die amerikanischen Befunde zur Gültigkeit, Verlässlichkeit und Nützlichkeit studentischer Befragungen zur Lehre - vorausgesetzt, diese werden mit angemessenen Instrumenten durchgeführt, methodisch sinnvoll ausgewertet und in der Praxis der Hochschulen und Fachbereiche kompetent gehandhabt.

Verschiedene Gründe sprechen für sorgfältig ausgearbeitete Fragebogen. Marsh (1982b) und Aleamoni und Thomas (1980) haben nachgewiesen, daß detaillierte und multidimensionale Fragebogen weniger anfällig für externe Einflüsse sind, verglichen mit Fragebogen, die sich auf nur wenige globale Items stützen. Über mögliche Faktoren, die sich auf die Ergebnisse studentischer Lehrbewertung auswirken können, wird in Kapitel fünf berichtet.

Anhand bisheriger Untersuchungsergebnisse und -erfahrungen werden Kriterien zusammengestellt, die für die Erstellung von Fragebogen der studentischen Veranstaltungskritik beachtenswert sind. Da die meisten dieser Untersuchungen in den Vereinigten Staaten durchgeführt wurden, fehlen an einigen Punkten noch für die deutsche Hochschullandschaft notwendige Untersuchungen.

In Kapitel sechs und sieben werden Prinzipien und Konstruktionskriterien für die Entwicklung und Verwendung von Instrumenten studentischer Veranstaltungskritik aufgezeigt. Veranschaulicht werden soll damit, daß eine sorgsame Fragebogenkonstruktion notwendig ist, zumal wenn dieses Beurteilungsverfahren dazu dienen soll, die Qualität verschiedener Veranstaltungen und Lehrender zu vergleichen.

Damit stellt sich zwangsläufig die Frage nach bereits bestehenden Instrumenten zur Messung der Effizienz und Qualität in den verschiedenen Hochschulbereichen. Diese befinden sich zum großen Teil allerdings noch in der Entwicklungs- oder Erprobungsphase.

In Kapitel acht werden einige deutschsprachige Instrumente der studentischen Veranstaltungskritik vorgestellt. Ausgewählt wurden Fragebogen, die einen allgemeinen Anspruch haben oder bei hochschulübergreifenden Evaluationsprojekten verwendet wurden. Damit soll keine Herabwürdigung oft sorgsam erstellter fach- oder hochschulspezifischer Instrumente verbunden sein. Es würde jedoch den vorhandenen Rahmen sprengen, die ganze Bandbreite der verwendeten Instrumente aufzunehmen.

Obwohl in Deutschland bei den meisten Projekten die Revisionszeitpunkte noch nicht erreicht sind, läßt sich doch vermuten, daß eine optimale Nutzung der Evaluationsdaten mehr Steuerungsmechanismen benötigen würde als bis jetzt vorhanden sind. Für eine effektive Nutzung müssen die erhobenen Beurteilungen in ein Gesamtsystem der Verbesserung der Lehre eingebettet werden. Ein kurzer Anriß dieser Diskussion und Überlegungen zu einem Gesamtkonzept sind in Kapitel neun zu finden.

Lehrevaluation durch studentische Veranstaltungskritik kann dabei nicht mehr als die Sammlung notwendiger Daten und Befunde bedeuten. An diese Phase der Datenerhebung und ihrer Analyse müssen sich entsprechende Verwertungsstrategien anschließen (vgl. auch Kromrey, 1995b; Barz, 1995). Im vorliegenden Bericht sollen die verschiedenen Möglichkeiten der Datenverwendung mit ihren Vor- und Nachteilen dargelegt werden. Es wird der Versuch unternommen, diese Möglichkeiten zu einem Gesamtansatz zusammenzuführen. Daran schließen sich Überlegungen zur Einführung eines solchen Gesamtkonzepts an.

Sollen studentische Veranstaltungskritiken als Teil der Evaluationsmaßnahmen an Hochschulen eine Verbesserung der Lehre zur Folge haben, müssen Rahmenbedingungen geschaffen werden, die einer zielgerichteten und akzeptierten Verwendung der erhobenen Daten dienlich sind. Dreh- und Angelpunkt ist dabei die Stärkung der Wichtigkeit der Lehre an den Hochschulen (vgl. Wissenschaftsrat, 1996). Bei der Einführung eines "sensiblen" Instruments wie der studentischen Veranstaltungskritik kommt deshalb der Einführungsstrategie eine besondere Bedeutung zu.

## 2 Ansätze und Projekte der Lehrevaluation in Deutschland

Die Projekte zur Evaluation der Lehre an Universitäten und Fachhochschulen haben die unterschiedlichsten Initiatoren und Initiatorinnen. Dies kann in einem Fall das Bundesbildungsministerium, in einem anderen ein Hochschuldidaktisches Zentrum und in einem dritten eine studentische Fachschaft sein.

Es würde den Rahmen dieses Berichts bei weitem sprengen, würde versucht, alle diese Aktivitäten darzustellen. Eine Einschränkung muß somit vorgenommen werden. Es werden daher nur solche Vorhaben aufgenommen und vorgestellt, die von staatlichen Stellen initiiert und/oder begleitet werden. Weiter werden Vorhaben einbezogen, die über eine einzelne Hochschule hinausgehen oder einen hochschulvergleichenden Ansatz haben. Zusätzlich wird auf jene Projekte eingegangen, die von Institutionen betreut werden, deren Arbeitsschwerpunkt u.a. im Bereich der Lehrevaluation liegt und die hochschulübergreifende Projekte durchgeführt haben.

Schließlich wird aufgrund ihrer besonderen Rolle auch auf Aktivitäten von Studierenden verwiesen. Dabei wurde eine Ausnahme von dem Grundsatz gemacht, nur hochschulübergreifende Projekte darzustellen. Zwar wurde auf die Beschreibung von einzelnen Fällen (z.B. von Fachschaften) verzichtet; koordinierte Arbeiten, die sich auf eine gesamte Hochschule beziehen, wurden jedoch aufgenommen. Für ehrenamtlich tätige Studierendenvertreter und -vertreterinnen ist eine große Hochschule bereits ein weites und arbeitsintensives Feld.

Aus den genannten Prämissen für die berücksichtigten Projekte und Vorhaben ergibt sich die Struktur für dieses Kapitel: Als erstes werden vom Bundesbildungsministerium (BMBW bzw. BMBF) geförderte und begleitete Tagungen und Projekte aufgeführt. Daran schließen sich Projekte der einzelnen Bundesländer an, wie sie auf Anfrage von diesen mitgeteilt wurden.

Es folgen Projekte der Hochschulrektorenkonferenz (HRK), der Kultusministerkonferenz (KMK) und des Wissenschaftsrates. Diese Institutionen bemühen sich, gewisse Standards im Bereich der Lehrevaluation zu diskutieren und durchzusetzen (vgl. HRK, 1995c; Wissenschaftsrat, 1996). Es haben sich zudem bereits einige Kooperationen von Hochschulen gebildet, die im Anschluß genannt werden. Diese dienen meist dazu, die Erfahrungen anderer Hochschulen mit der Lehrevaluation nutzen zu können.

Den Abschluß bilden die zum Teil sehr umfänglichen, überregionalen Vorhaben zur Lehrevaluation verschiedener Institutionen, wie des Interdisziplinären Zentrums für Hochschuldidaktik (Bielefeld), der Hochschul-Informationssystem GmbH (Hannover) und des Centrums für Hochschulentwicklung (Gütersloh) sowie von studentischen Initiativen und Verbänden.

## 2.1 Projekte des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF, früher BMBW)

Ende 1991 wurde auf Initiative des Staatssekretärs im Bildungsministerium, Dr. Fritz Schaumann, mit einem Programm zu 'Qualität und Wettbewerb in der Lehre' begonnen. Dieses Programm hatte zum Ziel, Impulse im Bereich der Lehrbewertung zu setzen und Projekte der Lehrvaluation zu fördern.

Gefördert wurden im Rahmen dieses Programms in den letzten Jahren vor allem eine Reihe von Tagungen, durchgeführt von verschiedenen Organisationen und Verbänden:

1991

- Die Sommerschule der Gewerkschaft Erziehung und Wissenschaft (GEW);  
Thema: „*Qualität von Studium und Lehre*“.
- Das Symposium der Liberalen Hochschulgruppen (LHG);  
Thema: „*Die Evaluation von Lehrleistungen*“.
- Das Seminar des Deutschen Hochschulverbandes (DHV);  
Thema: „*Universitäre Lehre - Wirklichkeit und Erwartungen*“.
- Die Jahrestagung der Arbeitsgemeinschaft für Hochschuldidaktik (AHD);  
Thema: „*Qualität der akademischen Lehre*“.
- Die Zusammenkunft des Hochschullehrerbundes (HLB);  
Thema: „*Weiterentwicklung der Lehre*“.

1992

- Die Tagung des Hochschul-Information-Systems (HIS),  
Thema: „*Evaluation der Lehre im internationalen Vergleich*“.
- Das Arbeitstreffen an der Freien Universität Berlin;  
Thema: „*Evaluation der Lehre*“.
- Die Jahrestagung der Arbeitsgemeinschaft für Hochschuldidaktik (AHD);  
Thema: „*Verbesserung der Qualität der Lehre durch Evaluation*“;
- Das Fachgespräch der AG Hochschulforschung (Konstanz) und des BMBW;  
Thema: „*Studiensituation - Effizienz und Qualität*“.

1993

- Die Arbeitstagung der FU Berlin;  
Thema: „*Evaluation der Lehre*“.
- Das Forum der Hochschulrektorenkonferenz;  
Thema: „*Qualität von Lehre und Studium*“.

1994

- Tagung der Evangelischen Akademie Bad Boll;  
Thema: „*Hochschulökonomie*“.

1995

- Das Symposium der Medizinischen Fakultät der Universität Köln;  
Thema: „*Qualität der Lehre in der Medizinausbildung*“.

Weiter wurden seit 1991 mehrere Projekte ganz- oder teilfinanziert, die zum Teil die Entwicklung von Instrumenten für Befragungen von Studierenden zur Lehr-evaluation zum Ziel hatten:

- Professorenbefragung an der Universität Bochum, die sich mit dem Selbstverständnis der Dozenten als Lehrende auseinandersetzt.
- Befragung der Studierenden zum Thema „*Mehr Ehre für die Lehre*“ und „*Prüf den Prof*“ des RCDS.
- Befragung der Studierenden zum Thema „*Lehre unter der Lupe*“ der LHG.
- Wiederholte Befragung junger Beschäftigter zur Bewertung des Lehrangebots der betriebswirtschaftlichen Fachbereiche an Universitäten und Fachhochschulen durch das Institut der deutschen Wirtschaft.
- Methoden der Lehrberichtserstellung wurden vom Interdisziplinären Zentrum für Hochschuldidaktik (IZHD) der Universität Bielefeld zusammengestellt.
- Erprobung eines maschinenlesbaren Veranstaltungsfragebogens an der Technischen Universität Berlin.
- Erprobung eines Berichtssystems „*Studium und Lehre*“ von HIS an der Universität Hannover, Fachbereiche Gartenbau und Landschaftsarchitektur.
- Die Studie „*Qualität der Lehre*“ an der Universität Kaiserslautern.
- Das Pilotprojekt „*Profilbildung*“ der Hochschulrektorenkonferenz.
- Das Pilotprojekt „*Qualitätsbewertung im Hochschulbereich*“ der Hochschulrektorenkonferenz im Rahmen einer EU-Studie.
- Survey „*Studiensituation und studentische Orientierungen*“ (Erhebung an 22 Hochschulen) der AG Hochschulforschung an der Universität Konstanz.

Zu den meisten dieser Projekte liegen Veröffentlichungen vor, zum Teil von den Autoren, zum Teil durch das BMBW bzw. BMBF in seinen Schriftenreihen. Darüberhinaus hat das damalige BMBW im Jahre 1993 eine Zwischenbilanz: „*Qualität und Wettbewerb in der akademischen Lehre*“ vorgelegt (Bundesministerium für Bildung und Wissenschaft, 1993).

## 2.2 Projekte der Bundesländer

Die für die Hochschulen zuständigen Ministerien der Bundesländer haben prinzipiell drei Möglichkeiten, die Lehrevaluation zu beeinflussen: durch gesetzliche Regelungen, durch Bereitstellung von finanziellen Mitteln und durch Anregung und Entwicklung von Modellprojekten.

In den letzten Jahren sind eine Reihe von Änderungen zu den Hochschulgesetzen verabschiedet worden. Den Hochschulen ist in einigen Bundesländern eine 'Lehrberichtsspflicht' auferlegt worden (z.B. in Baden-Württemberg, Nordrhein-Westfalen, Sachsen). In bestimmten Zeitabständen (häufig ein bis drei Jahre) sollen sie Daten aus dem Bereich der Hochschullehre veröffentlichen (vgl. Kapitel 3). In einigen Bundesländern gibt es die Möglichkeit sowohl für ganze Hochschulen als auch für einzelne Fachbereiche oder gar studentische Initiativen, Projekte der Lehrevaluation, vor allem der studentischen Veranstaltungskritik, fördern zu lassen (z.B. Baden-Württemberg, Hessen, NRW, Rheinland-Pfalz).

Einige wenige Bundesländer haben auch Modellprojekte ins Leben gerufen. Diese werden meist mit Hilfe von außeruniversitären Institutionen durchgeführt und von den Länderministerien finanziert.

Evaluation der Lehre ist nicht in allen Wissenschaftsministerien ein gewichtiges Thema. Die neuen Bundesländer, die in den letzten Jahren gravierende Umstrukturierungen ihres Hochschulwesens durchgeführt haben, tun sich aus verschiedenen Gründen zum Teil schwer mit Projekten dieser Art. Andere Themenbereiche erscheinen problematischer, deren Lösung gilt als dringender. Manche Ministerien sind schlicht überlastet; manchmal klingen Bedenken durch, die Hochschullehrenden schon wieder beurteilen zu lassen.

Dennoch wird auch in den Wissenschaftsministerien im Osten Deutschlands über die Lehrevaluation diskutiert. In **Mecklenburg-Vorpommern** besteht vom Kultusministerium aus Interesse, sich mittelfristig an der niedersächsischen Evaluationsagentur zu beteiligen. In **Sachsen** ist das Staatsministerium für Wissenschaft und Kunst bestrebt, einen Minimalstandard für die dort vorgeschriebenen Lehrberichte einzuführen.

Aber auch in den alten Bundesländern läuft die Entwicklung sehr unterschiedlich. In **Saarland** bestehen z.Z. keine Projekte. Das **bayerische** Wissenschaftsministerium hat als ersten Schritt das Staatsinstitut für Hochschulforschung und Hochschulplanung beauftragt, eine Erhebung zu einem möglichen Instrumentarium der Lehrevaluation durchzuführen.

In **Berlin** liegt ein besonderer Fall vor. Dort wurden bereits frühzeitig Beschlüsse in den Hochschulkuratorien und Senaten zugunsten einer Lehrevaluation herbeigeführt. Von einer Festlegung auf eine Lehrberichtsspflicht wurde danach abgese-

hen. Im Frühjahr 1991 hatte Berlin bereits eine Landeshochschulstrukturkommission eingesetzt, die einen umfangreichen Bericht erarbeitete (Landeshochschulstrukturkommission Berlin, 1992). Auf Grundlage dieser Empfehlungen wurde im Sommer 1993 ein Hochschulstrukturplan im Senat beschlossen (Senatsverwaltung für Wissenschaft und Forschung, 1993). In diesem Bericht sind sowohl die künftigen Lehrberichte als auch Rahmenbedingungen für ein stärkeres Engagement in der Lehre genannt (Senatsverwaltung für Wissenschaft und Forschung, 1993, S. 72).

**Niedersachsen** ist bereit, die von der Landesrektorenkonferenz beschlossene Evaluationsagentur finanziell zu fördern. In **Rheinland-Pfalz** können aus dem Haushaltstitel zur Förderung und Verbesserung der Lehre auch Mittel für Evaluationsprojekte abgerufen werden, was in beachtlichen Umfang genutzt wird. **Bremen** fördert ebenfalls Projekte im Rahmen eines Programms zur Effizienzsteigerung in der Lehre.

In **Nordrhein-Westfalen** haben zwei vom Wissenschaftsministerium eingesetzte Arbeitsgruppen (Professoren, Studierende, Wissenschaftsministerium) den Entwurf eines 1990 von Seiten des Ministeriums vorgestellten Programms „Qualität der Lehre“ diskutiert. Im Mai 1992 erschien der Abschlußbericht (Ministerium für Wissenschaft und Forschung, 1992), in dem detailliert Möglichkeiten der Lehrverbesserung beschrieben werden. Nordrhein-Westfalen hat mit seinem Programm „Qualität der Lehre“ auch Finanzierungsmöglichkeiten für einzelne Projekte der Lehrbewertung vorgesehen, die ausgiebig genutzt werden. Das Angebot an die Universitäten, ein landesweites Modellprojekt durchzuführen, wurde von diesen noch nicht angenommen. Ein Modellprojekt an den Fachhochschulen hat dagegen bereits begonnen.

In **Baden-Württemberg** und in **Hessen** können zum einen Mittel für Projekte der Lehrevaluation abgerufen werden, zum anderen sind in beiden Ländern zusammen mit anderen Institutionen Modellprojekte initiiert worden.

Auch Hessen hat, im Herbst 1992, eine Hochschulstrukturkommission eingesetzt. In deren Bericht werden jedoch Themen wie Lehrverbesserung und -evaluation nur ganz am Rande erwähnt (Hessisches Ministerium für Wissenschaft und Kunst, 1995). Eine Broschüre des Ministeriums macht jedoch die Maßnahmen Hessens im Bereich der Lehrevaluation und -verbesserung deutlich (Hessisches Ministerium für Wissenschaft und Kunst, 1994).

**Schleswig-Holstein** und **Hamburg** waren die ersten beiden Länder, in denen Modellprojekte durchgeführt wurden. Die Planungsphase begann bereits 1992, die Evaluation wurde Anfang 1994 durchgeführt. In diesem Jahr soll in einem Gespräch bilanziert werden, was von den damaligen Empfehlungen zur Verbesserung der Lehrqualität umgesetzt wurde.

In der nachfolgenden Übersicht sind die Projekte und deren finanzielle Unterstützung der sechzehn Bundesländer zusammengestellt. Die Zusammenstellung beruht auf den Angaben der zuständigen Ministerien. Fünf Ministerien teilten mit, daß zur Zeit keine spezifischen Projekte vorliegen.

<b>Übersicht 1</b> <b>Projekte und Vorhaben der Bundesländer zur Lehrevaluation</b> (nach Auskunft der zuständigen Ministerien; Stand: Dezember 1995):	
<b>Bundesland</b>	<b>Projekte, Programme und finanzielle Unterstützung</b>
<b>Baden-Württemberg</b> Ministerium für Wissenschaft und For- schung	Modellprojekt mit HIS (Hannover): Stuttgart und Karlsruhe (Informatik), Tübingen und Heidelberg (Anglistik). Seit 1994 Arbeitsgruppe mit der LRK zur Vorberei- tung des Modellversuchs. Für personelle Unterstützung der Evaluationvorhaben ca. ½ Mill. DM zur Verfügung gestellt.
<b>Bayern</b> Staatsmin. für Unter- richt, Kultus, Wissen- schaft und Kunst	Das Staatsinstitut für Hochschulforschung und Hochschulpla- nung wurde beauftragt, eine Erhebung zu einen möglichen In- strumentarium der Lehrevaluation zu ermitteln.
<b>Berlin</b> Senatsverwaltung für Wiss. und Forschung	Über die Kuratorien wurden Beschlüsse erwirkt (1993), Lehrbe- richte und -evaluationen durchzuführen. Empfehlungen einer Hochschulstrukturkommission liegen vor (1993)
<b>Brandenburg</b> Ministerium für Wis- senschaft, Forschung und Kultur	Zur Zeit liegen noch keine Projekte vor.
<b>Bremen</b> Senator für Bildung, Wissenschaft, Kunst und Sport	Zur Zeit liegen noch keine Projekte vor.
<b>Hamburg</b> Behörde für Wissen- schaft und Forschung	Mit Schleswig-Holstein zusammen wurde ein Pilotprojekt von HIS (Hannover) an den Universitäten in Hamburg und Kiel durchgeführt und mit 90.000 DM unterstützt (Gutachter, Reise- kosten, Fachbereiche). Des weiteren wurden zusätzliche Evaluationsprojekte im Rah- men des Fonds für innovative Maßnahmen in der Lehre in den Jahren 1993 und 1994 gefördert.

<b>Hessen</b> Ministerium f. Wissenschaft und Kunst	1993 wurde eine „exemplarische Fachanalyse“ der Biologie und der Germanistik eingeleitet, 1994 kamen Erziehungswissenschaften, Physik und Wirtschaftswissenschaften hinzu. Das Ministerium finanziert ein Projekt zur Erprobung von Lehrberichten an Fachhochschulen mit IZHD (Bielefeld). 1993 wurden 482.000DM für Evaluationsmaßnahmen bewilligt, 1994 waren es 438.000 DM.
<b>Mecklenburg-Vorpommern</b> Kultusministerium	Es besteht Interesse, sich mittelfristig an der Niedersächsischen Evaluationsagentur zu beteiligen
<b>Niedersachsen</b> Ministerium f. Wissenschaft und Kultur	Mittel für Evaluationsprojekte können über die Niedersächsische Evaluationsagentur beantragt werden. Es stehen jährlich bis zu 900.00 DM bereit.
<b>Nordrhein-Westfalen</b> Ministerium f. Wissenschaft und Forschung	Ein Projekt mit dem IZHD (Bielefeld) zur Erprobung von Lehrberichten an den Fachhochschulen wird durchgeführt. Mittel für Evaluationsprojekte können aus dem QdL-Programm beantragt werden. Von den 2 Mill. DM, die bereitgestellt wurden, sind 1993 und 1994 jeweils eine Mill. für Projekte der studentischen Veranstaltungskritik bewilligt worden. Allen Hochschulen, die sich an einem Evaluationsprojekt mit CHE (Gütersloh) beteiligen, wird 40.000 DM Unterstützung angeboten.
<b>Rheinland-Pfalz</b> Ministerium f. Wissenschaft und Weiterbildung	Aus dem Haushaltstitel zur Förderung und Verbesserung der Lehre (900.000 DM) können auch Mittel für Evaluationsprojekte abgerufen werden. 1993 wurden 481.000 DM bewilligt, 1994 390.000 DM.
<b>Saarland</b> Ministerium f. Wissenschaft und Kultur	Zur Zeit liegen noch keine Projekte vor.
<b>Sachsen</b> Staatsministerium für Wissenschaft u. Kunst	Bestrebungen, einen Minimalstandard für die Lehrberichte einzuführen.
<b>Sachsen-Anhalt</b> Kultusministerium	Zur Zeit liegen noch keine Projekte vor.
<b>Schleswig-Holstein</b> Ministerium f. Wissenschaft und Forschung	Modellprojekt gemeinsam mit Hamburg unter Leitung von HIS (Hannover) an den Universitäten Kiel und Hamburg. Das Ministerium hat dafür 120.000 DM bereitgestellt.
<b>Thüringen</b> Ministerium für Wissenschaft, Forschung und Kultur.	Zur Zeit liegen noch keine Projekte vor.

## **2.3 Hochschulrektorenkonferenz, Kultusministerkonferenz, Wissenschaftsrat**

### **Hochschulrektorenkonferenz (HRK)**

Nachdem die Hochschulrektorenkonferenz viele Jahre lang Reformmaßnahmen von Seiten der Hochschulen ohne vorherige Erhöhung der Finanzmittel abgelehnt hatte, wurde 1993 ein Wendepunkt deutlich. Ihr Präsident, Prof. Hans-Uwe Erichsen, erklärte in seinem Jahresbericht 1993: "In den Hochschulen hat sich demgegenüber zunehmend die Einsicht durchgesetzt, daß die Verfolgung als richtig erkannter Ziele nicht vom Tun oder Unterlassen und damit auch nicht vom Versagen des Staates abhängig gemacht werden darf. Viele Hochschulen waren daher bestrebt, auf der Grundlage einer ihrem Selbstverständnis als autonome, mit dem Recht der Selbstverwaltung ausgestattete Einrichtungen adäquaten, fortschreitenden Emanzipation vom Staat wieder stärker als in der Vergangenheit das Gesetz des Handelns in die Hand zu bekommen" (Erichsen, 1993, S. 2).

Hier wurde eine Meinungsänderung deutlich. Reformen werden als notwendig betrachtet, selbst wenn von staatlicher Seite keine zusätzlichen finanziellen Mittel zu erwarten sind. Dieser neue Ansatz hat sich in der Folge auf die Einstellung der Hochschulen zur Lehrevaluation ausgewirkt.

Die HRK hat eine Arbeitsgruppe "Qualität der Lehre" eingesetzt. Als Schwerpunkte von Evaluationsmaßnahmen werden vor allem die Selbstevaluation der Fachbereiche sowie regelmäßige hochschulübergreifende Evaluationen "in Verantwortung der Hochschulen" angestrebt.

Wegen befürchteter staatlicher Einflußnahme wird immer wieder betont, daß: "... die Evaluation eine eigene Aufgabe der Universität und der an ihr Lehrenden und Lernenden ist, keine Aufgabe der Hochschulverwaltung, der Wissenschaftsministerien oder besonderer externer Evaluationsinstanzen. Dennoch kann Hilfe von außen nützlich (gelegentlich auch notwendig) sein, um den entscheidenden Anstoß zu liefern, um Austausch von Erfahrungen zu verbessern und schließlich um die Umsetzung zu beschleunigen ... (Hochschulrektorenkonferenz, 1994, S. 29).

Die HRK hat schließlich im März 1995 eine Empfehlung zur Umsetzung der Lehrberichtsverordnungen und zur Evaluation der Lehre im allgemeinen an alle Rektorate geschickt (Hochschulrektorenkonferenz, 1995a).

Diese Empfehlung besteht aus mehreren Teilen:

1. Allgemeine Erläuterung zur Lehrevaluation nach dem niederländischen Modell,
2. Kostenfrage,

3. Muster zur Gliederung der Lehrberichte,
4. Fragebogen für Studierende zu Vorlesungen,
5. Dozentenfragebogen.

Weiter wurde die Einrichtung einer Evaluationsagentur bei der HRK vorgeschlagen.

Die Stellungnahmen der Hochschulen dazu lassen sich in drei Gruppen gliedern: (1) Eindeutige Zustimmung zur Empfehlung, gerade von den Hochschulen des norddeutschen Evaluationsverbundes; (2) eindeutige Ablehnung an zwei Universitäten wegen möglicher Bürokratisierung der Lehrevaluation und aus Angst vor ausufernder Gutachtertätigkeit und (3) grundsätzliche Zustimmung von dem überwiegenden Teil der Hochschulen mit Verbesserungsvorschlägen.

Am 3.7.1995 verabschiedete das Plenum der Hochschulrektorenkonferenz eine Vorlage zur Lehrbewertung, die sich für eine „unabhängige und staatsferne Durchführung der Evaluation in Eigenverantwortung der Hochschulen, Fakultäten und Fachbereiche“ ausspricht (Hochschulrektorenkonferenz, 1995b).

Vorgesehen ist ein zweistufiges Verfahren (orientiert am niederländischen Modell) mit interner Selbstevaluation und späterer Beurteilung durch Fachkollegen (externer Evaluation). Für die interne Evaluation ist die Erstellung eines Lehrberichts vorgesehen, der auch Daten aus der Befragung von Lehrenden und Studierenden beinhaltet (Hochschulrektorenkonferenz, 1995c).

### **Kultusministerkonferenz der Länder (KMK)**

Gemeinsam verabschiedeten **Kultusminister- und Hochschulrektorenkonferenz** 1993 eine Erklärung zur Studienstrukturreform (Kultusminister- und Hochschulrektorenkonferenz, 1994). Dort wurde erstmalig darauf hingewiesen, daß Lehrdefizite zu einer Studienzeitverlängerung führen können, und daß die Gründe sowohl in der "mangelnden Abstimmung im Lehrangebot" als auch in der "mangelnden didaktischen Befähigung der Lehrenden" liegen. Eine Evaluation der Lehre wurde zum ersten Mal als "unerlässlich" bezeichnet.

Die HRK/ KMK-Arbeitsgruppe „Weiterentwicklung der Struktur des Hochschulwesens“ hat Empfehlungen zu quantitativen Indikatoren in den Hochschulen beschlossen (6.12.94/ 8.3.95). Empfohlen wird ein tiefgegliedertes Berichtssystem auf Hochschulebene. Diese Daten sollen in einem zweiten Schritt von den Statistischen Landesämtern gesammelt und aufbereitet und auf Bundesebene vom Statistischen Bundesamt zusammengeführt werden.

Trotz dieser Arbeitsgruppe sind die in einigen Gesetzen oder Verordnungen geforderten quantitativen Indikatoren zur Lehrqualität relativ unterschiedlich. Eine

Absprache, die sich auch auf die Gesetzgebung zur Lehrberichtserstattung auswirkt, wäre wünschenswert (vgl. Kapitel 3).

### **Wissenschaftsrat**

Der Wissenschaftsrat hat an verschiedenen Hochschulen ein Modellprojekt mit zwei beteiligten Fächern (Physik und Betriebswirtschaftslehre) mit dem Ziel durchgeführt, ein exemplarisches Verfahrensinstrumentarium zu erproben. Das angewendete Verfahren war angelehnt an das niederländische Modell zur Lehr-evaluation.

Die Begehungen der Fachbereiche (jeweils 1,5 Tage) fanden im Wintersemester 94/95 und im Sommersemester 95 statt. Das Verfahren ist inzwischen abgeschlossen. Als problematisch stellte sich die große Unterschiedlichkeit der einzelnen Hochschulen heraus, ihre unterschiedlichen Standards und die unterschiedliche Problemwahrnehmung.

Die an Modellprojekten im Fach *Betriebswirtschaftslehre* beteiligte Hochschulen sind:

- Fachhochschule Frankfurt
- Fachhochschule für Wirtschaft Berlin
- Universität München
- Universität Köln
- Universität Kiel
- Universität (GHS) Siegen

Die an Modellprojekten im Fach *Physik* beteiligten Hochschulen sind:

- Universität Freiburg
- Technische Universität Berlin
- Universität Saarbrücken
- Universität Osnabrück
- Universität Leipzig

Auf Grundlage der empirischen und methodischen Basis sowie der praktischen Erfahrungen, die in diesen Projekten gewonnen wurden, veröffentlichte der Wissenschaftsrat im Januar 1996 die „Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation“ (Wissenschaftsrat, 1996). Im Anhang dieser Empfehlungen ist beispielhaft ein recht umfänglicher „Leitfaden zur Evaluation von Studium und Lehre“ angeführt.

## **2.4 Kooperationen und Verbände**

### **Verbund der norddeutschen Hochschulen**

1994 beschlossen die akademischen Senate der Universitäten Kiel, Rostock, Hamburg, Bremen und Oldenburg für drei Jahre die Evaluation von Studium und Lehre im Verbund zu erproben. Hilfestellung versprach dabei die Universität Groningen (Niederlande) zu leisten.

Der Verbund dieser Hochschulen aus fünf norddeutschen Bundesländern, auch 'Nordverbund' genannt, nutzte die Ergebnisse des mit dem Hochschul-Informationssystem (HIS) durchgeführten Modellprojektes an den Universitäten Hamburg und Kiel in den Fächern Biologie und Germanistik.

Die anderen Partnerhochschulen evaluierten ebenfalls diese beiden Fächer. Im Sommersemester 1995 wurden an den genannten Hochschulen zusätzlich Wirtschaftswissenschaften und Informatik mit einem ausgesprochen umfangreichen Fragebogen bewertet.

Anders als in vielen anderen Projekten sollen beim Evaluationsverfahren des „Nordverbundes“ Studierende in den Gutachterkommissionen mit vertreten sein. Ebenfalls sehr spezifisch ist der Gedanke einer auswertenden Konferenz, auf der nach Beendigung der externen Evaluation Fachvertreter und -vertreterinnen die Möglichkeit haben, sowohl die Empfehlungen als auch Selbstverpflichtungen zu diskutieren.

### **Niedersächsische Evaluationsagentur**

Die Niedersächsische Evaluationsagentur wird von den niedersächsischen Hochschulen getragen. Das Land hat die Finanzierung der Stellen- und Sachausstattung sowie des laufenden Mittelbedarfes der Agentur übernommen. Die zentrale Evaluationsagentur der niedersächsischen Hochschulen hat ihre Arbeit im Oktober 1995 aufgenommen.

Die Evaluationsagentur hat die Aufgabe, eine systematische und hochschulübergreifende Evaluation von Studiengängen auf Fachbereichsebene zu initiieren und zu begleiten.

Sie soll die an der Evaluation Beteiligten zusammenführen und im Benehmen mit den Hochschulen die Auswahl der externen Gutachter und Gutachterinnen übernehmen. Zu ihrem Tätigkeitsbereich gehört auch die Veröffentlichung von zusammenfassenden Evaluationsberichten (im Sinne eines Gesamteindrucks).

## **Netzwerk Bildungsforschung**

Eine Zusammenarbeit über Ländergrenzen hinweg zeichnet sich auch im Bereich der Lehrevaluation ab. So sind z.B. am 'Netzwerk Bildungsforschung' Bremen, Nordrhein-Westfalen, Niedersachsen, Niederlande und Flandern (Belgien) beteiligt. Zur Aufgabe hat sich dieses Netzwerk gesetzt, Daten in den beteiligten Ländern im Bereich der Hochschule zu sammeln und einen intensiven Informationsaustausch voranzutreiben.

Für diese Aufgabe wurden vier Arbeitsgruppen gebildet: Hochschulfinanzierung, Hochschulstruktur (Studienangebote, etc.), Übergang Schule/ Hochschule und Lehrevaluation. Dabei sind auch andere Institutionen beteiligt. So arbeitet z.B. HIS indirekt über Niedersachsen mit, das Wissenschaftliche Sekretariat für Hochschulreform direkt über Nordrhein-Westfalen.

### **2.5 Lehrevaluation mit externen Institutionen**

Institutionen, die bereits in der einen oder anderen Weise in der Hochschulforschung verankert sind, aber auch neugegründete Organisationen, haben die Lehrevaluation als ein maßgebliches Thema der nächsten Jahre begriffen. Gemeinsam mit den Hochschulen, teils mit, teils ohne Beteiligung eines Ministeriums, werden Modellprojekte und Beratungen sehr unterschiedlichen Umfangs durchgeführt.

#### **Interdisziplinäres Zentrum für Hochschuldidaktik (IZHD)**

Das Projekt des IZHD der Universität Bielefeld „Entwicklung und Erprobung von konsensfähigen Parametern für die Selbstevaluation von Fachbereichen (als Grundlage von Lehrberichten)“, unterstützt vom BMBF, hat im März 1993 den Abschlußbericht vorgelegt (Webler et. al., 1993).

Daran schloß sich von August 1993 bis Januar 1995 das Projekt „Modellhafte Erprobung von Lehrberichten als Teil der Organisationsentwicklung“ an. Beteiligt waren sechs universitäre Fachbereiche (I).

Ab dem Frühjahr 1994 bis zum März 1995 begann das Projekt „Erprobung von Lehrberichten an hessischen Fachhochschulen“ (II).

Ab 1.12.94 wurde mit dem Projekt „Erprobung und Einführung von Lehrberichten an Fachhochschulen in NRW“ begonnen, das bis zum 30.9.95 abgeschlossen sein soll (III). Einige weitere Projekte kamen hinzu (IV).

Übersicht beteiligter Hochschulen und Fächer am Projekt „Erprobung von Lehrberichten“ durch das IZHD in Bielefeld:

(I)	Uni Bielefeld	Philosophie
		Fakultät für Soziologie
	GH Kassel	Fachbereich Sozialwesen
		Architektur
		Stadt- und Landschaftsplanung
	TU Braunschweig	Bauingenieur- und Vermessungswesen
		Elektrotechnik
(II)	FH Gießen-Friedbg.	FB Elektrotechnik II
		FB Techn. Gesundheitswesen
	FH Wiesbaden	FB Weinbau- und Getränketechnik
		FH Frankfurt
		FB Bauingenieurwesen
	FH Darmstadt	FB Information und Dokumentation
(III)	FH Aachen	FB Wirtschaft
	FH Bielefeld	FB Architektur / Bauingenieurwesen
	FH Bochum	FB Maschinenbau
	FH Dortmund	FB Informatik
	FH Iserlohn	FB Elektrotechnik
	FH Münster	FB Chemieingenieurwesen
	FH Niederrhein	FB Textil- und Bekleidungstechnik
	FH Gelsenkirchen	FB Versorgungs/ Entsorgungstechnik
	FH Köln	FB Sozialpädagogik
		eine zehnte Fachhochschule soll hinzukommen
(IV)	GH Siegen	FB I Sozialwissenschaften
	Hamburg	Hochschule für Wirtschaft und Politik

## **Hochschul-Informationen-System (HIS)**

Die Hochschul-Informationen-System GmbH (HIS) wandte sich frühzeitig der Lehrevaluation als neuem Handlungsfeld zu und erstellte eine zweibändige Dokumentation zu Projekten und Instrumenten der Lehrbewertung an Hochschulen (Reissert, 1992). Es folgte 1993 ein Pilotprojekt zur fachbereichsinternen Selbstevaluation an der Universität Hannover, später das Pilotprojekt an den Universitäten Hamburg und Kiel (Reissert, 1994). In Baden-Württemberg werden zwei Studiengänge an zwei Hochschulen evaluiert.

Übersicht zu den Evaluationsprojekten von HIS:

Hannover	Gartenbau, Landschaftsarchitektur
Hamburg	Biologie und Germanistik
Kiel	Biologie und Germanistik
Stuttgart und Karlsruhe	Informatik
Tübingen und Mannheim	Anglistik

Nach den Erfahrungen von HIS ist die Veränderungsbereitschaft, vor allem auf der relativ niedrigen organisatorischen Ebene der Fachbereiche, an den Hochschulen höher als erwartet. Sobald jedoch auf der Inhaltsebene Veränderungsbedarf entstünde, u.U. zusätzlich noch finanzielle Umverteilungen notwendig würden, seien Veränderungen schwierig zu erreichen.

## **Centrum für Hochschulentwicklung (CHE)**

Das 1995 neu eingerichtete Centrum für Hochschulentwicklung (CHE) in Gütersloh hat u.a. eine „Moderatorrolle“ im Evaluationsverbund norddeutscher Hochschulen (Bremen, Hamburg, Kiel, Oldenburg und Rostock) übernommen. Ziel dieser Zusammenarbeit ist die Qualitätssicherung durch interne und externe Leistungsbewertung. Da an den genannten Hochschulen die Datenerhebungen bereits vorliegen, geht es gegenwärtig vor allem um die Entwicklung einer Struktur zur Umsetzung der Evaluationsergebnisse und -empfehlungen.

Darüberhinaus versteht sich das CHE auch im Bereich der Lehrevaluation als Informationsdrehscheibe, um auf Anfragen hin Kontakte und Informationen zu vermitteln. In diesem Sinn arbeiten CHE und HIS in einem vom Stifterverband für die deutsche Wissenschaft geförderten Projekt 'Qualitätsförderung in der Lehre' zusammen, das eine Bestandsaufnahme der aktuellen Praxis zur Aufgabe hat. Im Rahmen des Projekts sollen unter anderem Strukturelemente der Berichterstattung zu Studium und Lehre bestimmt werden.

## 2.6 Studentische Evaluationsprojekte

Im Bereich der Lehrevaluation sind Studierende vor allem in der praktischen Durchführung aktiv. Von den über 60 Evaluationsprojekten, die das Hochschul-Informationssystem (1992) in seiner Dokumentation zusammentrug, stammt der überwiegende Teil von Fachschaften und studentischen Initiativen.

Als Studentenverbände haben der Ring Christlich-Demokratischer Studenten (RCDS) und die Liberalen Hochschulgruppen (LHG) vom BMBW finanziell unterstützte Projekte studentischer Veranstaltungskritik durchgeführt (siehe Bundesministerium für Bildung und Wissenschaft, 1993; Ring Christlich-Demokratischer Studierender, 1992; Liberale Hochschulgruppen, 1992).

Gelegentlich bilden sich mehr oder weniger langfristige Initiativen, die versuchen, die studentischen Evaluationsprojekte an der jeweiligen Hochschule zu koordinieren (z.B. das Studentische Evaluationsprojekt an der FU Berlin, das Studentische Koordinationszentrum in Hamburg oder die Koordinationsstelle Studentischer Evaluation in Hannover). Diese Initiativen sind häufig bei dem entsprechenden Allgemeinen Studierendenausschuß (AStA) angesiedelt und stark personengebunden. Verläßt der oder die Studierende die Hochschule oder kann das Projekt aus anderen Gründen nicht fortsetzen, bedeutet das häufig ein zumindest vorübergehendes Aussetzen.

Studierende sehen die studentische Lehrbewertung jedoch nicht als einzige Möglichkeit, die Lehrqualität zu verbessern, wie die Aktion „Professoren-TÜV“ am Otto-Suhr-Institut der FU Berlin zeigt, bei der die Fachschaft Politikwissenschaften eine Lehrenden-Beratung mit intensiven Einzelgesprächen organisierte. In ähnlicher Weise wurde von Studierenden an der Universität Bremen die Form einer „dialogischen Evaluation“ erprobt (vgl. Handbuch Hochschullehre, Bd. 1, D3.1, 1995).

Die Koordination studentischer Aktivitäten im Bereich der Lehrevaluation außerhalb der einzelnen Hochschule auf Landes- oder Bundesebene erscheint schwierig. Dennoch gibt es immer wieder Ansätze dazu. So lud das Studentische Evaluationsprojekt (StEP) der FU Berlin im Frühjahr 1993 zu einer Evaluationstagung der Studierenden ein. Beim 'freie zusammenschluß der studierendenschaften' (fzs), der Nachfolgeorganisation des VDS, verglich eine Arbeitsgruppe die Gesetzgebung der Länder zum Thema Lehre und Lehrevaluation.

Außer auf die immer wieder mehr oder weniger spontanen Forderungen oder Aktionen Studierender ist als besondere studentische Initiative auf die Bemühung zur Gründung einer Deutschen Lehrgemeinschaft zu verweisen (AStA der Universität Münster).

### 3 Grundlagen der Lehrevaluation in den Bundesländern

Ähnlich unterschiedlich wie die Beteiligung der einzelnen Bundesländer an Projekten zur Lehrevaluation sind deren gesetzliche Grundlagen in bezug auf Umfang und Detailliertheit für die Lehrberichtspflicht.

Nachfolgend wird im Überblick festgehalten, ob in den einzelnen Bundesländern eine Verpflichtung zur Lehrberichterstattung festgeschrieben ist und wie diese geregelt wird. Das betrifft zum Beispiel die Verantwortlichkeit für diese Berichterstattung, die vorgesehenen Zeiträume und die verlangten Daten und Informationen.

Danach sind die gesetzlichen Bestimmungen oder Erlasse der Bundesländer zu den Lehrberichten zusammengestellt, wobei die jeweilige Gesetzesquelle angeführt und die entsprechenden Paragraphen wiedergegeben werden.

Schließlich wird ein Fazit über die länderspezifischen Bestimmungen zu den Lehrberichten und Evaluationsdaten gezogen, wobei sich deren Uneinheitlichkeit als Problem herausstellt.

#### 3.1 Verpflichtungen zur Lehrberichterstattung

In folgenden Bundesländern gibt es keine Verpflichtung, einen Lehrbericht vorzulegen: in **Bayern**, **Brandenburg** und **Thüringen**. Das Land Brandenburg plant allerdings, eine Lehrberichtspflicht noch im Jahre 1995 per Erlaß festzulegen.

Auch **Bremen** hat zwar keine gesetzlich festgelegte Lehrberichtspflicht, es besteht jedoch eine Vereinbarung über die Vorlage von Lehrberichten zwischen dem Senator und den Hochschulen.

In **Berlin** haben Kuratoriumsbeschlüsse einen Lehrbericht verpflichtend vorsehen, in **Hessen** wurde das durch einen Erlaß erwirkt.

**Hamburg** hat im Grunde genommen die älteste „Berichtspflicht“. Bereits im Jahre 1978 verpflichtete die damalige Fassung des Hochschulgesetzes die Hochschulen, in dreijährigen Abständen über die Studienreform zu berichten.

Dieses Überbleibsel der siebziger Jahre, das auch noch in der jetzigen Fassung zu finden ist, könnte eine Grundlage zu einer Lehrberichtspflicht abgeben, wird aber nicht in diesem Sinne genutzt. Jedoch wird künftig im Rahmen der „Erläuterungen zum Haushaltsplan“ eine Vielzahl an statistischen Daten von den Hochschulen gefordert.

In **Nordrhein-Westfalen** ist gesetzlich festgelegt, daß der Dekan/ die Dekanin für die Erstellung eines Lehrberichts verantwortlich ist. Dies deckt sich mit der Politik des Wissenschaftsministeriums, die Position des Dekan/ der Dekanin zu stärken und sie gleichzeitig in größerem Maße für den Lehrbetrieb verantwortlich zu machen. Eine nähere Regelung zu Inhalt, Daten und Abfolge des Lehrberichts existiert jedoch nicht.

In **Baden-Württemberg** liegt die Betonung dagegen auf dem Fakultätsrat. Dort sollen auch Lehrbewertungen besprochen und mit Hilfe einer Studienkommission Verbesserungsvorschläge entwickelt werden. Ausschließlich wird festgelegt, daß im Falle von durchgeführten Lehrbewertungen diese samt einer Stellungnahme der Lehrenden im Lehrbericht aufgeführt werden sollen.

Im **Saarland** wurde im Juni 1994 beschlossen, daß sowohl die einzelnen Fachbereiche als auch die Hochschulen *jährlich* einen Lehrbericht vorzulegen haben.

In **Rheinland-Pfalz** wurde im Mai 1995 eine gleiche Festlegung getroffen, wonach Lehrberichte vorzulegen sind.

Andere Länder haben bereits in den Gesetzestexten detailliert ausgeführt, welche Daten, aufgearbeiteten Ergebnisse und Kommentare sie in den Lehrberichten erwarten.

**Niedersachsen** legte im Januar 1994 fest, daß die dortigen Hochschulen *in angemessenen Zeitabständen* einen Bericht über die Situation und Entwicklung im Bereich von Lehre und Studium vorlegen müssen. Darin sollten folgende Daten enthalten sein: *Entwicklung der Zahl der Studierenden und der Hochschulabschlüsse* und über die *Studienzeiten* und *Studienbedingungen*.

In **Sachsen** müssen die Hochschulen bereits seit dem August 1993 *jährlich* Lehrberichte vorlegen. Außer über allgemeine Maßnahmen zu Lehrevaluationen und der Stellungnahme der Lehrenden über Verbesserungsmaßnahmen, muß für jeden Studiengang über die *Zahl der Studienbewerber, Studienanfänger* und *Absolventen* sowie über die *Zahl der erfolgreichen und nicht erfolgreichen Abschluß- und Zwischenprüfungen* berichtet werden.

In **Sachsen-Anhalt** wurde im Oktober 1993 eine im Wortlaut fast gleiche Lehrberichtspflicht wie in Sachsen vorgesehen, die jedoch von einem zweijährigen Abgabeturnus ausgeht.

**Schleswig-Holstein** sieht seit dem April 1995 einen Lehrbericht im Turnus von *zwei Studienjahren* vor. Allgemein ist genannt, daß er für jeden Studiengang die für eine „Beurteilung von Lehre, Studium, Hochschulprüfungen und staatliche Prüfungen wesentlichen strukturellen und quantitativen Daten und Aussagen“ enthalten soll.

Detailliert sollen Angaben zur Organisation der Lehre, zu den *Prüfungen und ihrer Dauer*, sowie jeweils *getrennt nach dem Geschlecht* zum *Studienerfolg*, zur *Studiendauer*, zum *Studienabbruch* und zum *Fachwechsel* enthalten sein.

**Mecklenburg-Vorpommern** hat im Februar 1994 einen Lehrbericht im Abstand von *zwei Jahren* beschlossen. Außer einer allgemeinen Aufforderung, über die Lehrqualität zu berichten, werden nach Studiengang geordnet folgende Daten verlangt: die *Zahl der Studienbewerber*, der *Studienanfänger* und *Absolventen* sowie der *erfolgreichen und nicht erfolgreichen Zwischen- und Abschlußprüfungen*, *Zahl der Studienabbrecher* und die *Kapazitäten im Lehr- und Prüfungsbetrieb*.

Neben diesen deskriptiven Zahlen wird auch eine gewisse statistische Ausarbeitung verlangt: Anteil der Studenten, die in der *Regelstudienzeit* das Studium abgeschlossen haben, und das *Verhältnis* der *Studienanfänger* zu der Zahl der *Absolventen*.

**Hessen** hat per Erlaß im März 1994 eine Lehrberichtspflicht eingeführt. Die ersten Berichte mußten danach im Februar 1995 erfolgen. Das geforderte Datenmaterial ist umfangreich: zum einen sind es u.a. Beschreibungen von Maßnahmen zur Verbesserung der Lehre sowie zu innovativen Maßnahmen der curricularen Erneuerung und, falls bekannt, Informationen über den Verbleib der Absolventen.

Weiter werden verschiedene statistische Daten eingefordert: *Gesamtzahl der Studierenden eines Studienganges*, *Anzahl der Studierenden in der Regelstudienzeit*, *Anzahl der Studienanfänger im 1. Fachsemester pro Studienjahr*, *Fachstudien-dauer bis zur Diplomvorprüfung oder Zwischenprüfung*, *durchschnittliche Studiendauer und Median*, *Gesamtzahl der Absolventen pro Studiengang und Studienjahr*, *Schwundquote* ("Übergangsquote", erstellt nach dem "Hamburger Verfahren"), *Absolventenquote* ("Erfolgsquote") und *Anzahl der Promotionen pro Studienfach*. Statistische Auffälligkeiten sollen von den Hochschulen im Bericht kommentiert werden.

### 3.2 Überblick zu den gesetzlichen Bestimmungen

Auf den folgenden Seiten wird ein Überblick über die gesetzlichen Bestimmungen zu den Lehrberichten in den einzelnen Bundesländern gegeben (Stand Dezember 1995).

Eine Bilanz dieser Bestimmungen, vor allem unter dem Gesichtspunkt der Vergleichbarkeit der Verfahren, der verlangten Daten und Berichtszeiträume, schließt sich an diese Übersicht an.

## Übersicht 2 Gesetzgebung der Länder zu Lehrberichten

<i>Bundesland</i>	<i>Gesetzgebung zum Thema Lehrberichte</i>
Baden-Württemberg	<p>Im <b>Gesetz über die Universitäten im Lande Baden-Württemberg</b> (UG) vom 10.01.95 ist im §25 <b>Fakultätsrat</b>, Absatz 4 u.a. festgelegt: „Zu den Aufgaben der Studienkommission gehört es insbesondere, Empfehlungen zur Weiterentwicklung von Gegenständen und Formen des Studiums ... zu erarbeiten sowie Verfahren zur Bewertung und Verbesserung der Qualität der Lehre unter Einbeziehung studentischer Veranstaltungskritik zu entwickeln. Die Studienkommission erarbeitet in regelmäßigen Abständen einen Bericht über die Entwicklung von Lehre, Studium und Prüfungen.</p> <p>Der Bericht enthält für den Berichtszeitraum auch Angaben über die Bewertung des Lehrangebotes in den einzelnen Studiengängen insbesondere über Befragungen der Studierenden zur Qualität der Lehre und die Stellungnahme des Lehrkörpers zu den Ergebnissen der Befragung: der Bericht bezieht auch die Ergebnisse externer Bewertungen ein. Der Fakultätsrat gibt der Fachschaft Gelegenheit, zu dem Bericht Stellung zu nehmen. Der Fakultätsrat erörtert den Bericht der Studienkommission und ergreift geeignete Maßnahmen zur Verbesserung der Lehre.“</p>
Bayern	Keine Festlegung über Lehrberichte.
Berlin	Alle Berliner Hochschulen verfügen über Senats- und Kuratoriumsbeschlüsse, meist aus dem Jahre 1993, legen eine Lehrberichtspflicht verbindlich fest.
Brandenburg	Erlaß vermutlich noch in diesem Jahr (1995).
Bremen	Keine Festlegung über Lehrberichte.
Hamburg	Das <b>Hamburgische Hochschulgesetz</b> von 1994 legt in §66 <b>Aufgaben</b> , Abschnitt 3 fest: „(3) Die Hochschulen berichten in dreijährigen Abständen über die Studienreform. Die zuständige Behörde kann darüber hinaus zu einzelnen Bereichen Berichte anfordern.“
Hessen	Der <b>Erlaß W I 1 - 907/10 - 65</b> - (Einführung von Lehr- und Studienberichten an den hessischen Universitäten) vom 05.04.94 sieht erstmals Lehr- und Studienberichte vor; sie sollten bis zum 15.02.95 dem Ministerium vorliegen. Vorgesehen ist, daß die einzelnen Fakultäten eigene Lehrberichte anfertigen, die dann auf Hochschulebene in einen zusammenfassenden Lehr- und Studienbericht eingehen. Die Hochschulen sollen sich eigene Vorgaben setzen, um eine Vergleichbarkeit zu ermöglichen.

<p>Hessen (Fortsetzung)</p>	<p>Der Lehr- und Studienbericht soll folgende Angaben beinhalten:</p> <ul style="list-style-type: none"> <li>- Gesamtlehrkapazität unter besonderer Berücksichtigung der Entwicklung des Betreuungsverhältnisses Lehrende/ Lernende,</li> <li>- Sicherstellung des Lehrangebotes und der Einhaltung der Lehrverpflichtung,</li> <li>- Schwerpunkte der Maßnahmen zur Verbesserung der Lehre sowie zu innovativen Maßnahmen der curricularen Entwicklung,</li> <li>- und falls bekannt über den Verbleib der Absolventen.</li> </ul> <p>Weiter werden einige statistische Daten eingefordert:</p> <ul style="list-style-type: none"> <li>- Gesamtzahl der Studierenden eines Studienganges,</li> <li>- Anzahl der Studierenden in der Regelstudienzeit,</li> <li>- Anzahl der Studienanfänger im 1. Fachsemester pro Studienjahr,</li> <li>- Fachstudiendauer bis zur Diplom-Vorprüfung oder Zwischenprüfung, durchschnittliche Studiendauer und Median,</li> <li>- Gesamtzahl der Absolventen pro Studiengang und Studienjahr,</li> <li>- Schwundquote ("Übergangsquote", gemäß "Hamburger Verfahren"),</li> <li>- Absolventenquote ("Erfolgsquote"),</li> <li>- Anzahl der Promotionen pro Studienfach.</li> </ul> <p>Statistische Auffälligkeiten sollen von den Hochschulen im Bericht kommentiert werden.</p>
<p>Mecklenburg-Vorpommern</p>	<p>Das <b>Gesetz über die Hochschulen des Landes Mecklenburg-Vorpommern</b> (Landeshochschulgesetz LHG) vom 9.2.94 legt in § 11 <b>Lehrangebot, Lehrbericht</b>, Absatz 3 fest: „Die Hochschule überprüft in regelmäßigen Abständen, mindestens alle zwei Jahre, unter Beteiligung von Studenten und Absolventen die Qualität der Lehrtätigkeit ihrer Fachbereiche und berichtet darüber (Lehrbericht). Der Bericht soll über Maßnahmen zur Bewertung und Verbesserung der Lehre berichten. Er soll für jeden Studiengang insbesondere:</p> <ol style="list-style-type: none"> <li>1. Aussagen über die Qualität der Lehrleistungen,</li> <li>2. die Zahl der Studienbewerber, der Studienanfänger und Absolventen sowie der erfolgreichen und nicht erfolgreichen Zwischen- und Abschlußprüfungen,</li> <li>3. das Verhältnis der Zahl der Studienanfänger zu der Zahl der Absolventen,</li> <li>4. den Anteil der Studenten, die in der Regelstudienzeit das Studium abgeschlossen haben,</li> <li>5. die Zahl der Studenten, die das Studium abgebrochen haben,</li> <li>6. die Kapazitäten im Lehr- und Prüfungsbetrieb</li> </ol> <p>ausweisen. Die Landesrektorenkonferenz kann weitere Elemente des Berichtes und Einzelheiten der Berichterstattung vorschlagen. Die Kultusministerin kann durch Rechtsverordnung den Mindestinhalt des Berichts bestimmen. Der Lehrbericht ist im Senat zu behandeln, der Kultusministerin vorzulegen und zu veröffentlichen.“</p>

Niedersachsen	<p>Das <b>Niedersächsische Hochschulgesetz</b> (NHG) vom 21.1.94 legt in § 10 <b>Zentrale Aufgaben</b> fest:</p> <p>„(1) Die Hochschule legt in angemessenen Zeitabständen einen Bericht über die Situation und Entwicklung im Bereich von Lehre und Studium vor. Hierbei unterrichtet sie insbesondere:</p> <ol style="list-style-type: none"> <li>1. über die Entwicklung der Zahl der Studierenden und der Hochschulabschlüsse sowie</li> <li>2. über die Studienzeiten und Studienbedingungen in den einzelnen Fächern, und zwar auch im Hinblick auf die Entwicklung an anderen Hochschulen.</li> </ol> <p>(2) Die Hochschule überprüft die Lehrangebote und die Studienzeiten in den einzelnen Fächern. Hierzu führt sie insbesondere Evaluationen des Lehr- und Studienbetriebs durch und berät die wissenschaftlichen Einrichtungen ... “</p>
Nordrhein-Westfalen	<p>Das <b>Gesetz über die Universitäten des Landes Nordrhein-Westfalen</b> (UG) vom 3.8.93 legt in § 27 <b>Dekanin oder Dekan</b>, Absatz 1 fest:</p> <p>„ Die Dekanin oder der Dekan leitet den Fachbereich und vertritt ihn innerhalb der Hochschule. Sie oder er ist insbesondere verantwortlich für die Vollständigkeit des Lehrangebotes, für die Studien- und Prüfungsorganisation sowie die Erstellung des alle zwei Jahre vorzulegenden Lehrberichts.“</p>
Rheinland-Pfalz	<p>Das <b>Landesgesetz über die Universitäten in Rheinland-Pfalz</b> (UG) vom 23.5.95 legt in § 20 <b>Lehrangebot</b>, Absatz 3 fest:</p> <p>„Die Hochschule darf für ihre Aufgaben in der Lehre die Studierenden anonym über die Art und Weise der Vermittlung von Lehrinhalten in den Lehrveranstaltungen befragen und die gewonnenen Daten verarbeiten. Die Ergebnisse dürfen, soweit sie Namen von Lehrenden enthalten, nur hochschulöffentlich mitgeteilt werden.“</p> <p>In § 80 <b>Aufgaben</b> wird in Absatz 2 festgelegt:</p> <p>„Der Fachbereich hat insbesondere ... 2. das erforderliche Lehrangebot zu gewährleisten (§20) und jährlich dem Präsidenten einen Lehrbericht vorzulegen ...“</p> <p>In § 74 <b>Aufgaben des Präsidenten</b> wird in Absatz 1 genannt:</p> <p>„Der Präsident leitet die Hochschule und vertritt sie nach außen, sorgt für ein gedeihliches Zusammenwirken der Organe und der Mitglieder der Hochschule und unterrichtet die Öffentlichkeit von der Erfüllung der Aufgaben der Hochschule durch die Veröffentlichung des Jahresberichts.“</p>

Saarland	<p>Das <b>Gesetz über die Universität des Saarlandes</b> (UG) vom 1.6.94 legt in § 38 <b>Fachbereichsvorsitzender</b>, Absatz 3 fest:</p> <p>„(3) Der Fachbereichsvorsitzende legt dem Universitätspräsidenten jährlich einen Bericht über die Situation der Lehre vor (Lehrbericht); die Fachschaft nimmt zu dem Bericht Stellung.</p> <p>Auf der Grundlage der Lehrberichte der Fachbereichsvorsitzenden legt der Universitätspräsident dem Ministerium für Wissenschaft und Kultur jährlich einen zusammenfassenden Lehrbericht vor, der von der Zentralen Studienkommission vorzubereiten ist.</p> <p>Das Ministerium für Wissenschaft und Kultur kann durch Rechtsverordnung den Mindestinhalt der Lehrberichte bestimmen.“</p>
Sachsen	<p>Das <b>Gesetz über die Hochschulen im Freistaat Sachsen</b> (SHG) vom 4.8.93 legt in §14 <b>Lehrberichte</b> fest: „Die Hochschulen legen jährlich Lehrberichte vor. Darin berichten sie insbesondere</p> <ol style="list-style-type: none"> <li>1. für jeden Studiengang über die Zahl der Studienbewerber, Studienanfänger und Absolventen sowie über die Zahl der erfolgreichen und nicht erfolgreichen Abschluß- und Zwischenprüfungen,</li> <li>2. über die von der Hochschule und ihren Fakultäten oder Fachbereichen getroffenen Maßnahmen zur inhaltlichen und didaktischen Qualität der Lehre, zur Betreuung der Studenten und zur Einhaltung der Regelstudienzeit,</li> <li>3. über Befragungen der Studenten zur Qualität der Lehre und die Stellungnahme des Lehrkörpers zu den Ergebnissen der Befragung,</li> <li>4. über die Ergebnisse von Evaluationen.</li> </ol> <p>Näheres kann durch eine Rechtsverordnung des Staatsministeriums für Wissenschaft und Kunst geregelt werden.“</p>
Sachsen-Anhalt	<p>Das <b>Hochschulgesetz des Landes Sachsen-Anhalt</b> von 7.10.93 legt in § 8 <b>Lehrberichte</b> fest: „Die Hochschulen legen alle zwei Jahre Lehrberichte vor. Darin berichten sie insbesondere:</p> <ol style="list-style-type: none"> <li>1. für jeden Studiengang über die Zahl der Studienbewerber, Studienanfänger und Absolventen sowie über die Zahl der erfolgreichen und nicht erfolgreichen Abschluß- und Zwischenprüfungen,</li> <li>2. über die von der Hochschule und ihren Fachbereichen getroffenen Maßnahmen zur inhaltlichen und didaktischen Qualität der Lehre, zur Betreuung der Studenten und zur Einhaltung der Regelstudienzeit,</li> <li>3. über Befragungen der Studenten zur Qualität der Lehre und die Stellungnahme des Lehrkörpers zu den Ergebnissen der Befragung,</li> <li>4. über die Ergebnisse von Evaluationen.</li> </ol> <p>Näheres kann durch eine Verordnung des Ministeriums geregelt werden.“</p>

Schleswig-Holstein	<p>Das <b>Gesetz über die Hochschulen im Lande Schleswig-Holstein (HSG)</b> vom 27.4.1995 legt in § 56 <b>Dekanat</b>, Absatz 1 fest: „...Das Dekanat führt auf Antrag der Vertreter und Vertreterinnen der Mitgliedergruppe der Studierenden im Fachbereichskonvent mindestens einmal im Semester in Angelegenheiten der Lehre, des Studiums oder der Prüfungen eine hochschulöffentliche Anhörung durch; der Antrag bedarf der Mehrheit von drei Viertel der Mitglieder der Gruppe der Studierenden.“</p> <p>In § 81 <b>Studienreform</b>, Absatz 8 und 9 steht:</p> <p>„(8) Der Fachbereich erstellt für jeweils zwei Studienjahre einen Lehrbericht. Er soll für jeden Studiengang die für eine Beurteilung von Lehre, Studium, Hochschulprüfungen und staatliche Prüfungen wesentlichen strukturellen und quantitativen Daten und Aussagen über die zur Bewertung und Verbesserung der Lehre ergriffenen Maßnahmen enthalten. Der Lehrbericht enthält insbesondere Angaben zur Organisation der Lehre und der Prüfungen, zur Prüfungsdauer, zu Engpässen im Lehr- und Prüfungsbetrieb, zu Maßnahmen der Teilnahmebeschränkung nach §4 Abs.3 sowie jeweils getrennt nach dem Geschlecht zum Studienerfolg, zur Studiendauer, zum Studienabbruch und zum Fachwechsel. Ist eine im Studienplan vorgesehene Lehrveranstaltung nicht durchgeführt worden, nennt der Bericht hierfür die Gründe. Die Daten sollen überregionale Vergleiche ermöglichen. Die Hochschule erstellt auf der Grundlage der Lehrberichte der Fachbereiche einen Lehrbericht der Hochschule, der die nach Satz 2 bis 5 erforderlichen Angaben enthält, und legt ihn innerhalb eines Jahres nach Ablauf der beiden Studienjahre dem Minister oder der Ministerin für Wissenschaft, Forschung und Kultur vor. Die Vertreter und Vertreterinnen der Mitgliedergruppe der Studierenden im Fachbereichskonvent können sich zum Lehrbericht des Fachbereichs, die Vertreter und Vertreterinnen der Mitgliedergruppe der Studierenden im Senat können sich zum Lehrbericht der Hochschule äußern; sie können verlangen, daß ihre Äußerungen dem Lehrbericht beigefügt werden.</p> <p>(9) Der Fachbereich darf zur Wahrnehmung seiner Aufgaben in der Lehre die Studenten und Studentinnen über den Ablauf von Lehrveranstaltungen sowie über die Art und Weise der Darbietung des Lehrstoffs befragen und die Antworten auswerten. Die Studenten und Studentinnen sind zur Antwort nicht verpflichtet. Der Senat regelt die Erhebung und Weiterverarbeitung der personenbezogenen Daten dieser Befragung durch Satzung.“</p>
Thüringen	Keine Festlegung über Lehrberichte.

## Fazit: uneinheitliche Bestimmungen für Lehrberichte

- ⇒ **Studentische Befragungen** sind in keinem Bundesland verpflichtend, jedoch in Baden-Württemberg, Schleswig-Holstein, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt und Hessen als eine Möglichkeit der Datensammlung genannt.
- ⇒ Eine **Stellungnahme der Studierenden zu den Lehrberichten** ist ausdrücklich in Baden-Württemberg, Saarland und Schleswig-Holstein erwähnt, wobei Schleswig-Holstein auch eine Anhörung der Studierenden vorsieht.
- ⇒ Die Bundesländer haben keine oder unterschiedliche Forderungen nach bestimmten Zahlenmaterial festgeschrieben. Langfristig erscheint es daher notwendig, gemeinsam zu prüfen, welche Zahlen benötigt werden, und sich auf deren Erhebung zu einigen (z.B. über die KMK).
- ⇒ In den Gesetzen, die bestimmte Daten im Lehrbericht fordern, wird häufig auch nach der Anzahl der **Studienabbrecher und -abbrecherinnen** gefragt. Interessant dabei ist, daß dies bisher weder irgendwo erhoben wurde, noch aus den bisher bestehenden Daten errechnet werden kann. Die Hochschulen sind somit aufgefordert, bei der Exmatrikulation die Gründe derselben zu erfassen. HIS hat seit 1993 einige Projekte dazu durchgeführt, an denen sich die Hochschulen orientieren können (HIS, 1994, 1995).
- ⇒ Weiter erscheint die Verwendung mancher der **geforderten Daten** noch nicht eindeutig und damit problematisch, z.B. das Verhältnis zwischen Studienanfängern und Absolventen, um den Schwund zu beschreiben, der sich während eines Studienverlaufs in einem Fach ergeben hat. Da der bloße Anteil „Absolventen an Studienanfängern“ wenig besagt, wenn nicht geklärt ist, aus welchen Gründen die anderen Studierenden ihr Studium an der entsprechenden Hochschule nicht beendet haben, wäre eine zutreffende Erhebung von Fachwechslern, Studienortwechslern und Abbrechern verbunden mit der Erhebung maßgeblicher Gründe wichtig.
- ⇒ Die größte Schwierigkeit wird jedoch in der **Vergleichbarkeit der Daten** bestehen. Lehrberichte müssen je nach Land gar nicht, jährlich, alle zwei oder drei Jahre erstellt werden. Die ersten Lehrberichte in Deutschland werden zu ganz unterschiedlichen Zeiten (Abstand von einigen Jahren) fällig.
- ⇒ Einige Orientierungsmöglichkeiten für die Erstellung von Lehrberichten liegen bereits vor. Die Hochschulrektorenkonferenz hat ein „Muster zur Gliederung der Lehrberichte“ verabschiedet. Das Bundesministerium für Bildung und Wissenschaft hat einen Reader zum Thema „Lehrberichte“ herausgegeben (Webler et al., 1993).

In den nächsten Jahren wird man sich auch außerhalb der Gesetzgebung zumindest auf einen bestimmten **Mindeststandard der Lehrberichte** einigen müssen, falls eine Vergleichbarkeit der erhobenen Daten erreicht werden soll.

#### 4 Zur Validität studentischer Veranstaltungskritik

An Bewertungsinstrumente wird die berechtigte Forderung gestellt, daß sie objektiv, valide und reliabel zu sein haben. **Objektivität** gilt in den Sozialwissenschaften als gegeben, wenn Erfassung und Auswertung der Daten standardisiert sind. Sicherlich handelt es sich bei den Stellungnahmen von Studierenden zur Lehre oder zu Dozenten um persönlich-subjektive Aussagen und Wertungen. Diese lassen sich aber durchaus mit "objektiven" Instrumentarien erfassen. Studentischen Aussagen entsprechen in dem Maße der Forderung nach "Objektivität", in dem sie standardisiert erhoben und ausgewertet wurden.

Als **Validität** wird die Gültigkeit von Instrumenten und Skalen in bezug auf einen bestimmten Sachverhalt bezeichnet. Mit anderen Worten: Die Frage nach der Validität bezieht sich darauf, ob Instrumente und deren Teile tatsächlich das messen, was sie zu messen vorgeben. Im Fall der studentischen Veranstaltungskritik zielt die Messung auf „die Lehrleistung des Dozenten“ bzw. „die Lehrqualität einer Lehrveranstaltung“ aus studentischer Sicht. Dabei kann es aber zuweilen strittig sein, was Elemente der Lehrleistung sind und mit welchem Gewicht sie in einer Bilanz zu berücksichtigen wären.

**Reliabilität** meint die Meßgenauigkeit und Zuverlässigkeit der verwendeten Instrumente. Die Diskussion um Reliabilität wird bei der Veranstaltungskritik von Studierenden eher am Rande geführt, da viele Kritiker und Kritikerinnen von einer nicht gegebenen Validität ausgehen. Deshalb erübrige es sich, der Frage nachzugehen, ob das "Falsche" zudem noch ungenau gemessen wird. Läßt sich aber belegen, daß die Lehrleistung von Dozenten und Dozentinnen und die Qualität ihrer Veranstaltungen sich durchaus valide durch Studierende bestimmen läßt, dann wird die Frage nach der Meßgenauigkeit wichtig. Dies umso mehr, wenn mit der Erfassung Vergleiche verschiedener Lehrender oder Veranstaltungen beabsichtigt werden.

Reliabilität und Validität hängen zusammen. Denn erst, wenn ein Instrument den Sachverhalt überhaupt zutreffend erfaßt (d.h. valide ist), kann es genaue und zuverlässige Resultate liefern (d.h. reliabel sein). Und erst eine hohe Reliabilität ermöglicht eine gute Validität, d.h. nur ein genaues Instrument kann das gewünschte Kriterium zutreffend vermessen. Die große Auseinandersetzung findet jedoch weniger um die Reliabilität der Instrumente, um deren Meßgenauigkeit, statt, sondern ist weit umfassender. Es wird vor allem bezweifelt, ob es überhaupt möglich ist, die Lehrqualität und Lehrleistung mit Fragebogen der studentischen Veranstaltungskritik zu erfassen.

Die Kritik an der Validität studentischer Aussagen über die Qualität von Veranstaltungen und Lehrenden umfaßt eine Reihe von Argumenten und eine Vielzahl von angeführten Aspekten. Bei der Darstellung der empirischen Studien da-

zu ist es unvermeidlich, methodische Fachtermini zu verwenden, trotz aller Bemühungen, sie auf ein Mindestmaß zu beschränken. Zitate aus den amerikanischen Studien wurden der Lesbarkeit halber ins Deutsche übersetzt.

Im folgenden werden Studien aufgeführt, die systematisch die Zuverlässigkeit und Genauigkeit studentischer Bewertungen mit vorher festgelegten und akzeptierten Kriterien vergleichen. Die Frage nach der Gültigkeit studentischer Beurteilungen ist von besonderer Wichtigkeit, da "mangelnde Sachkenntnis" eines der häufigsten Gegenargumente auf die Forderung nach studentischer Veranstaltungskritik ist. Studierende seien nicht in der Lage, die Lehre zu bewerten, weil sie keine Beurteilungsgrundlagen für die Unterscheidung zwischen guter und schlechter Lehre hätten. Sie könnten frühestens nach Beendigung ihres Studiums einschätzen, was sie gelernt haben und welchen Stellenwert dies hat. Daraus wird im Gegenzug abgeleitet, daß nur Kollegen/ Kolleginnen die Lehre sachgerecht beurteilen können.

Dieses Argument läßt vermuten, daß zwischen der Selbstbewertung der Lehrenden oder von Kollegen und Kolleginnen und der studentischen Veranstaltungskritik große Differenzen bestehen müßten. Dies wäre auch darauf zurückzuführen, daß zur Bewertung von Lehrveranstaltungen die Lehrenden und Studierenden sich auf jeweils ganz andersartige Kriterien stützen.

Gegenüber dem Vorbehalt einer mangelnden Sachkenntnis der Studierenden und damit der ungenügenden Validität ihrer Urteile ist die Frage zu richten, wie ihr Vorhandensein oder Fehlen empirisch überprüft werden kann. Woran und wie läßt sich die Validität studentischer Veranstaltungskritiken klären?

Cohen (1981) nennt einige Kriterien, an denen studentische Bewertungen validiert, d.h. ihre Gültigkeit überprüft werden kann: "... (1) Bewertungen von Kollegen, (2) Bewertungen von Seiten der Verwaltung, (3) Selbstbewertungen, (4) Bewertung ehemaliger Studierender und (5) studentische Leistung" (Cohen, 1981, S. 283). Bis auf Punkt 2 (Vergleich zu Lehrbewertungen der Verwaltung) können zu allen anderen der von Cohen angeführten und einigen weiteren Kriterien empirische Untersuchungen herangezogen werden.

#### **4.1 Beurteilung der Lehre: Kollegen und Studierende im Vergleich**

##### **Wie sich Lehrende eine Meinung zur Lehrqualität bilden**

Mit dem Einwand der mangelnden Sachkenntnis der Studierenden wird zumeist, direkt oder indirekt, vorausgesetzt, daß nur Experten gleichen Ranges eine hinreichend qualifizierte Lehrbewertung vornehmen könnten. Es ist daher auf-

schlußreich, wie Lehrkollegen zu gegenseitigen Urteilen kommen und worauf sie diese gründen.

Eine Befragung von Aleamoni (1987) belegt, daß zwischen der Lehrbewertung von Kollegen/Kolleginnen und wissenschaftlicher Produktivität eine nur sehr geringe Korrelation (Übereinstimmung) von  $r=.07$  besteht, wohingegen die Korrelation zwischen der Lehrqualitätseinschätzung von Lehrenden und Studierenden mit  $.70$  hoch ausfällt.

In Gesprächen ermittelte der Autor, daß Lehrende ihre Informationen über die Lehrfähigkeit ihrer Kollegen und Kolleginnen nahezu ausschließlich von Studierenden erhielten. Dieser Punkt ist beachtenswert, da Aleamoni somit herausstellen konnte, daß Lehrende ihre Kollegen und Kolleginnen in der Lehre nicht nach dem Kriterium "Ausmaß wissenschaftlicher Produktivität" bewerten.

Ziel der Untersuchung von McGaghie (1975) ist es, die Beziehung zwischen der Lehrbewertung von Studierenden und Lehrenden zu klären. Lehrende nennen folgende Bewertungsdimensionen: Forschungsaktivitäten, Teilnahme an der "academic community", persönliche und wissenschaftliche Anerkennung, intellektuelle Größe, Beziehung zu Studierenden und Interesse an der Lehre. Es gäbe auch Hinweise, daß der akademische Rang die Bewertung von Kollegen und Kolleginnen beeinflusst.

Studierende würden dagegen die Lehre nach ganz anderen Kriterien bewerten: Kursorganisation, Präsentation, "Analytisch-synthetische Vorgehensweise" und Beziehung Kurs - Dozent/ Dozentin. Trotz derart unterschiedlicher Beurteilungskriterien waren sich die befragten Gruppen über den besten bzw. schlechtesten Lehrenden erstaunlich einig.

Die beiden Untersuchungen von McGaghie (1975) und Aleamoni (1987) beleuchten einen wenig untersuchten, dennoch außerordentlich wichtigen Aspekt der Lehrbewertung durch Kollegen oder Kolleginnen. Dabei kommt es sehr häufig zu einer Bewertung der Lehre, ohne daß dem eine "peer-evaluation" zugrundeliegt, d.h. ohne daß Kollegen und Kolleginnen eine oder mehrere Veranstaltungen besucht oder sich zumindest die Kursplanungen angesehen haben; sie schließen vielmehr aus verschiedenen anderen Indikatoren, die meist nicht definiert oder als Kriterien festgelegt sind, auf die Qualität der Lehre.

Dieser Aspekt der gegenseitigen Lehrbewertung durch Kollegen spielt in der wissenschaftlichen Forschung kaum eine Rolle. Das ist zu bedauern, da dieser Punkt in der hochschulpolitischen Diskussion häufig thematisiert wird. Es ist sogar zu vermuten, daß je weniger die Evaluation der Lehre in den Hochschulen etabliert und bereits Kriterien und Instrumente gefunden und akzeptiert sind, desto eher die Lehre auf eine Art und Weise in den Fachkollegien beurteilt wird, die kaum

kontrollierbar und oft von weit weniger "Sachkenntnis" bestimmt ist, wie sie von den Studierenden verlangt wird.<sup>1</sup>

### **Kriterien der Lehrbewertung von Lehrenden und Studierenden**

Es liegen einige Studien vor, die zu ermitteln versuchen, ob Studierende und Lehrende gleiche oder verschiedene Kriterien zur Lehrbewertung benutzen bzw. diese gleich gewichten. Fraglich bleibt jedoch, welchen Einfluß es auf die Veranstaltungskritik selbst hat, wenn die Befragten unterschiedliche Kriterien der Bewertung zugrunde legen.

Mit ihrer Studie wollten Marquez, David und Dorfman (1979) mögliche Differenzen in der Gewichtung von Lehraspekten durch Lehrende und Studierende auffinden. Die Befragten, 40 Studierende und 40 Lehrende, wurden um die Bewertung der Lehreffektivität von Lehrenden gebeten, die ihnen in 100 Profilen vorgestellt worden waren. Nach Studierenden und Lehrenden unterschieden, ergaben sich nur geringe, aber signifikante Unterschiede. Lehrende betonten dabei geringfügig stärker Bereiche, die der übergeordneten Dimension "Inhalt" zugeordnet werden können, während den Studierenden die Bereiche der übergeordneten Dimension "Stil" etwas wichtiger sind. Diese Unterschiede sind jedoch marginal. Dies kann z.T. an der großen Varianz zwischen den einzelnen Bewertern liegen. Eine Generalisierung der Ergebnisse ist bei dieser Studie aus verschiedenen Gründen schwer möglich.

In einer ähnlichen Studie von Baum und Brown (1980) hatten Studierende in sieben Kursen und 50 Lehrende die Gelegenheit, 100 Punkte über zehn Statements zu verteilen, die allgemeine Aspekte der Lehre beinhalteten. Die Verteilung der Punkte zwischen Studierenden und Lehrenden unterschied sich im gesamten, im einzelnen war sie für fünf Statements ebenfalls signifikant unterschiedlich. Die Autoren schließen daraus, daß Lehrende und Studierende "fundamental unterschiedliche Kriterien" zur Lehrbewertung heranziehen würden.

Es wurde in dieser Untersuchung aber nicht, wie in den folgenden Studien, die Übereinstimmung zwischen Studierenden und Lehrenden bei der Bewertung mit einem bestimmten Instrument ermittelt, sondern nur die Übereinstimmung in Bezug auf theoretische Lehraspekte und deren Wichtigkeit. Auf diese Weise lassen sich auch die Diskrepanzen zwischen dieser Studie und Untersuchungen mit relativ guter Übereinstimmung zwischen Studierenden und Lehrenden erklären.

---

<sup>1</sup> Ein Beispiel für diese problematische Handhabung ist der Artikel von Jürgen Kriz "Wie gut sind unsere Universitäten?". (In: Stern, 1993, Bd. 16, S. 171-184), dessen Datenbasis rein auf die subjektive Einschätzung von Professorinnen und Professoren aus den entsprechenden Fachbereichen anderer Hochschulen beruht.

## **Übereinstimmung von Lehrenden und Studierenden bei der Lehrevaluation**

Eine Studie, die eine gute Übereinstimmung zwischen Lehrenden und Studierenden bei der Einschätzung der Wichtigkeit von Items eines Instruments nachweist, ist die von Shatz und Best (1986), in der Items identifiziert und verglichen werden, die zur Kursevaluation genutzt werden. Als Grundlage diente der "Students' Evaluations of Educational Quality" (SEEQ). 45 Lehrende und 106 Studierende wurden aufgefordert, wichtige Items des SEEQ zu benennen und daraus die wichtigsten zu identifizieren.

Beide Gruppen identifizierten durchschnittlich 22 Items als wichtig. Die benannten Items korrelierten mit .59, die wichtigsten sogar mit .75. Die Autoren kommen zum Schluß: "Obwohl Lehrende und Studierende die Kurse mit unterschiedlichen Verantwortlichkeiten und Perspektiven betreten, scheinen sie in der Frage übereinzustimmen, welche Aspekte der Kurse im Evaluierungsprozeß am wichtigsten sind" (Schatz und Best, 1986, S. 241).

Paula L. Stillman et al. (1983) ließen 85 Lehrende eines Jahres und 67 des nächsten Jahres mit einem 7 Item-Fragebogen und zwei offenen Fragen bewerten. Davon unterrichteten 41 Lehrende den gleichen Kurs zum zweiten Mal. Außerdem wurde eine "peer evaluation" durchgeführt, bei der ein Lehrender/ eine Lehrende die Kurse besuchte und den gleichen Fragebogen beantwortete. Die Korrelationen zwischen studentischer Veranstaltungskritik und der "peer evaluation" rangierte zwischen .37 und .67 und war signifikant. Für die Lehrenden korrelierten die Ergebnisse zwischen .60-.90.

In der Studie von Drews, Burroughs und Nokvich (1987) bewerteten Studierende und Lehrende in vier sehr unterschiedlichen Kursen 30 Prozent der Kursstunden mit einem kurzen 15-Item-Fragebogen jeweils am Ende einer Stunde. Die Korrelationen der Itembeantwortungen der Lehrenden und der Studierenden waren bei insgesamt elf Items signifikant. Items zum Kursmaterial wurden sehr ähnlich eingeschätzt, gefolgt von Dozent-/Dozentinbewertung und Gesamtbewertung. Am eklatantesten unterschied sich die Einschätzung, ob die Unterrichtsstunde spannend oder einschläfernd war.

### **4.2 Vergleich zwischen Studierenden und anderen Beobachtern/Beobachterinnen**

Die Beurteilung von Lehrveranstaltungen durch Kollegen und Kolleginnen unterscheidet sich von der durch andere Beobachter und Beobachterinnen vor allem in zwei Punkten: Zum einen müssen die Beobachtenden nicht unbedingt Kollegen und Kolleginnen sein (vorstellbar wären z.B. Mitglieder eines Hochschuldidaktischen Zentrums). Zum anderen hat dieses Verfahren den Vorteil, daß Werte von

Beobachtern/Beobachterinnen mit Hilfe von präzise festgelegten Kriterien und Beobachterschulung leichter zu validieren sind.

McKeachie und Lin (1978) führten die einzige dazu vorliegende Studie durch. 20 Lehrende mit Kursen von durchschnittlich 25 Studierenden wurden sechs mal in vierzehn Semesterwochen von trainierten Beobachtern in Beobachtungskategorien eingeteilt. Zwei Kategorien wurden für die Studie von McKeachie und Lin (1978) erneut verwendet: Wärme und Akzeptanz. Studierende füllten eine Lehrevaluation "Student Perception of Teaching and Learning" aus.

Die drei Items des Faktors „Rapport“ (erlaubend, freundlich, fordert zu Kritik auf) wurde mit der Beobachtereinschätzung von „Wärme und Akzeptanz“ korreliert. Eine signifikante Korrelation ergab sich zwischen dem Fragebogenitem „freundlich“ und der beobachteten „Akzeptanz“. Auch korrelierte der Gesamtfaktor „Rapport“ signifikant mit der „Akzeptanz“. Die Autoren kommen zu dem Schluß: „Diese Studie leistet eine gewisse empirische Unterstützung der Annahme, daß studentische Bewertung auf dem Verhalten der Lehrenden beruht“ (McKeachie und Lin, 1978, S. 46 ).

Eine solche Schlußfolgerung erscheint jedoch allein aufgrund dieser Studie zu weitreichend. Die Kategorien der geschulten Beobachter und Beobachterinnen ließen sich alle auf einer "Beziehungsebene" anordnen, Kategorien einer didaktisch-sachlichen Ebene fehlten völlig. Dennoch läßt sich nach dieser Untersuchung feststellen, daß die studentische Bewertung, wenn sie mit relativ konkreten Items ermittelt wird (wie dies in der vorliegenden Studie der Fall war), zumindest hinsichtlich der "Beziehungsdimension" das Verhalten von Lehrenden zutreffend wiedergibt.

### **4.3 Selbstbeurteilungen von Lehrenden und studentische Veranstaltungskritik**

Die Übereinstimmung zwischen der Selbstbeurteilung von Lehrenden und der Kursbewertung durch Studierende ist relativ leicht zu überprüfen. Diese Übereinstimmung hat eine größere praktische als wissenschaftliche Relevanz, da es zwar eine gute Methode ist, Lehrende davon zu überzeugen, daß studentische Veranstaltungskritik nützliche Ergebnisse liefert, jedoch die Frage nicht geklärt ist, inwieweit Selbstbeurteilungen ihrerseits valide sind (z.B. in Zusammenhang mit dem Kriterium Lernzuwachs).

Eine der wohl größten Validitätsuntersuchungen in diesem Bereich ist von Marsh (1982b) durchgeführt worden. Er untersuchte die konvergente und diskriminante Validität zwischen den Selbstbeurteilungen und den Lehrbewertungen. 181 Lehrende mit 329 Kursen füllten Bögen zur Lehrevaluation aus, wobei ihnen explizit

mitgeteilt wurde, daß ihre Bewertungen sich durchaus von denen der Studierenden unterscheiden könnten. Der gleiche Fragebogen wurde auch von den Studierenden bearbeitet. Die Ergebnisse der Korrelationsuntersuchungen ergaben gute Übereinstimmungen.

Nach dieser umfänglichen Studie läßt sich folgern: Erstens unterscheiden Lehrende und Studierende in nahezu übereinstimmender Weise die verschiedenen Elemente der Lehrqualität (Dimensionen oder Faktoren). Zweitens weisen ihre Urteile eine hohe Übereinstimmung auf.

#### **4.4 Qualität der Lehre aus der Sicht junger Berufstätiger und Ehemaliger**

Als Argument gegen die Bewertung der Lehre durch Studierende wird immer wieder deren fehlende Distanz zum Untersuchungsgegenstand genannt. Deshalb wird des weiteren behauptet, daß erst im Laufe ihrer späteren Berufstätigkeit Studierende einschätzen können, was sie gelernt haben.

Ziel der Untersuchung von Firth (1979) war es, gerade dieses Argument zu überprüfen. Dafür versuchte er, die Langzeitstabilität studentischer Evaluation zu ermitteln. Die Studie bestand aus insgesamt zwei Untersuchungsteilen. Zunächst wurden 216 Lehrende aus 83 Kursen mit einem Fragebogen von allen Studierenden des letzten Studienjahres der Wirtschafts-, Rechts- und Erziehungswissenschaften bewertet. Diese Fachbereiche wurden gewählt, da man dort den größten Zusammenhang zwischen dem Stoff und den späteren Arbeitsanforderungen vermutete. Ein Jahr später wurden die dann bereits Graduierten angeschrieben und um eine Einschätzung der Lehreffektivität gebeten, die sie in Anbetracht ihrer jetzigen Berufserfahrung geben sollten.

Festgestellt werden konnte, daß die Bewertung zwischen dem letzten Jahr an der Universität und dem Jahr danach nur für 13 der 216 Lehrenden signifikant unterschiedlich war.

Mit Hilfe des Fragebogens "Student Assessment of Instruction" (SAI) bewerteten 1.374 Studierende (Rücklaufquote 85%) bei Overall und Marsh (1980) über drei Jahre hinweg Betriebswirtschaftskurse. Ein Jahr nach Abschluß ihres Programms (somit 1-3 Jahre nach den besuchten Kursen) beantworteten noch einmal 65 Prozent der angeschriebenen Ehemaligen einen Fragebogen zur Lehrbewertung.

Ergebnisse: Die relative Übereinstimmung (d.h. die Übereinstimmung für jedes Item bei den 1.374 Studierenden), betrug  $r=.83$  und die absolute Übereinstimmung (d.h. die Übereinstimmung des Durchschnittswerts bei jedem Item für die beiden Bewertungen) betrug 58 Prozent.

Daher kommen die Autoren zu der Auffassung: "Die Ergebnisse dieser Studie beinhalten eine große Unterstützung für die vermutete Stabilität studentischer Bewertung ihrer Kurse und Lehrenden und lassen vermuten, daß diese Stabilität nicht systematisch mit Kursniveau oder -inhalt variiert" (Overall und Marsh 1980, S. 324).

Die Langzeitstabilität der Bewertungen von Studierenden und Ehemaligen (jungen Berufstätigen) und ihre Übereinstimmung konnte in beiden aufgeführten Studien nachgewiesen werden. Ob dies daran liegt, daß junge Berufstätige die verschiedenen Lehraspekte im gleichen Licht sehen, oder daran, daß sie sich an ihre früheren Bewertungen erinnern und sie nur wiederholen, ist dabei unerheblich, da die angeführte Kritik hinsichtlich der mangelnden Sachkenntnis der Studierenden davon nicht berührt wird.

#### **4.5 Lernzuwachs als Maß der Lehrqualität und -effektivität**

Die Validierung studentischer Lehrbewertungen am Ausmaß des gelernten Wissens gilt als eines der besten, wenn nicht als das beste Kriterium, insofern sich die Lehrleistung der Dozierenden in einem möglichst hohen Lernzuwachs erweisen sollte.

Es erscheint nachvollziehbar, daß Lehrende, die Wissen, Strukturen, Zusammenhänge etc. zu vermitteln haben, genau daran gemessen werden sollten, inwieweit sie dieser Aufgabe gerecht werden. Aufgrund der Unterschiede in der Kursorganisation zwischen den Vereinigten Staaten und Deutschland<sup>2</sup> stellt sich jedoch die Frage, inwieweit Lehrende hier überhaupt dafür verantwortlich gemacht werden können, was ihre Studenten und Studentinnen gelernt haben. Cohen wendet gegen diese Art der Validierung ein: "Dies ist höchstens ein grober Index der Lehreffektivität, da eine Anzahl von Faktoren außerhalb der Kontrolle des Lehrenden - studentische Fähigkeit und Motivation z.B. - den Lernzuwachs beeinflussen" (Cohen, 1981, S. 281). Er trifft damit sicherlich einen Teil der Problematik. Dieser könnte jedoch anhand von Vergleichen zwischen randomisierten Kursen kontrolliert werden. Die Fragestellung einer solchen Untersuchung müßte lauten: Haben die Studierenden eines Kurses beim Dozenten A weniger gelernt als im vergleichbaren Kurs von Dozentin B?

Die größere Schwierigkeit besteht in der Operationalisierung des Lernzuwachses, der auf einen bestimmten Kurs oder Lehrenden zurückzuführen ist, und in der Vergleichbarkeit und Wertigkeit dieses Lernzuwachses. Hat zum Beispiel Dozent B, der seine Studierenden dazu gebracht hat, eine Unmenge von Fakten zu lernen,

---

<sup>2</sup> In den Vereinigten Staaten sind Kurse häufig kleiner, stärker strukturiert und arbeitsaufwendiger als in Deutschland.

mehr erreicht als Dozentin A, die den Studierenden vor allem Strukturzusammenhänge vermittelt hat, oder Dozent C, der in seiner Vorlesung das Interesse in sein Fachgebiet nachhaltig wecken konnte?

Cohen (1981) führte eine Meta-Analyse von 41 Studien durch. Eine der wichtigsten Voraussetzungen für die Auswahl der Studien war, daß die Daten Vergleiche zwischen Klassen enthielten. Folgende Ergebnisse konnten vom Autor als Einflußfaktoren auf die Kursleistung (meist Klausurergebnisse oder Note) ermittelt werden: Große Effekte hatten Geschicklichkeit in Lehre und Kursstruktur, mittlere Effekte wurden bei der Beziehung Lehrende/Studierende, Rückmeldung und Bewertung gefunden. Die Kursschwierigkeit hatte keine Bedeutung. Insgesamt stellte Cohen fest: "Wir können relativ sicher sein, daß Dimensionen des Gesamtkurses und des Lehrenden ziemlich eng mit der Leistung des Studierenden zusammenhängen" (Cohen, 1981, S. 298).

Untersucht wurde dann, inwieweit ein Zusammenhang zwischen Charakteristika der einzelnen Studien und den gefundenen Effekten in 67 Kursen besteht. Drei Variablen korrelierten signifikant mit der Effektgröße: Kontrolle möglicher Einflußfaktoren auf die Kursbewertung, Zeitpunkt der Kursbewertung (vor oder nach Bekanntgabe der Noten) und die Erfahrung des/der Lehrenden. Diese drei Variablen machten gemeinsam 31 Prozent der Varianz in der Korrelation zwischen studentischer Veranstaltungskritik (verwandt wurde nur die Gesamtbewertung der Lehrenden) und der Leistung aus.

Die Korrelation zwischen Lehrbewertung und Noten betrug .85, wenn die Noten bekannt, und .38, wenn die Noten nicht bekannt waren. Dies kann als relativ eindeutiger Nachweis gelten, daß nicht nur vorheriges Interesse, sondern auch das direkte Wissen um die eigene Note die studentische Veranstaltungskritik beeinflusst. Der Autor zieht folgenden Schluß: "Die vorliegende Meta-Analyse beinhaltet eine große Unterstützung für die Validität studentischer Bewertung als Maß der Lehreffektivität" (ebd., S. 300).

Problematisch dabei erscheint nur, daß bei einer Definition des Lernzuwachses über Noten bzw. Ergebnisse von Abschlußklausuren es kaum möglich ist, zwischen Effekten aufgrund von Noten und Effekten aufgrund von Gelerntem zu trennen. Daher kann der Befund von Cohen auch interpretiert werden zugunsten des Vorwurfs, daß Lehrende, die weniger streng bewerten, selbst besser bewertet werden. Zwar analysierte der Autor nur Studien, die zwischen Kursen verglichen, dennoch kann gerade dabei ein derartiges "Bias" vermutet werden.

Um die Schwierigkeit der Schlußfolgerung des Lernzuwachses anhand von Noten zu umgehen und diesen präziser zu fassen, wurden von Prosser und Trigwell (1991) Daten von 122 Studierenden aus elf Kursen in "Kommunikation in der

Pflege" sowohl innerhalb des Kurses als auch zwischen den Kursen verglichen. Erhoben wurde:

- a) ihre "akademische" Leistungsfähigkeit, mit Hilfe des Higher School Certificate Score,
- b) quantitative Lernfortschritte,
- c) die Herangehensweise an das Lernen (tiefes Verständnis anstreben, Zusammenhänge erkennen, oberflächliches Verständnis aufbauen),
- d) zwei Gesamtbewertungen zu Kurs und Dozent/ Dozentin und
- e) qualitative Lernfortschritte (Aufsatz zum Kursziel), gemessen an fünf Kategorien von vorstrukturelles bis abstraktes Verständnis.

Bei der Auswahl der Datenklassen berufen sich die Autoren auf neuere Studien, die verstärkt die Qualität und weniger das Ausmaß des Gelernten in den Vordergrund stellen.

Ergebnisse dieser Untersuchung: Der Vergleich zwischen den Kursen erbrachte, daß die studentische Bewertung eines Kurses signifikant mit der Herangehensweise "tiefes Verständnis anstreben" und mit qualitativen Lernfortschritten korrelierte. Der Vergleich innerhalb der Kurse ergab positive Korrelationen zwischen der Gesamtkursbewertung und der Herangehensweise "tiefes Verständnis anstreben" und eine negative Korrelation zwischen der Gesamtbewertung des Kurses und der Note, beides signifikant. Die Autoren gehen deshalb davon aus, daß eine Validierung von Lehrbewertungen mit Hilfe von qualitativen Verfahren erfolgen müßte.

In der Arbeit von Katherine van Giffen (1990) korrelierte die Einschätzung von Lehreffektivität und die Selbsteinschätzung der Lernleistung durch die Studierenden positiv ( $r = .789$ ).

Alle drei Studien machen die Schwierigkeit der präzisen Definition des Kriteriums "Lernzuwachs" deutlich. Es gibt Anhaltspunkte, daß Studierende, die mehr gelernt haben, auch ihre Lehrenden besser bewerten. Wie Prosser und Trigwell (1991) jedoch zeigen konnten, kann aber die genauere Klärung dieser Tendenzen nur mit Hilfe aufwendiger Untersuchungen gelingen, die verschiedene Aspekte des "Lernzuwachses" beleuchten würden.

Wegen der grundsätzlichen Wichtigkeit des Zusammenhanges zwischen „Qualität der Lehrveranstaltungen“ und „Lernzuwachs der Studierenden“, auch über die Fragen der Lehrevaluation hinaus, wären weitere, differenzierte Studien in diesem Feld wünschenswert.

#### **4.6 Verbesserung der Lehrbewertung nach Veränderung kritischer Aspekte**

Eine besonders elegante Art der Validierung studentischer Veranstaltungskritiken stellt die Messung von Veränderungen in der Lehrbewertung dar, nachdem kritisch bewertete Aspekte der Lehre verbessert wurden. Auch zu diesem wichtigen Punkt, ob Änderungen in der Lehre sich auch im Urteil der Studierenden bemerkbar machen, liegt allerdings nur eine Studie vor.

Zwei Studienjahre eines vierwöchigen Medizinkurses wurden von Lynch, Tamburino und Nagel (1989) mit einem Fragebogen evaluiert. Drei Problembereiche wurden identifiziert und Verbesserungen durchgeführt.

Die Bewertungen der Kurse in den folgenden drei Semestern wurden erhoben. Signifikante Änderungen der Ergebnisse wurden für die modifizierten Bereiche ermittelt. Alle anderen Ergebnisse unterschieden sich nicht signifikant von den ursprünglichen. Da sich die Bewertung nur bei den modifizierten Bereichen änderte, kann dies als ein Zeichen externer Validität der studentischen Veranstaltungskritik betrachtet werden.

#### **4.7 Weiterempfehlung bzw. Besuch eines weiteren Kurses**

Eine weitere Möglichkeit, Validität zu erfassen - wenn als eines der Lehrziele gilt, "Interesse am entsprechenden Fach und Stoff zu wecken" - ist die Ermittlung, ob Studierende den Kurs weiterempfehlen und weitere Kurse in der entsprechenden Disziplin zu besuchen planen. (Letzteres erscheint allerdings etwas problematisch, wenn man das Ausmaß an Veranstaltungen bedenkt, die verpflichtend zu belegen sind.)

In der Untersuchung von Bruton und Crull (1982) stuften 1.314 Studierende ihren Kurs in eine der fünf Kategorien von weit unter- bis weit überdurchschnittlich ein; sie gaben zudem eine Gesamtbewertung für den Lehrenden ab. Mit Hilfe einer Pfadanalyse wurden Zusammenhänge analysiert. Obwohl der Kurs explizit unabhängig von den Lehrenden bewertet wurde, war der Dozent/die Dozentin doch der wichtigste Faktor der Kursbewertung.

Ein weiteres Ergebnis dieser Studie war, daß Studierende, die kleinere Veranstaltungen besucht haben, eher bereit sind, einen weiteren Kurs derselben Disziplin zu besuchen. Dies weist darauf hin, daß die "Kursgröße" u.U. mit der Gesamtbewertung "Dozent/Dozentin" interagiert. Um genauere Aussagen darüber zu ermöglichen, müßten Kursvergleiche so durchgeführt werden, daß der Varianzanteil bei einer Weiterempfehlung, der auf den/die Lehrende zurückzuführen ist, ermittelt werden kann.

## 4.8 Reliabilität studentischer Veranstaltungskritik

Eine Definition der Reliabilität für studentische Veranstaltungskritik nennen Overall und Marsh (1980): "Reliabilität, wie sie sich in den meisten früheren Studien darstellt, ist gefaßt als die relative Übereinstimmung zwischen den Beurteilungen verschiedener Studierender innerhalb derselben Klasse - unter der Annahme, daß jede Varianz des einzelnen Studierenden zufällig ist und als Fehlervarianz betrachtet werden kann" (Overall und Marsh, 1980, S. 324).

Bledsoe (1978) ermittelte eine Reliabilität von .93 für Klassen und .96 für Items in 42 Statistikkursen und damit einen hohen Grad an Übereinstimmung und Genauigkeit. Zu Reliabilitätskoeffizienten um .90 kommt auch Centra (1980).

Viele Instrumente, die in anderen Arbeiten untersucht wurden, kamen zu vergleichbaren Ergebnissen. Es ist zu vermuten, daß bei einer entsprechend sorgfältigen Erstellung solcher Instrumente die Reliabilität kein größeres Problem darstellen dürfte. Es erscheint somit möglich, Fragebogen zu gestalten, die das definierte Kriterium (Lehrqualität aus studentischer Sicht) in ausreichend präzisiertem Maße erfassen.

### Schlußfolgerungen

*Insgesamt kann festgestellt werden, daß zwar bereits einige empirische Untersuchungen existieren, die die Gültigkeit (Validität) studentischer Veranstaltungskritik an unterschiedlichen Kriterien überprüfen, die 'via regia' wurde jedoch noch nicht beschrritten. Dies kann jedoch nicht nur den Wissenschaftlern und Wissenschaftlerinnen angelastet werden, die diese Studien durchgeführt haben.*

*Um wirklich beurteilen zu können, ob Studierende eine gute Veranstaltung von einer schlechten unterscheiden können, wäre es notwendig vorab festzustellen, welche Ziele der Besuch entsprechender Kurse haben soll. Erst wenn klar abgegrenzte Kriterien definiert sind, kann ermittelt werden, ob sie von den Lehrenden erreicht werden konnten und ob die Studierenden sie richtig einschätzen können.*

*Die in den Studien verwendeten Kriterien stellen jedoch implizit Teilbereiche entsprechender Ziele dar und können in ihrer Gesamtheit eine gute Annäherung an 'optimale Kriterien' darstellen. Im einzelnen kann gesagt werden:*

*Es konnten keine Belege gefunden werden, die eine unzureichende Sachkenntnis der Studierenden vermuten lassen. Die Langzeitstabilität der Bewertungen ist beträchtlich. Studierende sehen die Lehrqualität der einzelnen Veranstaltungen nach Beendigung ihres Studiums kaum anders als während desselben.*

*Die unterschiedliche Einschätzung verschiedener Lehraspekte durch Studierende und Lehrende wirkt sich nicht auf die Beurteilung mit Hilfe entsprechender Fragebogen aus, wie die gute Übereinstimmung zwischen der Bewertung von Kollegen/ Kolleginnen und Studierenden zeigt. Eine Auswirkung "mangelnder Sachkenntnis" der Studierenden kann daher nicht angenommen werden.*

*Das Argument der "fehlenden Sachkenntnis" der Studierenden zur Beurteilung von Lehre setzt offensichtlich falsch an: Es unterstellt gleichsam, die Studierenden würden das "Wissen und die Kompetenz des Lehrenden" bewerten. Das trifft aber nicht zu, die Studierenden beurteilen vielmehr "die Vermittlung von Wissensbeständen" (nicht die Wissensbestände selbst)!*

*Daher ist gerade das Argument der mangelnden Sachkenntnis der Studierenden in doppeltem Sinne unzutreffend: Es trifft weder den Sachverhalt, um den es bei studentischer Veranstaltungskritik geht, noch finden sich in empirischen Studien stützende Belege.*

*Die Übereinstimmung zwischen der Selbstbeurteilung der Lehrenden bzw. von Beobachtungsergebnissen und dem studentischen Urteil kann als relativ hoch betrachtet werden.*

*Auch die Ergebnisse zur Verbesserung kritisierter Lehraspekte und der Weiterempfehlung von besuchten Kursen weisen darauf hin, daß Studierende durchaus in der Lage sind, die Lehrqualität von Veranstaltungen zu beurteilen.*

*Im übrigen ist auffällig, daß Untersuchungen, die relativ allgemeine Kriterien der Lehre und ihre Wichtigkeit prüfen, eher größere Diskrepanzen in der Bewertung zwischen Lehrenden und Studierenden ermitteln. Untersuchungen, die konkrete Sachverhalte an bestimmten Aspekten oder in direkten Bewertungen der Lehre ermitteln, kommen dagegen zu hohen Übereinstimmungen zwischen Lehrenden und Studierenden (vgl. auch Kapitel 6 und Kapitel 7).*

## 5 Einflüsse auf studentische Urteile zur Lehre

Die im folgenden dargestellten Ergebnisse zum Thema der Lehrevaluation durch Studierende sind Studien entnommen, die zum großen Teil aus den Vereinigten Staaten stammen. Daß gerade in den USA so viele Untersuchungen durchgeführt wurden, liegt an der langen Nutzungsdauer (bereits in den zwanziger Jahren wurde studentische Veranstaltungskritik eingeführt) und dem inzwischen weit verbreiteten Einsatz.

In Deutschland gibt es einige wenige Veröffentlichungen aus den siebziger Jahren (z.B. Keil/Piontkowski, 1973; Ipsen/Portele, 1976), die sich mit "Prozessen des Hochschulunterrichts" und der "Organisation von Lehre" befaßten und auf studentische Befragungen stützten, allerdings ohne als Arbeiten zur "Lehrevaluation" präsentiert und verstanden zu werden.

In den letzten Jahren wurde zwar manches zum Thema "Evaluation der Lehre" veröffentlicht, mit Ausnahme weniger Studien konnten jedoch keine deutschen empirischen Untersuchungen herangezogen werden, die sich eigens der Frage zur Validität studentischer Veranstaltungskritik und zu den Einflüssen auf studentische Urteile zur Lehre angenommen haben (Ausnahmen z.B. Daniel, 1995; Rindermann/Amelang, 1994; Kromrey 1995).

Dies wirft die Frage der Übertragbarkeit auf: Inwieweit können Ergebnisse, die in großer Fülle in den USA gesammelt wurden, auf hiesige Verhältnisse angewandt werden? Bekanntlich unterscheidet sich das nordamerikanische Hochschulsystem in verschiedenen Punkten deutlich vom deutschen System. Dies wirkt sich auf die untersuchten Themen aus, von denen einige in Deutschland nicht die gleiche Rolle spielen dürften wie in den Staaten (z.B. Noten), während zu anderen (z.B. Auswirkung der "Vermassung" von Kursen auf die Lehrbewertung) Untersuchungen fehlen.

Allerdings scheinen die typischen Einwände gegen Lehrbewertungen durch Studierende von seiten der Dozierenden durchaus vergleichbar zu sein. So nennt der deutsche Wissenschaftler Süllwold (1992) als mögliche externe Einflußgrößen, welche die Lehrbewertung verzerren: Fachsemester, geschlechtsspezifische Zusammenhänge, voruniversitäre Bildung, Intelligenzniveau, Fleiß, Leistungsmotivation und ideologische Bedürfnisse.

Dies deckt sich weitgehend mit den von Aleamoni (1987) aufgelisteten häufigsten Einwänden gegen studentische Veranstaltungskritik:

1. Studierende können die Lehre nicht konsistent beurteilen, da sie folgende Eigenschaften aufweisen: "Unreife, fehlende Erfahrung und Launenhaftigkeit".

2. Nur Kollegen/ Kolleginnen können die Lehre entsprechend kompetent bewerten.
3. Die meisten Bewertungsbogen sind nichts weiter als ein "Beliebtheitswettbewerb".
4. Studierende können die Lehre nicht bewerten, bis sie nicht genügend Distanz zum Kurs bzw. zur Lehranstalt besitzen.
5. Bewertungsbogen sind unreliabel und nicht valide.
6. Externe Variablen (z.B. Kursgröße) beeinflussen die Bewertung.
7. Noten (erwartete oder erhaltene) beeinflussen die Bewertung.

An dieser Aufstellung von Argumenten gegen studentische Veranstaltungskritik orientiert sich die Aufarbeitung jener Studien, die der Frage nachgegangen sind, welche Faktoren möglicherweise die studentische Lehrbewertung beeinflussen oder verzerren. Insgesamt wird auf elf Argumente eingegangen, die von unterschiedlicher Reichweite sind (vgl. Übersicht 3).

Grundsätzliche Einwände gegen die Gültigkeit und damit Verwendbarkeit studentischer Veranstaltungskritik heben in erster Linie die mangelnde Sachkenntnis und Unerfahrenheit der Studierenden hervor. Deren Urteile würden deshalb nicht den inhaltlichen Wert einer Veranstaltung oder die tatsächliche Lehrleistung erfassen, sondern vielmehr den Unterhaltungswert und die bloße "Beliebtheit" von Lehrenden.

Andere Argumente beziehen sich darauf, daß die Bewertung eines Lehrenden vom Kurs, den er abhält, abhängig ist. Dabei werden sehr verschiedene Aspekte angesprochen: Kursart und Kursgröße, Kursaufwand und Kurslevel, Pflicht- oder Wahlkurs, Kurs für Hauptfach- oder Nebenfachstudierende. Damit werden vor allem Fragen der Vergleichbarkeit der Bewertung der Lehrenden thematisiert.

Schließlich werden Eigenschaften der Lehrenden wie der Studierenden angeführt, die in Verdacht stehen, die Bewertungen der Lehre zu verzerren. Darunter spielen Eigenschaften wie Strenge der Lehrenden, insbesondere in der Notengebung, oder die Leistungsmotivation der Studierenden eine Rolle. Aber auch das Geschlecht und der Status des Lehrenden (Professor, Assistent, Tutor) würden die Beurteilung des Lehrangebotes beeinflussen, da sich die Studierenden in der Bewertung stärker daran als an der zu erfassenden Lehrleistung und Lehrqualität orientierten.

Inwieweit empirische Untersuchungen derartige Vorbehalte gegenüber studentischer Veranstaltungskritik stützen, einschränken oder widerlegen, das soll im Folgenden geprüft werden. Es geht um die Qualität des "Produktes Lehrleistung" und deren Beurteilung durch Studierende.

## Übersicht 3

### **Argumente gegen die Validität der Bewertung von Lehrleistungen durch Studierende**

#### **1 Der Unterhaltungswert, nicht der Inhaltswert steht für Studierende im Vordergrund**

Studierende würden nur oder überwiegend einen spannenden und witzigen Vortragsstil beachten, dem wichtigen Inhaltswert der Lehre schenken sie kaum Aufmerksamkeit.

#### **2 Die Beliebtheit, nicht die Lehrleistung wird von Studierenden beurteilt**

Studierende würden vor allem emotional die Beliebtheit eines Lehrenden bewerten und nicht sachlich dessen Lehreffektivität und -qualität.

#### **3 Unerfahrenheit und Unreife der Studierenden**

Studierenden fehle es an Erfahrung und zum Teil am notwendigen Ernst. Diese Unreife führe zu beliebigen und beeinflussbaren Urteilen.

#### **4 Nicht der Lehrende, sondern der Kurs wird bewertet**

Lehrende würden in verschiedenen Kursen unterschiedlich bewertet, so daß ihre "Lehrqualität" vom Kurs abhinge, den sie abhalten. Deshalb sei eine vergleichende Lehrbewertung unzulässig.

#### **5 Arten und Formen von Kursen verzerren die Lehrbeurteilung**

Art, Größe, Anforderungen, Aufwand und Besucher verzerren die Bewertung der Veranstaltungsqualität.

#### **6 Der Status der Lehrenden verzerrt die Lehrbewertung**

Studierende lassen sich vom Status und Ruf eines/einer Lehrenden beeindrucken, so daß sie weniger die konkrete Lehrleistung beurteilen.

#### **7 Persönliche Eigenschaften von Lehrenden verzerren die Bewertung**

Studierende bewerten nicht die Leistungen in der Lehre, sondern statt dessen die Herzlichkeit, Freundlichkeit und auch den Humor eines Lehrenden.

#### **8 Persönliche Eigenschaften der Studierenden verzerren die Bewertung**

Studierende mit unterschiedlich ausgeprägten Eigenschaften (wie Motivation oder Intelligenz) kommen zu anderen Beurteilungen; insofern hängt die "Lehrqualität" weniger vom Lehrenden als von den Studierenden (und ihren Eigenschaften) ab.

### **9 Übereinstimmungen zwischen Lehrenden und Studierenden verzerren die Bewertung**

Studierende, die in ihren Einstellungen und Haltungen mit denen der Lehrenden übereinstimmen, bewerten deren Veranstaltungen besser.

### **10 Das Geschlecht beeinflusst die Bewertung**

Sowohl die Lehrenden, die bewertet werden, als auch die Studierenden, die ihre Urteile abgeben, richteten sich nach dem Geschlecht des Gegenüber. Deshalb käme es zu unterschiedlichen Beurteilungen.

### **11 Noten und Strenge beeinflussen die Bewertung**

Bessere oder schlechtere Noten (erwartete oder erhaltene) ebenso wie der Eindruck der Milde oder Strenge des Lehrenden seien ausschlaggebend für die Lehrbewertung durch Studierende.

## **5.1 Erstes Argument: Der Unterhaltungswert dominiert den Inhaltswert**

Bedenken gegen die Lehrbewertung der Studierenden richten sich häufig darauf, daß Studierende nur oder überwiegend auf einen spannenden Vortragsstil achten und dabei dem Inhalt kaum Aufmerksamkeit schenken würden. Eine Reihe von Studien haben sich dieser Frage gewidmet: Bestimmen Ausdrucksstärke und Unterhaltungswert oder die Inhaltsdichte und der Lerngewinn das Urteil der Studierenden. Diese Arbeiten sind als „Dr. Fox-Studien“ bekannt geworden.

Der Begriff "Dr.Fox-Effekt" stammt aus der Studie von Naftulin, Ware und Donnelly (1973), die in sechs kurzen Videofilmen Inhaltsdichte (Anzahl genannter Fakten) und Expressivität (Unterhaltungswert) variierten. Der angebliche Dozent "Dr.Fox" wurde dabei von einem Schauspieler dargestellt. Mit diesem Effekt ist gemeint, daß eine Lehrveranstaltung von den Studierenden vor allem nach dem "unterhaltsamen" Vortrag bewertet wird, weniger nach der inhaltlichen Stoffdichte und dem Lernzugewinn.

Ware und Williams (1975) ließen 207 Studierende in sechs unterschiedlichen Kursen Videofilme (20 min) bewerten. Die Filme variierten bzgl. Inhaltsdichte und Vortragsart (expressiv, d.h. enthusiastisch, humorvoll, freundlich, ausdrucksstark und charismatisch in der einen Version und wenig expressiv in der anderen). Danach füllten die Studierenden einen Fragebogen zur Lehre und einen Wissenstest aus. Sowohl der Inhalt als auch der Vortragsstil hatten im einzelnen Auswirkungen auf die Testergebnisse. Diese fielen für den expressiveren Vortragsstil insgesamt besser aus. Auch eine Interaktion zwischen Inhalt und Vortragsstil hatte keine Auswirkung auf die Ergebnisse des Wissenstests.

Williams und Ware wiederholten 1977 ihre Studie von 1975, um zu überprüfen, inwieweit Goffmans These (1959) zutrifft: "...expressives Verhalten kann eine Zuhörerschaft so stark oder stärker beeinflussen als der Inhalt, wenn es wenig Zeit oder Grund für die Zuhörerschaft gibt, die Präsentation zu beurteilen" (zitiert nach Ware und Williams, 1977, S. 455). Aus diesem Grund ließen sie nicht nur ein zwanzigminütiges Videoband bewerten, daß in der üblichen Art variierte, sondern zwei aufeinanderfolgende Bänder, die nacheinander bewertet wurden, um die Beobachtungsdauer zu verlängern. Goffmans These konnte zumindest für dieses Untersuchungsdesign nicht bestätigt werden, aufgrund der kurzen Präsentationszeit aber auch nicht widerlegt werden.

Meier und Feldhusen (1979) konnten in ihrer Studie viele der Befunde von Williams und Ware (1977) replizieren.

Ziel der Studie von Marsh und Ware (1982) war ebenfalls die Überprüfung des "Dr. Fox"- Effekts mit Hilfe der Variation von Ausdruckstärke und Inhaltsdichte. Jedoch versuchten die Autoren dabei sowohl die Multidimensionalität der Lehrbewertung zu erreichen als auch ein natürlicheres Setting zu wählen. Ausgehend von der These, daß Studierende mit einer bestimmten Zielmotivation in den Lehrveranstaltungen sitzen (z.B. mit Blick auf anschließende Prüfungen), wurde versucht, die extrinsische Motivation der Stoffaufnahme zu erhöhen.

Für diese Überprüfung wurden eine Reanalyse zweier älterer "Dr.Fox"-Studien durchgeführt. Eine war die geschilderte Studie von Ware und Williams (1975), die andere eine von den beiden Wissenschaftlern 1976 durchgeführte frühere Untersuchung. Diese wurde mit der Erweiterung wiederholt, daß die teilnehmenden Studierenden eine extrinsische Motivation bekamen: 1\$ für die Teilnahme und 0,05\$ für jede richtige Antwort eines anschließenden Tests. Über die "Belohnung" wurde die Hälfte der Studierende vor der Filmpräsentation und die andere Hälfte der Studierende nach der Filmpräsentation informiert.

Bei der erneuten Analyse der Werte wurde der verwendete Fragebogen zur Bewertung der Lehre einer Faktorenanalyse unterzogen. Das Ergebnis ließ fünf gut interpretierbare Faktoren zu: Klarheit und Organisation, Gerichtetheit und Konzentration, Wissen und Kenntnisse, Begeisterung und Enthusiasmus, Lernanregung und Motivierung. Diese Faktoren wurden dann als abhängige Variablen angesehen, die den Einfluß experimenteller Manipulation demonstrieren sollten.

Die Untersuchungsanlage dieser Studie werten die Autoren wie folgt: "Diese Bedingungskonstellation ist besonders wichtig, da sie dem typischen universitären Kurs am nächsten kommt. Studierende wissen von Beginn des Kurses an, daß ihre Examensleistungen belohnt werden - Noten sind wahrscheinlich ein höherer Anreiz als das Angebot von 5 cents für jede richtige Antwort" (Marsh und Ware, 1982, S. 132).

## Übersicht 4

### **"Dr. Fox"-Studien: Inhaltsdichte und Unterhaltungswert - Befunde hinsichtlich der Effekte auf Lehrbewertung und Lernerfolgstest**

#### **1. Der erste Eindruck**

Der deutlichste Effekt auf die Ergebnisse der Lehrbewertung durch Studierende geht von der Variation der Ausdruckstärke aus. Der Unterhaltungswert erscheint wichtiger als der Inhaltswert.

Allerdings wird die Lehrbewertung auch von den Variablen Inhaltsdichte und Belohnung beeinflusst. Zwar hat die Änderung der Ausdruckstärke einen ziemlich substantiellen Einfluß, dieser variiert jedoch in hohem Maße unter den fünf untersuchten Faktoren wie Klarheit, Wissen, Lernanregung, Begeisterung oder Konzentration.

#### **2. Eine notwendige Differenzierung**

Besonders ausgeprägt ist der Einfluß des Unterhaltungswertes (Ausdruckstärke) auf die Dimension "Instructor enthusiasm" (Begeisterung), geringer bei allen anderen Faktoren.

Unter der Bedingung, daß die Belohnung bereits vorher angekündigt wird, hat die Ausdruckstärke auf keine andere Dimension Einfluß mit Ausnahme des Faktors "Instructor Enthusiasm". Die Variation der Ausdruckstärke beeinflusst demnach die Lehrbewertung nicht allgemein, sondern wirkt sich nur auf den damit verbundenen Faktor aus, was keine Verzerrung der Ergebnisse bedeuten würde.

#### **3. Ein zweiter Blick**

Der Einfluß der Variable Inhaltsdichte auf eine günstige Lehrbewertung ist zwar geringer, variiert aber ebenfalls über die verschiedenen (fünf) Dimensionen. Diese Variable hat aber kaum Einfluß auf die Bewertung des Faktors "Instructor Enthusiasm" (Begeisterung). Der Einfluß der Inhaltsdichte ist jedoch bei niedriger Ausdruckstärke größer.

#### **4. Aufklärung und Revision des ersten Eindrucks**

Die besten Ergebnisse im Wissenstest (Lernerfolg) werden unter der Bedingung erbracht: "Hohe Ausdruckstärke - Hohe Inhaltsdichte - Belohnungsankündigung vor Vortrag". Dabei hat der Zeitpunkt der Ankündigung der Belohnung einen hohen Einfluß auf die Testergebnisse. Liegt dieser Zeitpunkt nach der Präsentation des Vortrags, so hat die Ausdruckstärke einen stärkeren Einfluß auf die Ergebnisse als die Inhaltsdichte; wird die Belohnung jedoch vor dem Vortrag angekündigt, so ist der Einfluß der Inhaltsdichte hoch, der der Ausdruckstärke jedoch vernachlässigbar gering.

Die Schlußfolgerung, die Marsh und Ware aus ihrer Untersuchung für die Lehre ziehen: "Augenscheinlich ist die Nutzung von Humor, Stimmmodulation und Enthusiasmus wichtig, wenn die extrinsische Motivation niedrig ist, aber weniger bedeutsam, wenn die Studierenden bereits motiviert sind" (Marsh und Ware, 1982, S. 133). Besonders interessant ist diese Schlußfolgerung, da Ware, der 1973 als Mitautor den Dr. Fox-Effekt mit"kreierte", damit seine früheren Ergebnisse deutlich relativiert. - Die Entwicklung der Einschätzung des „Dr. Fox-Effektes“ aufgrund der empirischen Studien ist in Übersicht 4 zusammengefaßt.

Die Befunde machen deutlich, daß der Dr. Fox-Effekt vor allem in Situationen auftritt, die in der Hochschule selten sind, nämlich der Stoffpräsentation eines (wie in den älteren Dr.Fox-Studien) fachfremden Themas, dem das Auditorium in keinsten Weise zweckgerichtet zuhört. Mit den überzeugenden Ergebnissen von Marsh und Ware (1982), die auf alle drei der vorher dargestellten Studien angewendet werden können, dürfte der fast ein Jahrzehnt durch die amerikanische Literatur zur studentischen Veranstaltungskritik geisternde Effekt stichhaltig aufgeklärt und in seiner Bedeutung zutreffend eingeordnet worden sein.

### **Schlußfolgerungen**

*In früheren, undifferenzierten Studien konnte der Eindruck entstehen, als würde der Unterhaltungswert (die Ausdrucksstärke) eines Vortrages von Studierenden mehr in ihre Urteile einbezogen als dessen Stoff- und Inhaltsdichte.*

*Weitere, differenzierende Studien haben diesen ersten Eindruck (als Dr. Fox-Effekt bekannt geworden) jedoch korrigiert. Dies aus zwei Gründen: Erstens trennen die Studierenden selbst sehr deutlich zwischen diesen beiden Aspekten (Dimensionen); zweitens achtet eine interessiert-motivierte Hörerschaft vermehrt auf den inhaltlichen Wert und Nutzen.*

*Die Befürchtung, daß Studierende vor allem den witzigen und unterhaltsamen Vortrag positiv bewerten, konnte nicht bestätigt werden. Studierende, die mit dem Ziel etwas zu lernen oder eine Prüfung zu bestehen, einen Kurs besuchen, bewerten diesen vor allem nach seinem Nutzen und nicht nach seinem Unterhaltungswert.*

*Im übrigen erscheint es durchaus angemessen, für die Qualität einer Lehrleistung unter anderem auch die Ausdrucksstärke und Unterhaltsamkeit des Vortrages einzubeziehen, da dadurch Aufmerksamkeit stabilisiert und die Lernmotivation erhöht wird. Dies begründet sich zudem dadurch, daß der Lernerfolg bei den Studierenden ganz offensichtlich dann am größten ist, wenn beides zusammenkommt: eine hohe Ausdrucksstärke und eine hohe Inhaltsdichte.*

*In Instrumenten zur studentischen Veranstaltungskritik wäre aber auf alle Fälle darauf zu achten, beide Aspekte einer Lehrveranstaltung getrennt zu erfassen, um mögliche Vermischungen zu vermeiden.*

## **5.2 Zweites Argument: Beliebtheitsurteil statt Leistungsurteil**

Wenn Lehrveranstaltungskritik der Vorwurf gemacht wird, nichts weiter als ein "Beliebtheitswettbewerb" zu sein, heißt das, daß ein starker "halo"- Effekt vermutet wird: Studierende würden überwiegend aufgrund der Beliebtheit eines Dozenten/einer Dozentin die Fragen beantworten. Marsh (1982) vermutet, daß dies vor allem dann der Fall sei, wenn es sich um kurze einfache Fragebogen handelt, die sehr global die Lehreffektivität zu ermitteln versuchen.

Die Studie von Barbara R. Sherman und Blackburn (1975) gehört zu den wenigen, die direkt mit dem Konstrukt "Beliebtheit" operieren. Sie wurde an einem College durchgeführt, wo üblicherweise die Studierenden die Mitglieder des Lehrkörpers mit einem "Institutional Research Instrument" bewerteten, das sowohl die Präsentation des Lehrstoffs als auch das Verhalten des Veranstaltenden gegenüber den Studierenden erfaßt. Weiter standen die Daten einer von Studierenden selbst entworfenen und durchgeführten Evaluation zur Verfügung, bei der vor allem Fairneß, Nützlichkeit und Wichtigkeit einer Veranstaltung beurteilt wurden. Als drittes Instrument wurde (in Anlehnung an Osgood) ein Semantisches Differential (Pequod) verwandt, bei dem die Studierenden anhand von 30 bipolaren Begriffspaaren, die sich sowohl auf die Persönlichkeit als auch auf die Lehreffizienz bezogen, ihre Hochschullehrer und - lehrerinnen beurteilten.

An Ergebnissen ist aus dieser Studie festzuhalten:

1.) Die Korrelation zwischen den Werten aus dem "Institutional Research Instrument" und dem Polaritätenprofil betrug  $r = .77$ . Zwischen dem "Institutional Research Instrument" und den Eigeneinstufungen der Studierenden betrug die Korrelation  $r = .73$  und zwischen dem Polaritätenprofil und dem "Pequod" war  $r = .86$ . Alle Korrelationen sind signifikant.

2.) Auch die Multiple Regressionsanalyse ergab Regressionswerte zwischen  $R = .74$ - $.88$  für die Vorhersagbarkeit der Werte des "Pequod" und des "Institutional Research Instruments" aus den Ergebnissen des Polaritätenprofils, was nach Ansicht der Autorin und des Autors die Möglichkeit demonstriert: "... den Erfolg eines Lehrenden im Kurs auf Grundlage seiner Persönlichkeitseigenschaften vorherzusagen" (Sherman, Barbara R. und Blackburn, 1975, S. 130).

Barbara R. Sherman und Blackburn definieren leider an keiner Stelle ihres Berichtes ihre Vorstellung des Begriffes "Persönlichkeitscharakteristika" im Ver-

gleich zu den von ihnen verwendeten Konzept des Verhaltens eines/einer Lehrenden in Bezug auf "Kursfunktion und -aktivität". Somit bleibt die berechnete Frage, ob mit dem an Osgood angelehnten Instrument nur "Persönlichkeitscharakteristika" oder aber auch andere Parameter des "funktionalen Verhaltens" in der Lehrsituation erfaßt werden. Ist letzteres der Fall, so wäre nicht mehr genau zu bestimmen, welches Ausmaß der vorhersagbaren Varianz wirklich auf Persönlichkeitsfaktoren zurückzuführen wäre, sondern nur, daß die drei verwandten Instrumente gut übereinstimmen und auch wahrgenommene Persönlichkeitseigenschaften bei der Lehrbewertung eine Rolle spielen.

### **Schlußfolgerungen**

*Die verschiedenen Studien machen deutlich, daß die Operationalisierung des Konstrukts "Beliebtheit" nicht einfach ist. Schwierigkeiten bereitet in solchen Fällen, Ursache und Wirkung exakt zu trennen und zuzuordnen. Kommt es z.B. zu einer hohen Beliebtheit, weil jemand ein guter Lehrer/ eine gute Lehrerin ist, oder erhält jemand gute Bewertungen, weil er/ sie beliebt ist?*

*Offensichtlich hängen Urteile zur Lehrqualität einer Veranstaltung und wahrgenommene Eigenschaften von Lehrenden miteinander zusammen. Allerdings bleibt zu bedenken, daß ein solcher Zusammenhang nicht gegen die Validität der studentischen Veranstaltungskritik spricht, insofern für eine gute Lehrleistung bestimmte Eigenschaften vorauszusetzen sind bzw. in sie einfließen.*

### **5.3 Drittes Argument: Unreife und Unerfahrenheit der Studierenden**

Manche Lehrende bezweifeln, daß Studierende die Lehre beurteilen können, da sie zu unreif und unerfahren seien, manche unterstellen sogar, daß ihnen der notwendige Ernst fehle, weshalb ihre Urteile "beliebig" ausfallen und starke Schwankungen aufweisen würden. Im folgenden werden daher Studien zur Konsistenz der Beurteilung bei studentischer Veranstaltungskritik und zu Beurteilungsfehlern aufgrund von "Beeinflussung durch ältere Semester" oder "Beeinflussung durch expressiven Vortragstil" aufgeführt. Die Annahme, daß diese beiden Punkte auf die Lehrbewertung Einfluß haben, kann als vermuteter Mangel an Erfahrung (bzw. fehlende Beobachtungsschulung) gewertet werden.

#### **Konsistenz oder Beliebtheit**

Konsistenz kann definiert werden als die Übereinstimmung von Lehrbewertungen für denselben Lehrenden bzw. dieselbe Lehrende in verschiedenen Kursen mit denselben Studierenden, zu verschiedenen Zeiten oder zu verschiedenen Themen.

Marsh (1982) vergleicht Daten aus 8.277 Kursen, die in 35 verschiedenen Fachbereichen über vier Jahre erhoben wurden. Das verwendete Instrument war der Fragebogen "Students' Evaluations of Educational Quality" (SEEQ), der 35 verschiedene Items und einige demographische Fragen enthält.

Marsh fand Korrelationen von  $r=.70$  für die Bewertung jener Lehrender, die den gleichen Kurs zu zwei verschiedenen Zeitpunkten abhielten. Dies drückt eine hohe Übereinstimmung und Konsistenz aus und gibt keine Hinweise auf eine Beliebbarkeit studentischer Urteile aufgrund ihrer "Unreife" oder Unerfahrenheit.

Seiler, Weybright und Stang (1977) nennen Korrelationen aus den von ihnen untersuchten studentischen Veranstaltungskritiken:

- für Lehrende, die gleiche Kurse im gleichen Jahr jedoch zu verschiedenen Zeitpunkten gaben:  $r=.48$ ;
- für Lehrende, die den gleichen Kurs in verschiedenen Jahren gaben:  $r=.38$ ;
- für Lehrende, die unterschiedliche Kurse im gleichen Jahr unterrichteten:  $r=.30$ ;
- und für Lehrende, die unterschiedliche Kurse in unterschiedlichen Jahren unterrichteten:  $r=.20$ .

Während eine noch hinreichende Konsistenz in dieser Untersuchung für gleiche Kurse zu verschiedenen Zeitpunkten oder Jahren gefunden wurde, fällt sie bei unterschiedlichen Kursen in unterschiedlichen Jahren stark ab.

### **Beeinflussung durch ältere Semester**

Terry und McIntosh (1988) untersuchen den Einfluß, den Kurserwartungen von Psychologiestudierenden auf die Mitte- und Endsemesterbewertung haben. Insgesamt dreimal beurteilten 61 Studierende Kurs und Dozenten/ Dozentin: Vor Kursbeginn wurden die Erwartungen an den Kurs erfaßt. In einer Zusatzfrage gaben 81 Prozent der Studierenden an, daß sie etwas bzw. einiges über den Kurs gehört hätten, 85 Prozent davon gaben an, daß ihre Informationen von anderen Studierenden stammen. Der Fragebogen wurde Mitte und Ende des Semesters wieder verteilt.

Insgesamt fanden die Autoren geringe, jedoch signifikante Korrelationen zwischen der ersten und dritten Beurteilung ( $r=.23$ ) und zwischen der ersten und zweiten Beurteilung ( $r=.22$ ). Die Korrelation zwischen der Mitte- und der Endsemesterbewertung dagegen betrug  $.61$ .

Die Studie läßt die Frage offen, ob Vorinformationen älterer Semester valide sind und es so zum ermittelten Zusammenhang kommt, oder aber, ob Studierende in ihrer Bewertung von "falschen" Erwartungen beeinflusst werden. Da die Korre-

lationen zwischen Vor-Kursbeginn und späteren Erhebungen nicht sonderlich hoch sind, kann angenommen werden, daß sich die Studenten und Studentinnen während des Semesters eine eigene Meinung über die Fähigkeiten des/ der Lehrenden bilden.

### **Schlußfolgerungen**

*Das Argument, die Studierenden seien zu unreif, um die Lehre zu bewerten, konnte nicht belegt werden. Die umfangreiche Studie von Marsh (1982) läßt sogar ein ausgesprochen konsistentes Verhalten der Studierenden bei der Lehrbewertung vermuten.*

*Die Studierenden scheinen sich nur in sehr geringem Umfang von 'Gerüchten' über die Lehrqualität der Lehrenden oder durch Informationen älterer Semester beeinflussen zu lassen. Diese Einflüsse erscheinen so niedrig, daß sie die Validität vergleichender Lehrbewertungen durch Studierende nicht bedeutsam vermindern.*

*Aufgrund der recht hohen Stabilität studentischer Urteile und ihrer geringen Beeinflußbarkeit kann daher ein Mangel an Erfahrung oder gar Unreife den Studierenden nicht unterstellt werden.*

#### **5.4 Viertes Argument: Nicht die Lehrenden, sondern der Kurs bestimmt die Lehrqualität**

Eine erstaunlicherweise nicht so häufig untersuchte Frage ist die nach dem Ausmaß des Kurseinflusses. Je nachdem, ob es sich um eine große dozentenorientierte Veranstaltung oder ein kleines Seminar handelt, das stark von der Beteiligung der Studierenden abhängt, sind Unterschiede vorstellbar.

Diese Unterschiede könnten sich auf so typische Fragen, wie die nach der Globalbewertung des Kurses oder auf die Einschätzung nach dem Ausmaß des Gelernten auswirken.

Romney (1976) versuchte, den Kurseffekt und den Effekt der/ des Lehrenden auf die Lehrbewertung mit Hilfe eines varianzanalytischen Verfahrens zu separieren. Ergebnis der varianzanalytischen Auswertung war, daß der Kurs die Lehrbewertung in gleichem Maße beeinflußt wie der Lehrende. Problematisch ist bei dieser Untersuchung der hohe Anteil an Kursen aus dem ersten Studienjahr.

In verschiedenen anderen Studien wurde die unterschiedliche Art der Bewertung von Studierenden aus niedrigen und höheren Semestern untersucht. Es wäre nicht

auszuschließen, daß höhere Semester es besser gelernt haben, den/ die Lehrende stärker vom Kurs zu trennen (zur methodischen Kritik an Romney vgl. Gillmore, 1977).

Kurseffekt und Lehrendeneffekt versuchte auch Marsh (1982) mit Hilfe einer Pfadanalyse zu trennen. Die Daten aus 8.277 Kurse wurden zu 341 Sets zusammengestellt. Jedes Set beinhaltete folgende Kombination von „Kurs“ und „Lehrenden“:

- Zwei Bewertungen des gleichen Kurses, unterrichtet von der/dem gleichen Lehrenden zu zwei verschiedenen Zeitpunkten.
- Bewertung verschiedener Kurse bei dem/ der gleichen Lehrenden.
- Zwei Bewertungen des gleichen Kurses von zwei verschiedenen Lehrenden.
- Zwei Kurse unterrichtet von verschiedenen Lehrenden.

Die Korrelationen im Set ergaben als Ergebnisse ganz unterschiedliche Stufungen der Übereinstimmung in der Beurteilung der Lehrenden:

- $r=.70$  für den gleichen Instruktor/ die gleiche Instruktorin bei gleichem Kurs zu unterschiedlichen Zeitpunkten; da diese Übereinstimmung geringer ist, als die Reliabilität des verwandten Instruments (.90), kommt Marsh zu der Schlußfolgerung: "Dies läßt vermuten, daß ein substantieller Teil der Varianz (über 20%) bei der studentischen Bewertung reliabel ist, aber einmalig auf den bestimmten Kurs zurückzuführen ist" (ebd., S. 51),
- $r=.52$  für den gleichen Instruktor/ die gleiche Instruktorin bei verschiedenen Kursen,
- durchschnittlich  $r=.14$  bei verschiedenen Lehrenden, die den gleichen Kurs unterrichten, wobei eine große Variabilität zwischen den einzelnen Faktoren festgestellt wurde.

Mit verschiedenen Pfadanalysen versucht Marsh, den Kurs- bzw. den Lehrendeneffekt auf die Lehrbewertung zu schätzen. Je nach Modell variiert die durchschnittliche Größe des Lehrendeneffekts zwischen .50 und .57 und die des Kurseffekts nur zwischen .10 und .18.

Mit dieser differenzierten Analyse konnte Marsh belegen, daß Lehrende in verschiedenen Kursen zwar nicht völlig gleich bewertet werden, daß aber die Rolle des Kurses gegenüber dem Gewicht des Lehrenden für dessen Evaluation weit geringer ist.

Diese Befunde sprechen auch dafür, daß das Lehrverhalten der Dozenten zumeist in verschiedenen Lehrveranstaltungen in gleicher Weise strukturiert ist. Deshalb scheint bei der praktischen Verwendung die Rolle des Kurses nahezu vernachlässigbar.

## **Schlußfolgerungen**

*Die aufgeführten Untersuchungen zum Ausmaß des Kurseinflusses auf die Lehrbewertung müssen im Zusammenhang mit den Konsistenzstudien gesehen werden (vgl. Abschnitt 5.3).*

*Insgesamt kann zwar davon ausgegangen werden, daß gewisse Varianzanteile existieren, die nur auf den entsprechenden Kurs zurückzuführen sind, doch der Anteil der konsistenten Beurteilung der Lehrenden liegt im Vergleich dazu deutlich höher und erscheint für eine vergleichende Lehrbewertung ausreichend.*

*Sicherlich ist es nach diesen Befunden nicht vertretbar, aufgrund der Evaluation eines einzelnen Kurses allgemeine Aussagen über die Lehrleistung des/der Dozentin vorzunehmen.*

*Ebensowenig sprechen aber unterschiedliche Bewertungen eines Lehrenden in verschiedenen Kursen oder zu verschiedenen Zeiten für eine geringe Validität der messenden Instrumente. Das würde nur dann zutreffen, wenn unterstellt wird, daß Lehrende in allen ihren Veranstaltungen eine für die Studierenden gleiche Lehrleistung erbringen.*

### **5.5 Fünftes Argument: Art und Form des Kurses beeinflussen die Bewertung**

Neben der globalen Frage, ob der "Kurs" wichtiger als der "Lehrende" für die Bewertung sei - dies kann deutlich verneint werden - bleiben selbstverständlich eine ganze Reihe von Fragen offen: Inwieweit beeinflussen Art, Form und Größe eines Kurses die Bewertung von dessen Qualität? Sechs derartige "Kursmerkmale" werden nachfolgend betrachtet. Mögliche Unterschiede nach Art und Form der Kurse in der "Lehrbewertung" würden dafür sprechen, dies bei einer vergleichenden Evaluation der Lehre von Dozenten zu berücksichtigen.

#### **Hauptfach versus Nebenfach**

Die Frage, ob z.B. Studierende eines Magisterstudiengangs einen Unterschied in der Bewertung ihrer Kurse im Hauptfach und im Nebenfach machen, wurde meines Wissens in Deutschland noch nicht untersucht. Dies erscheint jedoch gerade für die geisteswissenschaftlichen Fächer als eine relevante Forschungsfrage.

Danielsen und White (1976) analysierten die Befragungsdaten von über 5.300 Studierenden. Es konnten keine Unterschiede in den Bewertungen von Haupt- bzw. Nebenfachstudierenden festgestellt werden.

Felicita F. Romeo und Weber (1985) ermittelten demgegenüber, daß Kurse im eigenen Hauptfach signifikant schlechter bewertet werden.

Die Befunde zur Frage, ob Kurse im Haupt- oder Nebenfach unterschiedlich beurteilt werden, sind demnach nicht eindeutig. Ihr wäre demnach in weiteren Studien nachzugehen, wobei auch die Gründe für mögliche unterschiedliche Bewertungen aufzuklären wären.

### **Kurslevel bzw. Universitätsjahr der Studierenden**

Immer wieder wird vermutet, daß niedrigere Semester ihre Lehrenden signifikant schlechter bewerten als höhere Semester.

Danielsen und White (1976) und Patricia B. Elmore und Pohlmann (1978) konnten jedoch keine signifikanten Unterschiede in der Lehrbewertung unterschiedlicher Semester finden. Weder Blount et al. (1978) noch Felicita F. Romeo und Weber (1985) fanden Unterschiede in der Veranstaltungskritik von "graduate" und "undergraduate" Studierenden.

Richardson (1978) korrelierte die Antworten von 11.655 Studierende auf drei Globalitems "Lehrende", "Kurs" und "Examen" mit dem Kurslevel (Studienjahr). Alle drei Korrelationen waren signifikant, was aufgrund der großen Stichprobe schwer zu interpretieren ist, da große Stichproben dazu führen, daß auch geringe Differenzen "signifikant" im statistischen Sinne werden.

In der Studie von Bruton und Sue Crull (1982) hatte die Anzahl der Studienjahre einen geringen, aber doch signifikant negativen Einfluß auf die Kursbewertung. Allerdings waren die Kurse speziell für die Anfangssemester konzipiert.

Adriane Gaffuri et al. (1982) erhoben Daten in 85 Kursen. Es konnte nur eine Durchschnittskorrelation von .18 zwischen Items und dem Kurslevel gefunden werden.

Die meisten Studien erbringen keinen Zusammenhang zwischen Kurslevel bzw. Studienjahr und Lehrbewertung. Jedoch liegen einzelne Untersuchungen vor, bei denen zumeist geringe Unterschiede auftraten, wobei allerdings jeweils spezielle Bedingungen vorlagen.

### **Kursarten: Vorlesung, Übung, Seminar**

Die externe Einflußgröße der Kursart - Vorlesung, Übung, Seminar - wurde in den Vereinigten Staaten nicht untersucht. Sie scheint jedoch in Deutschland eine eindeutige Auswirkung auf die Kursbewertung zu haben.

Preißer (1992) konnte an der TU Berlin ermitteln, daß Seminare sowohl im Grund- als auch im Hauptstudium besser bewertet wurden als Vorlesungen. Gleiches konnte auch Daniel (1994) sowohl für die Wirtschafts- als auch für die Geisteswissenschaften an der Universität Mannheim ermitteln.

## **Kursgröße**

Eine Variable, die in den USA zwar häufig untersucht wurde, dort aber bei weitem nicht so relevant erscheint wie in Deutschland, ist die Kursgröße. Die Untersuchungen dazu beinhalten zwei Problembereiche. Einer davon ist eine mögliche Konfundierung zwischen Kursgröße und anderen Variablen. Große Kurse werden z.B. häufig in den ersten Semestern und in (ungeliebten) Pflichtkursen angeboten. Möglich erscheint auch, daß solche Kurse weniger gern unterrichtet werden oder an Lehrende mit niedrigerem akademischen Status und geringerer Lehrerfahrung abgegeben werden.

Der zweite Problembereich dieser Untersuchungen ist die Übertragbarkeit der Ergebnisse der amerikanischen Studien. Massenkurse, wie in einigen Fächern in Deutschland üblich, sind in den Staaten nahezu unbekannt. Die Definition "Großer Kurs" fängt dort schon bei der Mitgliederzahl einer großen Schulklasse an.

Ein extremes Beispiel für diese Problematik ist die Arbeit von Felicita F. Romeo und Weber (1985), die keine signifikanten Unterschiede für kleine vs. große Klassen fand. Dies könnte allerdings an ihrer Definition liegen, bei der "große Klassen" bereits bei mehr als 15 Studierenden beginnen.

Marcia K. Petchers und Chow (1988) ließen 856 Studierende des Fachbereichs Sozialarbeit ihre Kurse bewerten. Ein Einfluß der Kursgröße konnte nicht ermittelt werden.

Von einer negativen Korrelation mit der Klassengröße berichtet dagegen Scott (1977), d.h. in kleineren Kursen wurde die Lehre besser als in größeren beurteilt. Adriane Gaffuri et al. (1982) korrelierten für 85 Kurse die Fragen ihres Fragebogens mit der Kursgröße. Die Durchschnittskorrelation eines Items mit der Klassengröße war ebenfalls schwach negativ ( $r = -.13$ ).

Patricia B. Elmore und Pohlmann (1978) untersuchten eine ganze Reihe von möglichen Einflußvariablen. Die Korrelationen blieben insgesamt niedrig. Eine der noch am stärksten einwirkenden Variablen war die Kursgröße (negative Korrelation).

Für unterschiedliche Klassengrößen ermittelten Patricia Cranton und Schmith (1986) Effektgrößen zwischen .10-.20. Die Bewertungen für kleine Kurse bzw. für Lehrende mit mehr Erfahrung fielen besser aus. Bei der Untersuchung des

Effekts der Kursgröße nach einzelnen Semestern (aufgrund möglicher Konfundierungseffekte mit dem jeweiligen Semester) konnte nur für die Einschätzung des Gelernten ein Einfluß bestätigt werden. Am geringsten beeinflusste danach die Kursgröße die Bewertung der Effektivität des Lehrenden; dieser Effekt konnte aber nicht in allem Fachbereichen gefunden werden.

Eine Untersuchung, welche die Vermutung konfundierter Effekte unterstützt, haben Danielsen und White (1976) vorgestellt. Sie konnten einen signifikant positiven, allerdings nicht sehr großen Effekt für die Klassengröße ermitteln. Dies liegt jedoch u.U. an der Praxis der University of Georgia bzw. des College of Business Administration, wonach große Kurse nur von Lehrenden unterrichtet werden, die mit solchen Kursen gut zurecht kommen.

### **Kursschwierigkeit und Kursaufwand**

Lehrende befürchten häufig, daß schwierige oder aufwendigere Kurse schlechter bewertet werden. Leider liegen zu diesem Punkt nur wenige Untersuchungen vor; deren Ergebnisse sprechen allerdings nicht für einen Einfluß von Kursschwierigkeit oder Kursaufwand.

In der Studie von Firth (1979) wurde mit Hilfe der studentischen Einschätzung zum besten und schlechtesten Lehrenden nachgewiesen, daß das verwendete Evaluationsinstrument signifikante Unterschiede zwischen den beiden Extremgruppen (jeweils besten und schlechtesten 10% der Lehrenden) ermittelt. Dabei ermöglicht das Item "Schwierigkeit" und "Aufwand des Kurses" keine Unterscheidung zwischen den Gruppen. Das heißt, daß bei der Einschätzung eines Lehrenden die Kursschwierigkeit keine Rolle spielt. Centra (1977) konnte in einer ähnlichen Studie ebenfalls keinen signifikanten Einfluß ermitteln.

Coleman und McKeachie (1981) untersuchten die Auswirkung von Bewertungsergebnissen verschiedener Kurse auf die Kurswahl von Studierenden. Für den Kurs, der angeblich sehr positive Beurteilungen von früheren Studierenden erhalten hatte, aber als Kurs mit höherem Aufwand ausgewiesen war, hatten sich 40 Prozent mehr Studierende eingeschrieben als in thematisch vergleichbare Kurse.

### **Pflicht- oder Wahlkurs**

Brandenburg, Slinde und Batista (1977) untersuchten Korrelationen zwischen der Bewertung des "Illinois Course Evaluation Questionnaire" von 3.355 Studierenden und den Variablen: Klassengröße, Pflicht/ Wahlfach, erwartete Note, Kursniveau, Rang und Geschlecht des/der Lehrenden. Angesichts der Stichprobengröße verwundert es nicht, daß alle Effekte signifikant waren. Besonders hoch

fielen sie jedoch bei der erwarteten Note und der Variable Pflicht- oder Wahlfach aus.

Patricia B. Elmore und Pohlmann (1978) konnten keine Unterschiede für die Bewertung von Kursen ermitteln, die unterschiedliche Anteile an Pflichtbesuchern hatten.

In der Untersuchung von Marcia K. Petchers und Chow (1988) wurde der Frage nachgegangen, inwieweit externe Rahmenbedingungen Einfluß auf studentische Lehrbewertungen haben. Einzig der Faktor "Wahl- oder Pflichtkurs" beeinflusste zwei der drei Subskalen signifikant.

Eine mögliche Erklärung dafür liefern DuCette und Jane Kenney (1982), die herausfanden, daß in einigen Kursen stärker als in anderen die Lehrbewertung mit den Noten korrelierte. Dies sei z.B. in Statistikkursen besonders ausgeprägt. Dazu vermuten DuCette und Jane Kenney, daß diese Art von Kursen Pflichtkurse sind, an denen die Studierenden kein großes inhaltliches Interesse haben und Noten daher um so wichtiger werden.

Eine andere Erklärungsmöglichkeit wäre, daß Pflichtkurse häufig methodische o.ä. Kurse sind. Wie gut sie gemeistert werden, hängt dabei stärker als in anderen Kursen von Vorkenntnissen und Neigungen ab, die sich wiederum auf die Lernmotivation auswirken. Schätzen Studierende z.B. ihre erwartete Statistiknote gut ein, so kann man davon ausgehen, daß sie im entsprechenden Kurs gut mitgekommen sind. Gründe dafür sind außerdem ihre Vorkenntnisse und ihre Kursmotivation (vgl. auch Diehl, 1983).

Hofman (1988) untersuchte, inwieweit verschiedene Studienmotivationsmerkmale studentische Veranstaltungskritik beeinflussen. Insgesamt schließt Hofmann: "Die Hypothese, daß Merkmale der Studienmotivation eine bedeutsame Determinante der Beurteilung von Veranstaltungen darstellen, wird somit insgesamt gut bestätigt" (Hofman, 1989, S. 125).

Kromrey (1994a) geht sogar so weit, daß er die unterschiedlichen Lehrbewertungen vor allem auf eine unterschiedliche Motivation der Studierenden zurückführt. Daniel (1994) konnte jedoch ermitteln, daß weniger interessierte Studierende die Lehrenden zwar etwas schlechter bewerten, daß jedoch die Rangfolge der Bewertung mit denen interessierter Studierender identisch ist, sich somit nur parallel verschiebt.

Offenbar kann es sich in der Lehrbewertung niederschlagen, ob es sich bei einer Veranstaltung um einen Pflicht- oder Wahlkurs handelt. Dieser Faktor ist aber wohl hauptsächlich auf die unterschiedliche Motivation und das Interesse der Studierenden in diesen beiden Kursarten zurückzuführen.

## Schlußfolgerungen

*Es kann nicht pauschal ausgeschlossen werden, daß externe Einflußgrößen Auswirkungen auf die studentische Lehrbewertung haben. Diese müssen jedoch im einzelnen betrachtet werden.*

*Insgesamt gibt es kaum einen Nachweis dafür, daß niedrigere Semester schlechtere Bewertungen abgeben. In Deutschland, wo die Zusammensetzung der Erstsemester eine andere ist als in den Vereinigten Staaten (späterer Studienbeginn, heterogenere Alterszusammensetzung, z.T. verschiedene Vorerfahrungen: Lehre, Zivildienst, Bundeswehr, Praktika etc.) müßten eigene Forschungsprojekte mögliche Einflüsse, z.B. die einer abgeschlossenen Lehre, untersuchen.*

*Die Veranstaltungsart "Vorlesung" wird im Vergleich zu anderen Lehr-Lern-Formen relativ schlecht beurteilt. Dies müßte bei Kursvergleichen mitbedacht und berücksichtigt werden: Urteile zur Lehrqualität von Dozierenden sollten sich demnach bei Vergleichen nur auf die gleichen Veranstaltungsformen stützen.*

*Die in einigen Artikeln gefundenen Effekte zur Kursgröße sind alle eher niedrig. Dies kann jedoch auch an der verwendeten Definition eines „großen Kurses“, bzw. des Fehlens von "Massenkursen" in den USA liegen. Genauere Untersuchungen wären somit notwendig, um Aussagen für die Situation an deutschen Hochschulen machen zu können. Dabei müßte jedoch sichergestellt sein, daß die gefundenen Einflußgrößen tatsächlich auf die Kursgröße und nicht auf möglicherweise damit konfundierte Faktoren zurückzuführen sind.*

*Es wären ebenfalls weitere Studien an deutschen Hochschulen notwendig, um zu einer abschließenden Bewertung zu kommen, ob es eine Rolle spielt, daß ein Kurs in einem Haupt- oder Nebenfach angeboten wird.*

*Zusammengefaßt kann festgestellt werden, daß bei der Beurteilung von Lehrenden durch Studierende Kursschwierigkeit und -aufwand kaum Auswirkungen auf die Lehrbewertung haben.*

*Man kann allerdings davon ausgehen, daß der Faktor Motivation die Lehrbewertung beeinflusst. Vermutlich lernen weniger motivierte Studierende weniger im entsprechenden Kurs und bewerten ihn daher schlechter. Dies wäre somit sogar ein Zusammenhang, der sich mit den Überlegungen zu "Lernzuwachs und Validität" decken würde.*

*Insgesamt müßten noch vermehrt Studien zu einigen der genannten Einflußgrößen durchgeführt werden, um zu einer abschließenden Beurteilung zu kommen. Dies wäre vor allem dann wichtig, wenn einzelne Kurse verschiedener Lehrender miteinander verglichen werden sollen. Auf alle Fälle ist es wichtig, bei der Erfassung und Analyse von Lehrbewertungen die Art und Form der Kurse und die Mo-*

*tivation der Studierenden zu berücksichtigen. Damit wäre bei der Interpretation der Daten zu vermeiden, der Lehrqualität von Hochschullehrenden (positiv oder negativ) zuzuschreiben, was im Grunde auf die Kursart und die Zusammensetzung der Studierenden im Kurs zurückzuführen ist.*

## **5.6 Sechstes Argument: Der Status der Lehrenden verzerrt die Lehrbewertung**

Gelegentlich wird vermutet, daß auch der Lehrstatus sich auf die Lehrbewertung auswirkt. Entsprechende Korrelationen ließen sich entweder auf die längere Lehr- erfahrung von Lehrenden mit höherem Status zurückführen oder aber auf eine Urteilsverzerrung bei Studierenden, die sich durch den "Status" beeindrucken lassen und weniger die konkrete Lehrleistung beurteilen.

Patricia B. Elmore und Pohlmann (1978), Felicita F. Romeo und Weber (1985) und Marcia K. Petchers und Chow (1988) konnten jedoch keine Bewertungsunterschiede entsprechend dem Status der Lehrenden feststellen.

Die Autoren Meredith und Schmitz (1986) gingen der Frage nach, ob signifikante Unterschiede in der Bewertung der Lehre von Dozenten/ Dozentinnen und (stu- dentische) Tutoren/Tutorinnen existieren. Tutoren/ Tutorinnen erhielten sig- nifikant bessere Noten für "Feedback" und schlechtere für "Beherrschung der Materie", "Enthusiasmus", "Zusammenfassung am Ende einer Diskussion" und "Seminarthemen waren interessant". Signifikante Unterschiede im Lernausmaß konnten nicht ermittelt werden. Bei diesen Ergebnissen stellt sich die Frage, in- wieweit die Items "Enthusiasmus" und "Seminarthemen waren interessant" stär- ker wegen anderer Gründe (z.B. daß die von Tutoren/Tutorinnen unterrichteten Kurse oft Pflichtkurse des ersten Studienabschnitts waren) schlechter bewertet wurden.<sup>3</sup>

Centra (1980) bezieht sich u.a. auf eine eigene, frühere Studie und eine Untersu- chung von Creech (1976), in der eine schlechtere Bewertung für Tutoren/ Tuto- rinnen ermittelt wurde. In der Studie von Blount et al. (1978) konnten zwar nicht direkt bessere Ergebnisse für Lehrende mit höherem Status gefunden werden, je- doch ergaben paarweise untersuchte Werte (Newman-Keuls-Test), daß signifi- kant bessere Werte in einigen Dimensionen für Tutoren/Tutorinnen bzw. Assis- tenten und Assistentinnen als für Professoren und Professorinnen vergeben wur- den.

---

<sup>3</sup> Insgesamt ist die Frage nach der Qualität der Tutoren und Tutorinnen auch im Zusammen- hang mit der Diskussion um Ausweitung der Tutorenprogramme interessant. (Siehe: Ak- tionsprogramm "Qualität der Lehre", Abschlußbericht, Mai 1992, Hrsg. Ministerium für Wissenschaft und Forschung, NRW).

## **Schlußfolgerungen**

*Insgesamt kann man nicht davon ausgehen, daß Studierende einen Dozenten/ eine Dozentin höher bewerten, nur weil dieser/ diese einen höheren akademischen Titel hat. Jedoch erscheint es möglich, daß z.B. Aspekte, die mit akademischen Titel häufig korrelieren, wie z.B. Lehrerfahrung oder Art des Kurses, von Einfluß für die Lehrbewertung sind.*

### **5.7 Siebtes Argument: Eigenschaften von Lehrenden verzerren die Bewertung**

Wie die Diskussion um die Beliebtheit von Lehrenden (vgl. Argument 2, Abschnitt 5.2) geht die Frage nach dem Einfluß von Persönlichkeitseigenschaften wie Herzlichkeit oder Humor in eine ähnliche Richtung.

Der Kernpunkt beider Diskussionen ist, streng gesehen, eine Validitätsfrage: Werden Lehrende aufgrund ihrer Leistungen in der Lehre oder aufgrund ihrer Beliebtheit oder ihres Humors, d.h. unter der Einwirkung eines "Halo"-Effekts, eingeschätzt?

#### **Herzlichkeit**

Patricia B. Elmore und Pohlmann (1978) ermittelten, daß die "Selbstbeurteilung von Herzlichkeit" zwar nicht hoch, aber signifikant positiv mit der Lehrbewertung korreliert.

Die Herzlichkeit der Lehrenden wurde von Patricia B. Elmore und Karen A. La-Pointe (1975) als wichtige Variable identifiziert, die die studentische Bewertung der Lehreffektivität beeinflusst.

Terry & McIntosh (1988) fanden in ihrer Studie, die Herzlichkeit des oder der Lehrenden sei eine wichtige Erwartungsdimension der Studierenden, welche die spätere Bewertung beeinflusst.

Es ist zu vermuten, daß die Variable "Herzlichkeit" vor allem Einfluß auf Faktoren der Lehrbewertung hat, die die Beziehungsdimension der Lehre thematisieren und weniger auf andere Dimensionen der Lehrqualität (vgl. Kapitel 6).

#### **Humor und Hilfsbereitschaft**

In der Studie von Katherine van Giffen (1990) konnte ermittelt werden, daß der Einfluß von Humor insgesamt positiv mit der Lehreffektivität verbunden war ( r

= .532), ebenso Freundlichkeit ( $r = .639$ ) und Hilfsbereitschaft ( $r = .791$ ), wobei die letzten beiden mit  $r = .806$  miteinander korrelierten. Ähnlich wie beim Faktor der Herzlichkeit ist zu vermuten, daß Humor nur auf bestimmte Faktoren der Lehrbewertung, vor allem der Beziehungsebene, Einfluß hat.

## Schlußfolgerungen

*Bislang liegen nur wenige Studien zur Frage vor, ob persönliche Eigenschaften von Lehrenden die Bewertung ihrer Leistung in der Lehre beeinflussen. Sie sind zudem auf wenige Aspekte begrenzt. In der Regel handelt es sich um Faktoren der emotionalen Zuwendung. Für die bisher untersuchten Faktoren wie Herzlichkeit oder Humor ergaben sich Einflüsse auf die Bewertung der Lehre. Allerdings bleibt zu fragen, ob dabei von "Verzerrung" gesprochen werden kann, da ein förderliches soziales Klima auch die inhaltliche Qualität zu steigern vermag.*

*Deutlich werden in den vorliegenden Studien vor allem zwei Punkte: (1) Es ist meist unmöglich, eine eindeutige Kausalbeziehung zu bestimmen. (2) Die instrumentelle Trennung zwischen allgemeinen Persönlichkeitseigenschaften und Eigenschaften, die sich positiv auf die Lehre auswirken, scheint in keiner der Studien gelungen. Um letzteres genauer zu erläutern: Es erscheint nachvollziehbar, daß jemand, der die Eigenschaft hat, besonders zuverlässig zu sein, pünktlich in den Unterricht kommt.*

*Ob allgemeine Eigenschaften von Lehrenden die Lehrbewertung beeinflussen und somit ein "Halo"-Effekt vorliegt, kann nur mit Instrumenten erfaßt werden, die trennscharfe Faktoren besitzen. In einer solchen Anordnung ist dann belegbar, ob z.B. die Eigenschaftszuschreibung "aufgeschlossen" nicht nur mit adäquaten Items (z.B. behandelt auch neuere Themengebiete), sondern auch mit inadäquaten (z.B. kommt pünktlich) signifikant korreliert.*

## 5.8 Achstes Argument: Eigenschaften von Studierenden beeinflussen die Bewertung

Betrachtet man mögliche Auswirkungen von Eigenschaften der Lehrenden auf die Bewertung, so muß parallel dazu die Frage gestellt werden, ob unterschiedliche Studierendenpersönlichkeiten einen Einfluß auf die Ergebnisse studentischer Veranstaltungskritik haben. Anzunehmen ist, daß etwa leistungsmotivierte Studierende positiver urteilen könnten, aber auch die Zuschreibung von "Erfahrung" von Einfluß sein kann.

Ein genereller bzw. konsistenter Zusammenhang zwischen studentischer Veranstaltungskritik und Persönlichkeitseigenschaften der Studierenden konnte in einer

Untersuchung von Abrami et al. (1982) nicht gefunden werden. In zwei Laboruntersuchungen konnte kein Zusammenhang ermittelt werden zwischen dem manipulierten Lehrverhalten und studentischen Eigenschaften. Nur in der Feldstudie trat ein bedingter Zusammenhang zwischen studentischen Eigenschaften und studentischer Veranstaltungskritik auf.

Insgesamt kommen die Autoren zu dem Schluß, daß studentische Lehrbeurteilungen unabhängig von Persönlichkeitscharakteristika der Studierenden sind.

### **Studienerfolg und Ursachenzuschreibung**

Gelegentlich wird vermutet, daß Studierende mit gutem Studienerfolg (Noten) zu eher besseren Kursbewertungen kommen als Studierende mit schlechten Erfolgsergebnissen, da letztere den "Mißerfolg" eher dem Kurs zuschreiben dürften.

Snyder und Clair (1976) ermittelten an einer Stichprobe von 72 Studierenden, daß je besser die Note ausfiel, desto eher wurde internal attribuiert (eigene Leistung), je schlechter, desto eher kam es zu externalen Attributionen (äußere Faktoren).

Arkin und Maruyama (1979) ließen Studierende einen Fragebogen ausfüllen, der auch Skalen zur externalen/ internalen Attribution enthielt. Es konnte ermittelt werden, daß erfolgreiche Studierende eher internal attribuieren. Dagegen attribuieren weniger erfolgreiche stärker external als der Durchschnitt. Erfolgreichere Studierende machen im Vergleich zu weniger erfolgreichen nicht nur für sich, sondern auch für Durchschnittsstudierende stärker interne Gründe für Erfolg/ Mißerfolg verantwortlich und sehen dies auch etwas stabiler.

Darüberhinaus konnte festgestellt werden, daß die Selbstattribution weniger erfolgreicher Studierender nicht mit der Lehrbewertung korreliert, eine interne Selbstattribution aber positiv. Je stärker erfolgreiche Studierende internal attribuieren, desto besser bewerten sie die Lehrenden. Eine mögliche Erklärung geben die Autoren, indem sie einen zugrundeliegenden dritten Faktor vermuten.

Von Ames und Lau (1979) wurde ein Vergleich zwischen dem „Locus of Control“ und den Lehrbewertungen durchgeführt. Ein konsistent positiver Zusammenhang zwischen internaler Attribution und guter Bewertung und externaler Attribution und weniger guter Bewertung konnte ermittelt werden. Die Autoren folgern: "Die Ergebnisse zeigen eindeutig, daß attributionale Einschätzungen über die Gründe studentischer Leistung in einem Kurs signifikant mit der Lehrbewertung verbunden sind ... Der Bewertungsprozeß ist, wie auch immer, ein Wahrnehmungsvorgang, und die vorliegende Studie zeigt, daß attributionale Überzeugungen so maßgeblich bei der Lehrbewertung wie jeder objektive Unterschied zwischen Lehrenden sind" (Ames und Lau, 1979, S. 35).

Diese Ergebnisse decken sich z.T. mit der Studie von Arkin und Maruyama (1979), sind jedoch weitreichender, da in der vorherigen Studie keine Korrelation zwischen externaler Attribution und schlechteren Lehrergebnissen gefunden wurde. Dies kann daran liegen, daß bei Arkin und Maruyama der Erfolg bzw. Mißerfolg in die Bewertung einbezogen wurde.

Ames und Lau kommen somit zu einer Kausalattribution, die auch für die Diskussion um den Einfluß von Noten eine Bedeutung hat, indem sie die bessere Lehrbewertung von Studierenden nicht auf das bessere Ergebnis, sondern auf internale Attribution zurückführen. Ein anderes Erklärungsmuster bietet die unterschiedliche Leistungsmotivation, die auch mit unterschiedlichen Attributionsmustern zusammenhängen könnte.

### **Leistungsmotivation**

Mit seiner Untersuchung wollte Kovac (1976) ermitteln, inwieweit verschiedene Persönlichkeitsmerkmale die studentische Kursevaluation beeinflussen. Eine hohe Korrelation zwischen dem "Locus of Control" und dem "Leistungsanspruch" konnte festgestellt werden. Studierende mit einem hohen Leistungsanspruch bewerteten Inhalt und Struktur eines Kurses besser als Studierende mit einem geringeren Leistungsanspruch. Geht man davon aus, daß Studierende mit einem externalen Locus of Control einen geringeren Leistungsanspruch haben, so decken sich die Ergebnisse mit denen der vorherigen Studien.

Strom et.al. (1982) legten Studierenden Items zur Kursstruktur und Schwierigkeit vor. Der Fragebogen wurde mit den Ergebnissen eines anderen Fragebogens zu "Reizsuchverhalten", "Leistungsmotivation" und "Dogmatismus" korreliert. Die einzige signifikante Korrelation ergab, daß leistungsmotivierte Studierende weniger strukturierte und anspruchsvollere Kurse vorziehen.

Pamela J. Cooper et. al. (1982) ließen 240 Studierende einen Fragebogen beantworten und faßten die Fragebogen der verschiedenen Kurse für die jeweiligen Lehrenden zusammen. Im Vergleich zu den vorherigen Studien konnten keine Zusammenhänge zwischen der Lehrbewertung und dem Selbstkonzept bzw. der Leistungsmotivation der Studierenden festgestellt werden. Problematisch dabei ist eine sehr niedrige Rücklaufquote pro Kurs, die u.U. zu dem diskrepanten Ergebnis geführt haben könnte.

### **Lernstile und Vorlieben**

Man kann versuchen, Studierende nicht nur entsprechend ihrer Leistungsmotivation oder ihres Locus of Control einzuteilen, sondern z.B. auch nach unterschiedlichen kognitiven Stilen oder Lernstilen.

Meredith (1981) hat untersucht, inwieweit unterschiedliche "Lernerarten" die studentische Veranstaltungskritik beeinflussen. Sie gehen von zwei unterschiedlichen kognitiven Stilen aus: den "Focusern", die verstärkt Details und Fakten lernen, und den "Scannern", die sich eher auf das Begreifen von Konzepten und Prinzipien konzentrieren.

Die Ergebnisse veranlassen den Autor zu dem Schluß: "...das ist ein knapper Beweis, daß diese 'kognitiven Stile' in einem engen Zusammenhang mit der Lehr-/Kursbewertung stehen" (Meredith, 1981, S. 620). Dieser Punkt müßte allerdings mit solchen Themen wie Leistungsmotivation oder Studienmotivation gemeinsam untersucht werden.

In eine andere Richtung geht die Untersuchung von Drummond und McIntire (1977) zu "field dependent" Studierenden, die weniger strukturierte Kurse mit mehr Freiheitsgraden vorziehen, einer Mittelgruppe und "field independent" Studierenden, die das traditionelle System befürworteten. Zwischen den beiden Extremgruppen ergaben sich signifikante Unterschiede bei der Hälfte der Items zur Kursevaluation.

Allerdings muß bedacht werden, daß diese Items nahezu ausschließliche Vorlieben und Interessen der Studierenden abfragen. Es stellt sich somit die Frage nach der Generalisierbarkeit dieser Befunde. Denn es erscheint durchaus möglich, daß Studierende zwar unterschiedliche Gestaltungsvorlieben für einen Kurs haben, fraglich aber bleibt, inwieweit diese die Lehrbewertung beeinflussen.

Die Studie von Drummond und McIntire ist somit nur begrenzt dafür aussagekräftig, ob studentische Vorlieben die Lehrbewertung beeinflussen. Die von Meredith angesprochenen Zusammenhänge zu „kognitiven Stilen“ müßten dagegen näher untersucht werden, um Verallgemeinerungen zu ermöglichen.

## **Schlußfolgerungen**

*Eine allgemeine Schlußfolgerung aus vorliegenden Studien zu studentischen Lernstilen und Vorlieben erscheint schwierig, weil die Befunde uneinheitlich ausgefallen sind.*

*Es beeindruckt zwar, daß z.B. Studierende mit einer internalen Selbstattribution ihre Lehrenden möglicherweise besser bewerten als der Durchschnitt. Ob dies jedoch überhaupt in nennenswertem Grad geschieht, bleibt fraglich.*

*Die wenigen Studien zum Einfluß der Leistungsmotivation konnten keine prinzipiellen oder durchgängigen Unterschiede in der Lehrbewertung bedingt durch die Leistungsmotivation ermitteln.*

*Inwieweit unterschiedliche Lernstile und kognitive Stile der Studierenden für die Lehrbewertung eine Rolle spielen, läßt sich angesichts der Datenbasis und der vagen Befunde nicht abschließend beurteilen.*

*Spezifische Haltungen und Stile von Studierenden haben stets nur einen möglichen Einfluß auf einzelne Aspekte von Kursevaluationen, der zudem nicht sehr stark zu sein scheint. Eine generelle Verzerrung größeren Ausmaßes aufgrund unterschiedlicher Eigenschaften der Studierenden kann allerdings nicht angenommen werden.*

### **5.9 Neuntes Argument: Übereinstimmung von Lehrenden und Studierenden**

Aufgrund der Vermutung, daß Studierende jene Lehrenden mit einem ähnlichen Weltbild oder ähnlicher Persönlichkeitsstruktur bevorzugen, ist es sinnvoll zu überprüfen, inwieweit die Übereinstimmung von Lehrenden und Studierenden in Haltungen und Orientierungen die Bewertung der Lehre durch die Studierenden beeinflußt.

Eine derartige Untersuchung zur Kongruenz von Lehrenden und Studierenden und ihren Folgen für die studentische Veranstaltungskritik haben Hofman und Liya Kremer (1980) durchgeführt. Sie ließen elf Lehrende von 455 Studierenden bewerten, wobei zugleich deren jeweilige Einstellung zur höheren Bildung und zur Lehre ermittelt wurden. Anhand eines Ähnlichkeitsindex (SIMIL) wurde das Ausmaß an Kongruenz zwischen Lehrenden und Studierenden bestimmt. Damit ließen sich 17,7 Prozent der Kriteriumsvarianz dieser Lehrbewertung erklären. Autor und Autorin folgerten, daß die Gesamtbewertung sowohl von der studentischen Einstellung zu höherer Bildung als auch von der Kongruenz zwischen den Einstellungen von Studierenden und Lehrenden beeinflußt wird.

Der Kausalzusammenhang bleibt allerdings unklar: Entweder führt die Wahrnehmung guter Lehre dazu, daß Studierende davon ausgehen, daß die entsprechenden Lehrenden ihre bevorzugten Ansichten zur Lehre teilen; oder anders herum, daß Lehrende, von denen Studierende vermuten, sie würden ihre Ansichten teilen, bessere Bewertungen erhalten. Diese Studie ist in Zusammenhang mit der Diskussion über die mögliche Auswirkung von Meinungen über "gute Lehre" auf die tatsächliche Lehrbewertung mit einem bestimmten Instrument zu sehen.

Thomas, Ribich und Freie (1982) gingen von der Hypothese aus, daß die Identifikation mit dem/der Lehrenden sich sowohl auf die Lernmotivation als auch auf die Lehrbewertung auswirkt. Sie benutzten als Grundlage für ihre Studie ein klassisches psychoanalytisches Identifikationsmodell und kamen zu folgendem Ergebnis: Für fünf Faktoren (Kommunikation, Enthusiasmus, Organisation, Evalua-

tion und Lehreffektivität) wurden signifikante Werte erreicht. Eine Kausalbeziehung konnte aber nicht ermittelt werden. Demnach kann die Lehrbewertung entweder von einer Identifikation mit dem Lehrenden beeinflusst werden, oder die Identifikation führt zu einer höheren Lernmotivation und beeinflusst darüber die Lehrbewertung positiv.

Auch Morstain (1977) spricht von einer "schwachen Tendenz", daß Kurse mit einer übereinstimmenden "Orientierung" von Lehrenden und Studierenden besser bewertet werden.

Ein Teil der Befunde kann jedoch auch folgendermaßen betrachtet werden: Ausgehend davon, daß Studierende ein Verhalten wahrnehmen und die Einstellungen von Dozenten/Dozentinnen beurteilen, ist es wenig verwunderlich, daß Einstellungen, die sich auf die Lehre beziehen, hoch mit der Lehrbewertung korrelieren. Wenn Lehrende bestimmte Erwartungen der Studierenden an die Lehre erfüllen, dann ist es durchaus ersichtlich, daß dies zu einer positiveren Bewertung führt.

### **Schlußfolgerungen**

*Nach den vorliegenden Ergebnissen kann man nicht folgern, daß die Übereinstimmungen zwischen Lehrenden und Studierenden größere Auswirkungen auf die Lehrbewertung haben. Die vorhandenen Befunde zur "schwachen Tendenz" einer besseren Lehrbewertung bei Übereinstimmung in Haltungen und Orientierungen lassen sich auf unterschiedliche Weise verstehen und erklären.*

*Außerdem haben die vorhandenen Studien keinen direkten Zusammenhang ermitteln können. Zudem kann angenommen werden, daß die "Validität" der Instrumente dadurch nicht berührt wird.*

### **5.10 Zehntes Argument: Das Geschlecht beeinflusst die Bewertung**

Der Geschlechtseinfluß auf die Bewertung von Veranstaltungen und Lehrenden muß unter zwei Gesichtspunkten betrachtet werden: Macht es Unterschiede, ob ein Kurs von einem Mann oder einer Frau unterrichtet wird? Und: Werden Lehrende von Studenten anders bewertet als von Studentinnen?

### **Unterschiede in der Bewertung von Dozenten und Dozentinnen**

Patricia B. Elmore und Pohlmann (1978) konnten bei einer Untersuchung an 174 Kursen keine signifikant unterschiedlichen Bewertungen für Dozentinnen bzw. Dozenten ermitteln.

Die Autorinnen Linda Thurston und Glenn Littlepage (1979) führten zwei Experimente zu der Hypothese durch, daß Professorinnen niedrigere Bewertungen erhalten als Professoren, daß sie weniger populär sind und ihre Kurse weniger häufig gewählt werden. Diese Hypothese stützt sich auf die Vermutung, daß Professorinnen mehr leisten müßten, um die gleichen Bewertungen zu erhalten.

Es konnten jedoch keine Geschlechtsunterschiede gefunden werden. Insgesamt sahen die Autorinnen eine mögliche Erklärung für ihre Ergebnisse, die diskrepant zu den Ergebnissen anderer Studien stehen, in der größeren Genauigkeit und vor allem im Ausmaß an "outside definition of excellence", die in ihren beiden Experimenten durch die Beschreibung der wissenschaftlichen Leistungen der Lehrenden erreicht wurden.

Nach Durchsicht verschiedener Studien sind Bray und Howard (1980) der Ansicht, daß die Beziehung zwischen den Persönlichkeitseigenschaften von Studierenden und Lehrenden und der Lehreffektivität komplexer Natur sind. Faktoren wie Herzlichkeit, Interesse und Geschlecht wurden bereits in vielen anderen Studien untersucht. Bray und Howard gehen der Vermutung nach, daß dies nur verschiedene Aspekte eines globaleren Konstrukts, der "sex role orientation" (SRO), sind.

Aus dieser Studie sind folgende Ergebnisse zur Bewertung von „androgynen“, „femininen“ und „maskulinen“ Lehrenden festzuhalten:

1. Androgyne Lehrende wurden signifikant höher bewertet als maskuline Lehrende, die Studierenden waren mit ihnen auch zufriedener.
2. Insgesamt wurden androgyne Lehrende nicht signifikant höher bewertet als feminine. Allerdings berichteten Studierende von androgynen Lehrenden über geringfügig mehr Lernfortschritte.
3. Studierende, die von Hochschullehrerinnen unterrichtet wurden, die sich selbst als maskulin einstufen, gaben einen signifikant höheren Lernfortschritt an, als Studierenden, die von Hochschullehrern unterrichtet wurden, die sich als maskulin einstufen.
4. Von den Hochschullehrerinnen bekamen jene, die sich androgyn einstufen, die höchsten Zufriedenheitswerte.

### **Interaktion: Geschlecht - Humor**

Katherine van Giffen (1990) ermittelte, daß die Durchschnittswerte für die Lehreffektivität und für die Nutzung von Humor sowohl für Hochschullehrer und Hochschullehrerinnen als identisch betrachtet werden können. Weder ein Einfluß des Kursniveaus noch eine Wechselwirkung zwischen Kursniveau und Geschlecht konnte von ihr festgestellt werden.

### **Interaktion: Geschlecht - Herzlichkeit**

Bei den Auswertungen von Patricia B. Elmore und Karen A. LaPointe (1975) wurden keine signifikanten Zusammenhänge zwischen dem Geschlecht (Lehrenden und Studierenden) und der Herzlichkeit des/der Lehrenden festgestellt. Insgesamt scheinen die Ergebnisse weniger von der Selbstauskunft über die Herzlichkeit der Lehrenden als vom Geschlecht der Studierenden beeinflusst.

### **Unterschiede in der Lehrbewertung durch Studenten oder Studentinnen**

Patricia B. Elmore und Pohlmann (1978) fanden keinen globalen Geschlechtseinfluß der Studierenden auf die Lehrbewertungen, die sie abgeben.

Ashton (1975) fand ebenfalls keine Korrelationen zwischen der Einschätzung von Persönlichkeitseigenschaften und einem Fragebogen zur Lehrbewertung. Separat nach Geschlecht untersucht, erwiesen sich jedoch zwei Beziehungsebenen bei Studentinnen als signifikant mit der Lehrbewertung korrelierend: Konformität (-.54) und Unabhängigkeit (.48). Der Autor schließt daraus, daß zwischenmenschliche Werte höher mit der Lehrbewertung von Studentinnen in Beziehung stehen als daß sie die Bewertung von Studenten beeinflussen.

Dies würde sich auch mit den Ergebnissen von Daniel und Hornbostel (SPIEGEL-SPEZIAL, 1993, S. 157) decken. Einen Hinweis darauf, daß Studentinnen u.U. zu strengerer Bewertung neigen, hat Daniel gefunden. Er ermittelte in der zweiten SPIEGEL-Erhebung (DER SPIEGEL, 1994, S. 55), daß die Beurteilungen von Studentinnen signifikant schlechter als die ihrer Kommilitonen waren. Ob dies jedoch auch konkret bei der Beurteilung eines/einer bestimmten Lehrenden der Fall wäre, ist fraglich, da in der zweiten SPIEGEL-Erhebung nicht einzelne Lehrende bewertet wurden, sondern eher die Studienbedingungen, die von den Studentinnen als ungünstiger empfunden wurden als von Studenten.

Die Autorinnen Patricia B. Elmore und Karen A. LaPointe (1975) konnten weiter beobachten: Studentinnen vergaben bessere Werte bei dem Item "einzelne Ziele des Kurses" als Studenten. Ein Item ("zeigte Interesse an den Studierenden") wurde von Studentinnen signifikant höher für ihre lehrenden Geschlechtsgenossinnen eingeschätzt als für Hochschullehrer. Auch bewerteten Studentinnen ihre Hochschullehrerinnen insgesamt besser als Studenten. Dagegen wurden Hochschullehrer mit diesem Item von Studenten fast immer signifikant höher bewertet als von Studentinnen. Dies deutet auf geschlechtsspezifische Übereinstimmung bestimmter Aspekte hin.

Dieses Ergebnis deckt sich mit den Untersuchungen von Bray und Howard (1980), die eine signifikante Beziehung zwischen Geschlecht der Lehrenden und Geschlecht der Studierenden in Bezug auf den berichteten Lernfortschritt ermit-

telten. Dabei gaben Studentinnen einen höheren Lernfortschritt in Veranstaltungen an, die von Frauen gehalten wurden, im Vergleich zu den von Männern gehaltenen. Auch gaben Studentinnen im Vergleich zu Studenten einen höheren Lernfortschritt bei Hochschullehrerinnen an. Dies wäre eine Möglichkeit, die angeführten Befund von Elmore und LaPointe zu erklären: Geht man von Validierungsstudien aus, so wäre es schlüssig, daß Studierende, die einen höheren Lernfortschritt in bestimmten Kursen beobachten, diese auch höher bewerten - dies wären in diesem Fall Studentinnen in Kursen, die von Dozentinnen gehalten wurden.

## **Schlußfolgerungen**

*Nach den Studien zum Einfluß des Geschlechtes auf die Lehrevaluation ist zu vermuten, daß es keine direkt und prinzipiell unterschiedlichen Bewertungen zwischen Dozenten und Dozentinnen gibt (obwohl dieses Ergebnis kulturabhängig sein könnte), daß es aber möglicherweise zu Interaktionen zwischen Geschlecht und Persönlichkeitseigenschaften kommen kann.*

*Die untersuchten Interaktionen ließen jedoch keinen signifikanten Einfluß des Geschlechts der oder des Lehrenden auf die Ergebnisse der studentischen Veranstaltungskritik erkennen. Insgesamt ist somit zu vermuten, daß Dozenten und Dozentinnen keine signifikant unterschiedlichen Bewertungen allein aufgrund ihrer Geschlechtszugehörigkeit erhalten.*

*Die Untersuchungen über Studierende legen den Schluß nahe, daß Studentinnen nicht generell Hochschullehrer bzw. Hochschullehrerinnen anders bewerten als ihre männlichen Kommilitonen, sie aber einzelne Aspekte anders gewichten. Für genauere Aussagen müßten weitere Untersuchungen erfolgen. Da gerade Rollenbilder und -muster kulturabhängig sein können, wäre es notwendig, solche Untersuchungen zum Geschlechtseinfluß auf die Lehrbewertung auch in Deutschland, bzw. in vergleichbaren Ländern, durchzuführen, um verbindliche Aussagen treffen zu können.*

### **5.11 Elftes Argument: Noten (erwartete oder erhaltene) beeinflussen die Bewertung**

Noten sind wohl die am exzessivsten untersuchte mögliche Einflußvariable der studentischen Veranstaltungskritik. Die Angst vor der möglichen "Bestechlichkeit" der Studierenden ist auch in der Umfrage von Marsh (1982b) zu finden, in der mehr als zwei Drittel der befragten Lehrenden angaben, daß sie eine Verzerrung der Lehrbewertungen durch die Art der Notenvergabe vermuten. Diese Ängste können genauer beschrieben werden: Zum einen wird befürchtet, daß

strenger benotende Lehrende schlechter bewertet werden als mild bewertende Lehrende; zum anderen besteht die Angst, daß Studierende, die schlechter bewertet wurden, ihrerseits die Lehrenden schlechter bewerten.

### **Mild oder streng benotende Lehrende**

Um feststellen zu können, ob mild benotende Lehrende bessere Werte in der Lehrevaluation erhalten als streng bewertende, ist es zwingend notwendig, Kursunterschiede zu vergleichen und nicht innerhalb von Kursen zu korrelieren. Auf diese Art und Weise werden auch mögliche Interaktionen (z.B. zwischen Studienmotivation, Note und Lehrbewertung) vermieden. Bedauerlicherweise ist die Art der Korrelation häufig nicht angegeben.

Die Studie von Brown (1976) zeigte, daß die Durchschnittsnote eines Kurses der beste Prädiktor für die Bewertung des Lehrenden im Kurs war. Die Korrelation zwischen der Durchschnittsbewertung und der Durchschnittsnote betrug  $r = .35$ .

Auch in der Studie von Centra (1977) wird festgestellt, daß Korrelationen zwischen Leistungen und Gesamtlehrbewertung und in etwas geringerem Ausmaß zwischen Leistungen und Bewertung der Lehreffektivität sowie Qualität der empfohlenen Lektüre liegen. Am wenigsten korrelierte die Leistung mit der Kurschwierigkeit und der Arbeitsbelastung. Ein Problem liegt in der niedrigen Zahl der Studierenden in den Kursen, der Median lag bei acht Studierenden.

In der Studie von Patricia B. Elmore und Pohlmann (1978) wirkt von den untersuchten Variablen noch am stärksten die erwartete Note im Kurs auf die Lehrbewertung ein (positive Beziehung). Adriane Gaffuri et al. (1982) erhoben Daten über drei Trimester in 85 Kurse mit einem Fragebogen: Die Durchschnittskorrelation eines Items mit der Note betrug  $r = .35$ .

Die Studie von Chacko (1983) untersucht die Hypothese, daß Studierende, die eine schlechtere Bewertung in ihren "midterm examinations" erhielten, eher geneigt sind, die Effektivität der Lehrenden geringer zu bewerten, als Studierende, die keine ungünstigen Rückmeldungen erhalten haben. Zwar bewerten die Studierenden ihren Instruktor in einigen Teilbereichen tatsächlich schlechter als vor der Notenvergabe. Dies kann jedoch aufgrund des Untersuchungsdesigns auf unterschiedliche Gründe zurückzuführen sein.

### **Gut oder schlecht bewertete Studierende**

Viele Lehrende befürchten, Studierende, die eine schlechte oder eine schlechter als erwartete Note erhalten, dann ihrerseits es den Lehrenden "mit gleicher Münze heimzahlen" - und ihre Lehre schlechter beurteilen. Um zu ermitteln, ob weni-

ger gut bewertete Studierende ihrerseits den Dozenten/die Dozentin schlechter beurteilen, sind nicht die von den Kursteilnehmern und -teilnehmerinnen erwarteten, sondern nur die tatsächlich erhaltenen Noten heranzuziehen. Für eine Korrelation: erwartete schlechte Note - schlechte Lehrbewertung gäbe es andere plausible Erklärungen, z.B.: Der Student hat wenig verstanden, also schließt er daraus, daß der Lehrende den Stoff schlecht vermitteln kann.

Yates und Ksarmos (1971) überprüfen ihre Überlegung, ob Studierende ihre Bewertung ändern, nachdem sie ihre Noten erfahren haben. Studierende eines Kurses beantworteten zweimal innerhalb kürzester Zeit (ca. 1 Woche) den "Instructional Improvement Questionnaire", wobei ihnen der Zweck der zweimaligen Bewertung nicht bekannt war. Insgesamt kamen die Wissenschaftler zu dem Ergebnis, daß es zu einer sehr geringen, nicht signifikanten Verschlechterung der Bewertung kam. Ein Unterschied in der Art der Bewertung von gut vs. schlecht benoteten Studierenden konnte nicht gefunden werden.

Kritisch ist zu bedenken, daß die ähnliche Bewertung an der kurzen Zeit zwischen der ersten und zweiten Erhebung liegen könnte, was auch die Autoren als mögliche Begründung sehen. Es ist bedauerlich, daß bei dieser "quasi-experimentellen" Anordnung nicht eine möglichst adäquate Instruktion der Studierenden versucht wurde. Eine solche Untersuchung an einer größeren Zahl von Studierenden wäre eine gute Möglichkeit, dem vermuteten Zusammenhang zwischen Noten und Bewertung nachzugehen.

Snyder und Clair (1976) untersuchten den Noteneinfluß auf Lehrbewertungen und versuchen dabei Fähigkeits- und Leistungseinflüsse auszuschließen. Sie kamen zu folgenden Ergebnissen: 1. Je höher die erwartete Note, desto schlechter die Bewertung des Lehrenden. 2. Je höher die erhaltene Note, desto besser die Bewertung des Lehrenden. 3. Übertraf die erhaltene die erwartete Note, so gaben die Studierenden bessere Bewertungen ab. 4. Am schlechtesten fiel die Bewertung des Lehrenden dann aus, wenn die erhaltene Note die Erwartungen enttäuscht hatte.

Da diese Studie jedoch eher als Experiment denn als eine Untersuchung im natürlichen „Setting“ (Umfeld) gestaltet war, sind Verallgemeinerungen schwierig. Die Situation war an verschiedenen Punkten so künstlich angelegt, daß mögliche neue Einflüsse völlig ungeklärt sind. Es bleibt weiter untersuchenswert, warum erwartete gute Noten zu schlechterer Bewertung, erhaltene gute Noten jedoch zu besserer Bewertung führten.

Brandenburg, Slinde und Batista (1977) fanden signifikante Korrelationen zwischen der Bewertung des "Illinois Course Evaluation Questionnaire" (CEQ) und der erwarteten Note. Aleamoni und Hexner (1980) konnten dagegen keine nennenswerte Korrelation zwischen den Ergebnissen des CEQ, der vor dem Ab-

schlußexamen ausgefüllt wurde, und den späteren Noten ermitteln. In einer Untersuchung von Lester (1982) wurden ebenfalls keine signifikante Beziehung zwischen der durchschnittlichen Gesamteinschätzung des Lehrenden/ bzw. des Kurses und den Noten festgestellt.

Eine mögliche Erklärung für die Korrelation zwischen externen Variablen und der Kursbewertung könnte in der Art des Items bzw. der Fragebogenkonstruktion liegen. Dies wurde bereits von Marsh an den verschiedensten Punkten erwähnt. Eine Bestätigung dafür liefern auch Aleamoni und Thomas (1980), die in ihrer Untersuchung signifikant höhere Korrelationen zwischen den drei generellen Fragen zu Lehrenden, Kurs und Inhalt feststellen konnten als im Vergleich zu einzelnen Subskalen.

Nicht auszuschließen sind spezifische Interaktionen: Denkbar ist, daß nicht alle Studierenden sich im gleichen Maße von erhaltenen Noten beeinflussen lassen. Welche Persönlichkeitseigenschaften oder Denkstile sind dafür verantwortlich? Es liegen zwei Studien vor, die diesen Bereich untersucht haben.

Blass (1980) ermittelte in einem Kurs der Einführung in die Psychologie eine signifikant positive Korrelation zwischen sechs der insgesamt neun Items des Bewertungsbogens. Mit einem von ihm konstruierten Verfahren teilte er daraufhin den Kurs in relativ objektiv urteilende und relativ subjektiv bewertende Studierende. Die Korrelation zwischen Note und Lehrbewertung erwies sich bei den subjektiv bewertenden Studierenden als insgesamt deutlicher als für die Gesamtgruppe, während in der relativ objektiv urteilenden Kurshälfte nur noch Korrelationen für zwei der neun Items ermittelt werden konnten.

Meredith (1982) überprüfte, inwieweit die postulierte bipolare Dimension "Learning Orientation/ Grade Orientation (LOGO)" mit Noten zusammenhängen. Insgesamt schloß der Autor, daß die persönliche Wichtigkeit, die die Noten für einen Studierenden haben, sowohl die Bewertung der Lehrenden als auch die Kurszufriedenheit beeinflusst. Wobei zu vermuten ist, daß die Wichtigkeit der Note und das Leistungsmotiv eng miteinander verknüpft sind.

### **Erwartete Noten**

Überraschend viele Studien erheben auch erwartete Noten bzw. stellen es Studierenden frei, erwartete oder erhaltene Noten anzugeben. Werden diese zu Kursdurchschnitten zusammengefaßt, so können diese Werte verwendet werden, um zu überprüfen, ob als "milde bewertend" eingeschätzte Lehrende ihrerseits besser bewertet werden. Korreliert man jedoch die Note einzelner Studierender mit ihrer Lehrbewertung, so ist dies nicht möglich.

DuCette und Jane Kenney (1982) verwendeten die Daten von ca. 5.900 Studierenden, die einen Fragebogen zur Evaluation der Lehre beantwortet hatten. Einige Dimensionen der Lehrevaluation korrelieren positiv mit den erwarteten Noten, im extremsten Fall erklären sie jedoch weniger als 20 Prozent der Varianz. Die stärkste Korrelation wurde konsistent in der Dimension "Kurseffektivität" gefunden. Im Vergleich dazu korrelierte die Dimension "Lehreffektivität" weit aus geringer mit den erwarteten Noten.

Die Ergebnisse werteten DuCette und Jane Kenney als Beweis, daß ein "bias" als alleinige Erklärung für die Korrelation Note - Bewertung nicht genügt. Die Korrelation zwischen der zweiten Dimension und den Noten beurteilten sie als realistische Einschätzung der Studierenden. Jene, die im Kurs mehr gelernt hätten, erwarteten bessere Noten. Die Korrelation zwischen der ersten Dimension und der studentischen Veranstaltungskritik unterstützt eher eine "bias"-Erklärung.

Diese Studie macht deutlich, daß ein "bias" entweder nicht vorliegt bzw. in dieser Versuchsanordnung kaum zu ermitteln ist. Erhebt man erwartete Noten, so muß überlegt werden, was eigentlich damit erfaßt wird. Welche Faktoren führen dazu, daß Studierende eine bestimmte Note erwarten? Es erscheint wahrscheinlich, daß mit der erwarteten Note im Grunde genommen eine ganze Reihe anderer Faktoren erfaßt werden. Diese sind u.a. vermutlich Vorkenntnisse, Motivation, der Erfahrungswert, wie man sonst in solchen Prüfungen abschneidet u.ä.. Die Ergebnisse von DuCette und Jane Kenney sind daher am sinnvollsten mit Studien zur Motivation der Studierenden im Zusammenhang zu betrachten.

## **Schlußfolgerungen**

*Anhand der vorliegenden Studien über den Einfluß von Noten (erhaltene oder erwartete) auf die Lehrbewertung ist nur schwerlich zu folgern, daß strenger bewertende Lehrende häufiger schlechtere Ergebnisse erzielen, da zu vermuten ist, daß nicht alle Studien Werte **zwischen** Kursen korreliert haben. Außerdem ist es möglich, daß die ermittelten Zusammenhänge auf unterliegende Faktoren, wie z.B. unterschiedliche Motivation, zurückzuführen sind.*

*Aus den referierten Studien kann keine generelle Aussage abgeleitet werden. Es erscheint möglich, daß sich manche Studierende in einem begrenzten Ausmaß bei der Lehrbewertung von erhaltenen Noten beeinflussen lassen. Es besteht jedoch kein Nachweis, daß strenger bewertende Lehrende allein deshalb schlechtere Bewertungen erhalten.*

*Auch korrelieren erwartete oder erhaltene Noten in den Untersuchungen nur mit bestimmten Aspekten des Fragebogens zur Veranstaltungskritik. Um zu einem abschließenden Urteil zu kommen, scheinen zu dieser wichtigen und heiklen Problematik des Einflusses der Notenstrenge weitere Studien notwendig.*

## 5.12 Zusammenfassung und Folgerungen

Studien zur studentischen Veranstaltungskritik lassen sich verschiedenen Einwänden und Befürchtungen hinsichtlich der Validität von Bewertungen der Lehre durch Studierende zuordnen. Auf elf derartige Argumente wurde eingegangen, um zu klären, inwieweit sie berechtigt sind. Dabei handelt es sich einerseits um grundsätzlichere Vorbehalte, welche den Studierenden eine Urteilskompetenz überhaupt absprechen (z.B. Unerfahrenheit, mangelnde Sachkenntnis, falsche Kriterien), andererseits um Hinweise auf Faktoren, welche die studentischen Bewertungen verzerren oder beeinflussen könnten (wie Kursschwierigkeit, Motivation oder Notenstrenge).

Die gesichteten Studien und ihre Ergebnisse - insgesamt etwa 100 vor allem aus amerikanischen Hochschulen - lassen die Schlußfolgerung zu, daß es möglich ist, entsprechend gestaltete Fragebogen zur Bewertung der Lehrenden einzusetzen. Sie besitzen eine Validität, die es zuläßt, sie auch für den Vergleich von Lehrleistungen verschiedener Lehrender zu verwenden. Diese Bilanz sei nachfolgend für die verschiedenen Argumente und Faktoren zusammenfassend begründet und ergänzt.

Insgesamt kann gesagt werden, daß Studierende konsistente Urteile in der Lehrbewertung treffen können, die auch über Monate (oder Jahre) stabil bleiben. Von einer "mangelnden Distanz zum Bewertungsobjekt" (Aleamoni, 1987) kann daher nicht ausgegangen werden. Der Varianzanteil einer Bewertung, der auf den entsprechenden Kurs zurückzuführen ist, ist weitaus geringer als der Varianzanteil, der auf den entsprechenden Lehrenden zurückzuführen ist.

Der Verdacht, studentische Veranstaltungskritik komme eher einem bloßen Beliebtheitswettbewerb gleich, konnte nicht bestätigt werden. Es scheint möglich, in diesem Bereich Instrumente zu entwickeln, die reliabel und valide sind, und die Gefahr verzerrender "Halo"-Effekte minimieren oder zu kontrollieren erlauben.

Von der Einwirkung möglicher Einflußvariablen konnte keine direkte Ergebnisverzerrung aufgrund des Geschlechts der Lehrenden ermittelt werden. Es liegen jedoch Hinweise vor, daß Studentinnen einige Lehraspekte anders beurteilen als Studenten. Ebenfalls keine Auswirkungen haben Kursaufwand und -schwierigkeit. Die Angst der "Bestechlichkeit" von Studierenden erscheint somit unbegründet. Auch erhalten "milder" benotende Lehrende insgesamt keine besseren Bewertungen als "strenger" benotende. Keinen direkten Einfluß scheinen auch Persönlichkeitseigenschaften sowohl von Lehrenden als auch von Studierenden auf die Kursbewertung zu haben. Ein mittelbarer Einfluß über andere Faktoren ist jedoch möglich.

In den vorhandenen Studien wurden keine Einflüsse von Semesterzahl und Kursgröße gefunden. Diese Faktoren müßten jedoch in Deutschland unter den hiesi-

gen Bedingungen noch einmal genauer untersucht werden. Die wenigen deutschen Studien lassen vermuten, daß Vorlesungen insgesamt schlechter bewertet werden als Seminare.

Einen eindeutigen Einfluß hat der Faktor "Motivation" der Studierenden auf die Lehrbewertung. Dieser scheint jedoch kalkulierbar zu sein. Möglich ist auch, daß sich erhaltene Noten bei bestimmten Studierenden auf ihre spätere Lehrbewertung auswirken.

Zum Einfluß externer Faktoren auf die studentischen Bewertungen liegen wohl die meisten Studien vor: Es gibt wenig, von dem nicht vermutet wurde, daß es einen Einfluß auf die Lehrbewertung haben könnte. Bei dieser Art von Studien werden häufig entweder nur einzelne Aspekte untersucht oder aber es werden in großen Untersuchungen eine beträchtlichen Anzahl externer Variablen mit der Lehrbewertung korreliert.

Eine Ausnahme von dieser Herangehensweise ist die Studie von Scott (1977). Scott erhob Daten von 3.625 Studierende aus 195 Kursen (138 Lehrende), um anhand einer Liste häufiger Klagen von Lehrenden erst einmal zu ermitteln, welche Rahmenbedingungen u.U. zu einer niedrigeren Lehrbewertung führen könnten. Die genannten Schwierigkeiten waren: der Kurs wurde erstmalig unterrichtet, der Stoff sei zu schwierig, als daß man ihn adäquat im Kurs präsentieren könnte, die Klasse sei zu groß, der Kurs wurde nicht im eigenen Fachschwerpunkt unterrichtet, der Kurs war neu konzipiert, es wurden methodische Neuerungen eingeführt und eine inhaltliche Überarbeitung wurde versucht. Bei 83 Kursen wurde von Seiten der Lehrenden auf mindestens einen "mildernden Umstand" verwiesen. Korrelationsstudien ergaben, daß nur eine hohe Klassengröße einen signifikant (negativen) Einfluß auf die Durchschnittsbewertung des Kurses hatte. Der Autor sieht daher keinen statistischen Beweis für externe Einflußgrößen auf die Lehrbewertung.

An dieser Studie von Scott erscheint die Herangehensweise interessanter als das Ergebnis. Denn auf diese Art und Weise können Lehrbewertungen eine höhere soziale Validität erhalten und mögliche Einflüsse auf eine niedrigere Lehrbewertung aus der Sicht der Betroffenen gesammelt werden.

Es gibt die These, die von Marsh in verschiedenen Studien betont wird, daß sich der Einfluß externer Variablen nur auf bestimmte Teile eines Fragebogens beschränkt, nämlich auf diejenigen, die inhaltlich mit dieser Variablen zusammenhängen. Dieser Punkt soll an dieser Stelle noch einmal aufgegriffen werden.

Patricia B. Elmore und Pohlmann (1978) ermittelten, daß Items, die am besten aufgrund von Eigenschaften der Lehrenden und Studierenden vorhergesagt werden konnten, alle auf einer "Beziehungsdimension" lagen. Daß andere Dimen-

sionen der Lehrbewertung weniger oder gar nicht betroffen waren, werteten die Autorin und der Autor als Hinweis auf die diskriminante Validität.

Um konkretere Aussagen zu dem Einfluß "externer Faktoren" zu treffen, sollten künftige Analysen nicht nur global ausgelegt sein. Notwendig wäre, daß mit Fragebogen, die bereits als reliabel und valide gelten und multidimensional gestaltet sind, detailliert versucht wird zu ermitteln, welche Faktoren welche Fragebogenteile beeinflussen. Dabei ist es wichtig, nicht nur mit Vergleichen innerhalb der Kurse, sondern auch mit Kursdurchschnitten zu arbeiten, da damit viele "response biases" herausfallen. Zudem sind es Durchschnitte, Verteilungen und Profilvergleiche, die üblicherweise als Grundlage der Lehrbewertung verwandt werden.

Allerdings beinhaltet der Methodenstreit, ob eine vergleichende Lehrbewertung auf der Grundlage studentischer Urteile möglich ist oder nicht, auch die Frage nach dem politischen Willen, solche Instrumente einsetzen zu wollen. Die Frage nach der methodischen Validität kann auch als "Nebenkriegsschauplatz" von Abwehr und Verweigerung betrachtet werden, vor allem wenn die Anforderungen so hoch geschraubt werden, wie sie bei der Beurteilung von Studierenden durch Lehrende diesen nicht abverlangt werden.

Ähnlich sieht es auch Aleamoni: "Was wäre, wenn der Spieß umgedreht werden würde und die Bewertung der Studierenden in Zweifel gezogen würde? ... Wie viele Nachweise könnten wir vorzeigen, um die Studierenden zu überzeugen, daß die Beurteilungen ihrer Leistungen ausschließlich auf objektiven Kriterien beruhen? ... Wie viele Nachweise könnten wir vorzeigen, um die Studierenden zu überzeugen, daß Bewertungen ihrer Leistungen im Kurs sowohl valide als auch reliabel sind?" (Aleamoni 1987, S. 29-30).

Eine angemessene Erfassung von Lehrleistungen durch studentische Veranstaltungskritik ist mit mehrdimensionalen Fragebogen, welche die relevanten Sachverhalte ansprechen, möglich. In den Vereinigten Staaten und anderen Ländern, in den letzten Jahren vermehrt auch in Deutschland, sind derartige Instrumente zur Evaluation von Lehrveranstaltungen und für Lehrberichte entwickelt und geprüft worden. Sie genügen Ansprüchen an die Validität und Reliabilität von Messungen im sozialwissenschaftlichen Bereich, vermögen sie zum Teil sogar zu übertreffen. Deshalb ist ihre Verwendung im Rahmen der vergleichenden Lehr-evaluation und Lehrberichterstattung berechtigt. Einwände können sich jedenfalls nicht auf die unzureichende Validität studentischer Urteile und Veranstaltungskritiken berufen. Allerdings ist vorauszusetzen, daß in die Güte der Instrumente zur Lehr-evaluation, soll es sich um mehr als bloße Rückmeldungen zum internen Gebrauch handeln, einiges an Entwicklungsarbeit zu investieren ist.

## **6 Prinzipien der Entwicklung von Instrumenten zur Lehrevaluation**

Unter den Instrumenten zur Lehrevaluation sind sehr unterschiedliche Angebote und Verfahren zu finden. Dies können einerseits sehr elaborierte Instrumente sein, die sorgsam entwickelt wurden, andererseits Fragebogen, die fast beliebig zusammengestellt erscheinen. Es gibt sowohl spezielle Instrumente, die nur für bestimmte Fächer oder nur für bestimmte Kursarten gedacht sind, als auch Instrumente, die bereits im Routineverfahren eingesetzt werden.

Instrumente studentischer Veranstaltungskritik unterscheiden sich in Hinblick auf ihr Einsatzfeld, ihre Qualität und ihren Umfang. Auch in Deutschland existieren die verschiedensten Verfahren, von detailliert überprüften Instrumenten bis hin zu kurzfristig zusammengestellten Fragepaketen (vgl. die Zusammenstellung von Reissert, 1992).

### **6.2 Zwecke und unterschiedliche Anforderungen der Lehrevaluation**

Bei der Planung von Projekten zur Lehrevaluation hat man daher die Wahl, entweder auf ein vorhandenes Instrument zurückzugreifen, ein solches Instrument entsprechend den Bedürfnissen abzuwandeln oder aber einen eigenen Fragebogen zu entwickeln.

Bei der Frage, für welche Form man sich entscheidet, kommt es vor allem darauf an, zu welchen Zwecken die gesammelten Daten verwendet werden sollen. Wird nur ein Instrument benötigt, um Lehrenden eine kurze Rückmeldung zu geben, kann selbst ein einfaches Instrument mit relativ globalen Fragen nützlich sein. Ein solches Instrument sollte dann in erster Linie dazu dienen, eine breitere Datengrundlage für Gespräche über die Lehrqualität im Kurs bereitzustellen.

Wird jedoch angestrebt, die Lehrqualität in einem Fachbereich insgesamt zu erfassen und dafür Evaluationsdaten zu gewinnen, sollen Lehrende untereinander verglichen oder didaktische Maßnahmen geplant werden, so sind andere, anspruchsvollere Instrumente erforderlich. Für jede über das direkte "Kursfeedback" hinausgehende Absicht ist es sinnvoll, komplexere Instrumente zu verwenden.

Eine sorgfältige Fragebogenkonstruktion ist besonders dann zwingend, wenn ein Instrument zum Vergleich zwischen Lehrenden und ihren Lehrleistungen dienen soll, oder wenn die Befunde nicht nur zu einer besseren Kommunikation über die Lehrqualität führen sollen, sondern auch reale (z.B. finanzielle) Konsequenzen haben können.

Um den notwendigen Gütestandard zu erreichen, der vorauszusetzen ist, um Lehrende miteinander zu vergleichen, sind detailliertere und multidimensionale Fragebogen unerlässlich. Diese Instrumente sind weniger anfällig für mögliche externe Einflüsse. Fragebogen mit relativ wenigen globalen Items erreichen nicht das erforderliche Niveau an Zuverlässigkeit, Meßgenauigkeit und Objektivität. Wichtig ist deshalb bereits vor der Lehrbewertung zu klären und zu entscheiden, für welchen Zweck die Daten benötigt werden und wie sie verwendet werden sollen.

In den Vereinigten Staaten, die schon eine lange Tradition in der Lehrbewertung haben, existieren die unterschiedlichsten Untersuchungen zur Frage nach optimalen Kriterien für die Fragebogen studentischer Veranstaltungskritik. Im folgenden werden Ergebnisse derartiger Untersuchungen (vor allem aus den USA) vorgestellt. Daraus werden Kriterien abgeleitet, die für die Konstruktion von Fragebogen relevant sind, die einen weiterreichenden Anspruch haben, als Lehrenden eine kurze Rückmeldung zu geben, Fragebogen folglich, die eine möglichst hohe Qualität aufweisen sollen, um auch Veranstaltungen und Lehrende miteinander vergleichen zu können (vgl. Kapitel 7).

Vorab ist zu betonen, daß es sich um Konstruktionskriterien für Fragebogen zur studentischen Veranstaltungskritik handelt. Beurteilt wird mit ihnen nur die jeweilige Veranstaltung oder Veranstaltungsreihe. Aspekte der Lehre und der Lehrorganisation, die den Fachbereich oder die gesamte Hochschule betreffen, werden nicht weiter diskutiert. Zu vermuten ist, daß für diese Form der Evaluation zum Teil andere Kriterien eine Rolle spielen, daß z.B. für die Beurteilung der Lehrorganisation weitaus stärker auf bilanzierende und wertende (im Vergleich zu beobachtbaren) Daten zurückgegriffen werden kann als dies bei der Beurteilung einzelner Lehrveranstaltungen der Fall ist.

Die folgenden Ausführungen integrieren jeweils thematisch zusammenhängende Untersuchungsergebnisse und umfassen auch Schlußfolgerungen für die Entwicklung entsprechender Fragebogen. Der Aufbau wurde so gewählt, daß die einzelnen Aspekte bei der Konstruktion von Instrumenten studentischer Veranstaltungskritik nach ihrer Wichtigkeit aufgeführt sind.

Die *Auswahl und Abgrenzung verschiedener Dimensionen* stellen dementsprechend den wichtigsten "Merkposten" bei der Entwicklung dieser Verfahren dar. Die Multidimensionalität der Erfassungsmethode ist wohl der Grundstein für eine möglichst hohe Güte des Verfahrens und die Gültigkeit der Befunde.

Ein weiterer wichtiger Punkt, der die Qualität eines Fragebogens (vor allem seine Meßgenauigkeit) beeinflussen kann, ist die Klärung des *Vergleichsmaßstabs*, den die Beurteilenden bei der Beantwortung anlegen.

Dieser Vergleichsmaßstab ist bei der *Formulierung der Fragen und Vorgaben*, d.h. bei der Formulierung der sogenannten „Items“, zu beachten.

Im nächsten Kapitel werden Fragen behandelt, die stärker ins Detail gehen und den *Gesamtbogen*, die *Vergleichstabellen* und die *Ergebnisdarstellung* betreffen. Ein weiterer Abschnitt ist dem *Zeitpunkt und der Frequenz von Datenerhebungen* gewidmet.

Am Ende von Kapitel sieben ist eine *zusammenfassende Übersicht* zu finden, die alle genannten Kriterien im Sinne eines „Merkblattes“ auflistet, die für die Entwicklung, Gestaltung und Anwendung von Instrumenten der Lehrevaluation durch Studierende aufgrund vorliegender Untersuchungen und Erhebungen wichtig sind.

## **6.2 Zur Notwendigkeit multidimensionaler Lehrbewertungen**

Die Lehre ist ein vielschichtiger Vorgang mit vielen Aspekten und Facetten. Um sie möglichst effizient zu erfassen, ist daher ein multidimensionales Instrument notwendig. Aus diesem Grund werden nachfolgend Ansätze zur Ermittlung entsprechender Dimensionen vorgestellt und ihre Befunde diskutiert.

### **Komplexität von Lehrleistungen**

Angesichts der Komplexität von Lehrleistungen ist als erstes zu klären, welche Dimensionen einer Veranstaltung bzw. des oder der Lehrenden erfaßt und abgebildet werden sollen.

Marsh und Hocevar (1984) führten eine große Untersuchung (31.322 Studierende) zum Instrument „Students' Evaluation of Educational Quality“ (SEEQ) durch. Über den Vergleich von konvergenter und diskriminanter Validität zwischen den Faktoren versuchten sie, die Multidimensionalität der Lehre bzw. ihre Wahrnehmung durch die Studierenden zu bestätigen. Die konvergente Validität erwies sich als substantiell ( $r=.68$ ), die diskriminante lag weit niedriger.

Die neun konzipierten Faktoren des SEEQ konnten bestätigt werden. Sie sind als Dimensionen der Lehrqualität zu verstehen. Dabei handelt es sich um die folgenden Dimensionen:

- Wert der erworbenen Kenntnisse (Learning/ Value),
- Enthusiasmus (Enthusiasm),
- Kursorganisation (Organization),
- Verhalten zur Gruppe (Group Interaction),

- Zuwendung (Individual Rapport),
- Ermutigung (Breadth of Coverage),
- Fairness der Tests und Examina (Examinations),
- Fairness der Notenvergabe (Assignments),
- Kursschwierigkeit (Workload, Difficulty)

Dennoch werden die meisten Fragebogen zur Lehrevaluation nicht in diesem Sinne konzipiert bzw. ausgewertet (vgl. auch Scott, 1977, S. 746). Sie vernachlässigen die Dimensionalität der Lehre insgesamt oder lassen einzelne Dimensionen unberücksichtigt.

### **Globale Erfassung: Verzerrungen und Informationsverlust**

Problematisch an einer globalen Fragebogenkonstruktion für studentischen Veranstaltungskritiken ist die weitaus *höhere Anfälligkeit für Verzerrungen* durch äußere Einflüsse.

Eine ganze Reihe von Faktoren steht immer wieder in Verdacht, die Lehrbewertung unsachgemäß zu beeinflussen (z.B. Geschlecht, rhetorische Fähigkeiten des/der Lehrenden). Nach dem bisherigen Forschungsstand ist davon auszugehen, daß relativ wenig verzerrende Auswirkungen externer Einflüsse nachzuweisen sind. Diese Einflußgrößen können, falls sie empirisch bestätigt werden, meist einkalkuliert und die Ergebnisse entsprechend ausgewiesen werden.

Es gibt demnach zwei Möglichkeiten, den Einfluß externer, möglicherweise verzerrender Faktoren auf die studentische Veranstaltungskritik zu minimieren: zum einen die multidimensionale Gestaltung der Fragebogen und zum anderen die Bereinigung gewonnener Ergebnisse. Letzteres ist, nach dem bisherigen Stand, hauptsächlich für den Faktor „Motivation und Interesse am Kursthema“ notwendig. Der Einfluß aller anderen Faktoren ist entweder vernachlässigbar gering oder durch die mehrdimensionale Gestaltung der Erhebungsinstrumente aufzufangen.

Ein weiterer Nachteil einer globalen Erfassung und Interpretation von Lehrleistungen (z.B. durch Bildung eines Durchschnittswertes) ist der erhebliche *Informationsverlust*, der dadurch hingenommen werden muß. Beide Punkte, Anfälligkeit für Verzerrungen und Informationsqualität, werden im folgenden diskutiert.

Aleamoni und Thomas (1980) gehen sehr differenziert der Frage nach, welche Unterschiede zwischen globalen und differenzierten Vorgaben bestehen. Die beiden Autoren untersuchten, ob: a) studentische Charakteristika höher mit generellen als mit speziellen Items korrelieren, b) die Beantwortung der generellen Items ein guter Prädiktor für die Beantwortung spezieller Items ist, und c) spezi-

elle Items bessere Prädiktoren der Bewertung der Lehrenden für studentische Leistungen und Verhalten in einem Kurs sind als generelle Items. Insgesamt korrelierten die Autoren Daten von 7.242 beantworteten Bogen des Illinois Course Evaluation Questionares (CEQ).

Ergebnisse: Für alle externen Variablen mit Ausnahme von Kursgröße und Dozentenrang war die Korrelation mit generellen Items signifikant höher (1 Prozent Niveau) als mit speziellen Items. Keine der 56 Korrelationen mit den Subskalen des CEQ war größer als .19 (.013 -.18). Die Korrelationen mit den generellen Items betrug dabei  $r = .052$  -.37. Die Korrelation zwischen der Fähigkeitsbewertung durch Lehrende und des CEQ war für die Subskalen in allen Fällen signifikant höher als für die generellen Items. Die Korrelationen zwischen generellen Items und Subskalen lagen zwischen .29-.49.

Die Schlußfolgerung von Aleamoni und Thomas (1980) ist eindeutig: Die Korrelationen zwischen externen Variablen und Lehrrevaluation ist durch die Fragebogenkonstruktion bedingt. Lehrbewertungen sollten daher mit Subskalen arbeiten und nicht auf generellen Items beruhen.

### **Folgerungen für Instrumente und Auswertungen**

Diese Folgerung, die in der Literatur immer wieder bestätigt wird (z.B. durch das Auffinden verschiedener Faktoren in der Lehrbewertung), ist bei der Konzeption von Fragebogen und ihrer späteren Auswertung zu berücksichtigen. Vergrößert lassen sich die Ergebnisse von Aleamoni und Thomas wie folgt zusammenfassen: Benutzt man globale Beurteilungen für Kurs und Lehrende, dann können in viel stärkerem Maße "Beliebtheit" oder allgemeine Akzeptanz des Lehrenden erfaßt werden, als dies bei spezielleren Beurteilungen einzelner Lehraspekte der Fall wäre.

Die Problematik des Einflusses externer Variablen auf die Lehrbewertung versucht auch Marsh (1982b) in das Bewußtsein der Anwendenden solcher Instrumente zu bringen, indem er in allen seinen Artikeln betont: "Vieles der Vieldeutigkeit in den Untersuchungen zur studentischen Lehrbewertung stammt von der unglückseligen Praxis anzunehmen, daß studentische Bewertungen eindimensional seien" (Marsh, 1982b, S. 278). Außer Marsh sind auch eine große Anzahl anderer Autoren und Autorinnen der gleichen Ansicht, z.B. DuCette und Jane Kenney (1982).

Bei der Interpretation der Ergebnisse ist ebenfalls eine differenzierte Herangehensweise zu beachten. Es ist weder methodisch zulässig, noch inhaltlich aufschlußreich, Durchschnittswerte über Dimensionen zu bilden, die wenig miteinander korrelieren sollen, d.h. unterschiedliche Sachverhalte darstellen.

Ein wesentliches Ziel studentischer Veranstaltungskritik liegt in der Verbesserung der Lehre und der Qualifizierung der Lehrenden. Um eine solche Verbesserung und Qualifizierung erreichen zu können, muß den Dozierenden jedoch eine differenzierte Rückmeldung gegeben werden. Dies ist bei globalen Informationen nicht der Fall, weil offen gelassen wird, in welche Richtung die Verbesserung gehen sollte. Zwar wird in globalen Fragebogen oft versucht, dieses Manko durch die nachdrückliche Aufforderung zu ergänzenden Kommentaren der Studierenden zu mildern, diese sind jedoch aufwendig auszuwerten und können eine quantifizierbare Rückmeldung nicht ersetzen.

Aus den vorliegenden Ergebnissen kann die Schlußfolgerung gezogen werden, daß die Wahrung der Multidimensionalität eine der wichtigsten Voraussetzungen für ein objektives und informatives Instrument der Lehrbewertung ist. Sie ist deshalb bei der Entwicklung des Instruments wie bei der Auswertung der erhobenen Daten gleichermaßen zu beachten.

### **6.3 Auswahl und Abgrenzung verschiedener Dimensionen**

Nach der Feststellung, daß Lehre multidimensionale Aspekte hat, stellt sich die Frage: Welche Aspekte können ermittelt werden? Dabei muß zum einen nach den Verfahren gefragt werden: Wie können solche Dimensionen ermittelt werden? Zum anderen nach der Übereinstimmung bisher in empirischen Untersuchungen gefundener Dimensionen: Wie können sie miteinander verglichen und aufeinander bezogen werden?

#### **Wie können Dimensionen der Lehrbewertungen ermittelt werden?**

Grundsätzlich ließe sich bei der Ermittlung von Lehrdimensionen eine induktive und eine deduktive Vorgehensweise unterscheiden. Eine *deduktive Herangehensweise* ist in diesem Fall kaum möglich, da, wie Schott bereits 1973 anmerkte, keine "Theorie des Unterrichts" existiert. In den letzten zehn Jahren hat sich in diesem Punkt zwar einiges entwickelt, so wurde z.B. die kognitive Informationsverarbeitung aufgegriffen und im schulischen Lernen untersucht. Dennoch ist man in der Pädagogischen Psychologie von einer solchen "Theorie des Unterrichts" noch weit entfernt.

Weinert (1986) z.B. erhebt die Forderung: "So gesehen erschiene es ohne Nutzen, den Gegenstandsbereich der Pädagogischen Psychologie gegenüber der allgemeinen Lernforschung abzugrenzen. Was wir brauchen ist vielmehr eine theoretisch verzahnte Lern- und Instruktionspsychologie! Diese Forderung ist erst in Ansätzen erfüllt..." (S. 784). Als Grund für diesen Mangel sieht Weinert die "dis-

funktionale Arbeitsteilung" zwischen dem Auffinden elementarer psychischer Lernmechanismen und der angewandten pädagogischen Forschung.

Eine weitere Ursache für eine fehlende theoretische Grundlage im Bereich der Lehre spricht Heckhausen an: Die Bindung der Pädagogischen Psychologie an die Institution Schule und die "üblichen Schüler-Altersstufen" (Heckhausen, 1986, S.787). Um eine "Theorie des Unterrichts" an Hochschulen zu entwickeln, sind mit Sicherheit bereits genügend Grundlagen vorhanden. Eine solche Theorie existiert jedoch noch nicht. Folglich können Items zwar aus verschiedenen Ansätzen "herausgefiltert", gleichwohl nicht systematisch abgeleitet werden.

In allen vorliegenden Untersuchungen liegt daher überwiegend eine *induktive Herangehensweise* vor. Verwandt wurden z.B. Flanagans "Critical incidents-technique", "manuelle" Klusterung von Dozenten-/ Dozentinnenbeschreibungen zu homogenen Kategorien oder multidimensionale Skalierung. Es wurden auch Sekundäranalysen bereits gefundener Faktoren durchgeführt und diese zu gemeinsamen Dimensionen zusammengeführt.

Dabei muß bedacht werden, daß die meisten gefundenen Dimensionen Ergebnisse von Faktorenanalysen an entwickelten oder bestehenden Fragebogen sind. Es ist also theoretisch möglich, daß die "im Umlauf befindlichen" Items nur ganz bestimmte Arten von Dimensionen erfassen und andere dabei vernachlässigen. Um so mehr verdeutlicht dies die Notwendigkeit, pädagogische, psychologische und andere Theorien auf die Hochschullehre anzuwenden. Dies wäre auch dienlich, um das Mißtrauen von professoraler Seite gegen solche Instrumente abzubauen.

### **Einige theoretische Überlegungen zum 'Unterricht' an Hochschulen**

Bereits in der Studie von Müller-Wolf und Fittkau (1971) ist erkennbar, daß die Autoren sich nicht nur für das Aufdecken von Lehrdimensionen interessieren, sondern auch für ihre Wirkung. "Der deutlichste gefundene Zusammenhang besteht zwischen einem aktiven, dynamischen, anregenden, somit 'didaktisch effektiven' Lehrverhalten (...) und einer 'zufriedenen Stimulierung' der Studenten im Rahmen ihres Studiums (...):  $r=.67$ . Doch in fast demselben Maße hängt die Zufriedenheit der Studenten mit der Lehrveranstaltung, ihre Stimulierung und ihr Gefühl der Attraktivität des Fachgebiets (...) von dem 'emotional positiven, demokratischen Lehrverhalten' des Hochschullehrers ab:  $r=.63$ " (S. 176). Die gefundenen Interaktionen können kaum schon als Theorieansätze gewertet werden. Dennoch sind sie als ein Auftakt dieser Forschung in Deutschland zu sehen.

Erste theoretische Überlegungen zur Hochschullehre hat in Deutschland Schott (1973) formuliert. Er unterscheidet dabei einen informationspsychologischen und einen motivationspsychologischen Gesichtspunkt. Zu den informationspsycho-

logischen Aspekten gehören dabei z.B. Informationsvermittlung einerseits und Informationsaufnahme und Informationsverarbeitung andererseits. Zu den motivationspsychologischen Aspekt werden dagegen vor allem motivierende Handlungen des/ der Lehrenden gefaßt und ihre Wirkung auf die Studierenden.

Schotts Ableitungen aus der Allgemeinen Psychologie und der Motivationspsychologie und deren Anwendung auf die Lehre sind zwar interessant, führen aber nicht zu konkreten Ableitungen für die Lehre und bleiben, trotz ihrer Detailliertheit, abstrakt.

Dennoch sind die Schlußfolgerungen von Schott erste Ansätze für eine Theorie des 'Hochschulunterrichts'. Ob es allerdings sinnvoll ist, in diesem Sinne weiter zu forschen, auch mit der ganz speziellen Zielsetzung der Wissensvermittlung an Hochschulen, bleibt noch zu diskutieren. Zwar könnten Fragebogen der studentischen Veranstaltungskritik dann aus entsprechenden Theorien abgeleitet werden. Es stellt sich jedoch die Frage, inwieweit diese sich von den vorhandenen Fragebogen unterscheiden würden.

### **Welche Dimensionen der Lehrbewertung sind zu beachten?**

Um eine Aussage zu den Dimensionen von Fragebogen machen zu können, wurde in drei Schritten vorgegangen. Im ersten Schritt wurden die gefundenen Dimensionen gemeinsam mit mehr oder weniger namensgleichen Dimensionen aus bisher genannten Studien zusammengestellt.

Im zweiten Schritt wurden Studien hinzugezogen, deren Struktur mit den anderen nicht vergleichbar ist, die gefundenen Faktoren nicht kompatibel sind oder deren Dimensionen nicht auf einer Faktorenanalyse beruhen (z.B. Lita Linzer Schwarz, 1980). Im dritten Schritt wurden insgesamt ähnliche Dimensionen begründet zusammengefaßt.

Der Vergleich wurde über die Benennung der Dimensionen durchgeführt, was streng genommen nur als eine Nebeneinanderstellung gelten kann. Um die Dimensionen, die z.T. mit identischen Begriffen übertitelt sind, wirklich vergleichen zu können, müßte man die einzelnen Items, die zu der entsprechenden Dimension zusammengefaßt wurden, vergleichen bzw. die Faktoren, die sich bei der gleichen Stichprobe in den verschiedenen Instrumenten ergeben, miteinander korrelieren.

Da es sich um Aspekte der Lehre handelt, die leichter beobachtbar und abgrenzbar sind als viele andere psychologische Konstrukte (z.B. Extraversion), läßt sich vermuten, daß die gefundenen und zusammengefaßten Dimensionen, die gleichartig benannt wurden, im großen und ganzen auch vergleichbar in ihrem Inhalt wie in Ihren Vorgaben sind.

Klar zu trennen sind als erstes zwei große Bereiche: "... die Faktorenanalyse zeigt, daß Studierende zwischen Geselligkeit und Qualifikationsaspekten der Lehre unterscheiden können" (Beatty und Zahn, 1990, S. 281). Während eine dieser beiden übergeordneten Bereiche verhältnismäßig einfach zu der Dimension "Zuwendung" zusammengezogen werden kann, sind für die Beschreibung des zweiten übergeordneten Bereiches - die „Qualifikationsaspekte der Lehre“ - eine ganze Reihe von Dimensionen (bzw. Faktoren) notwendig.

Nachfolgend werden jene acht Dimensionen der Lehrbewertung aufgeführt und erläutert, die aus verschiedenen Studien zusammengestellt wurden und als typisch gelten dürfen. Bei Instrumenten zur Lehrevaluation wären sie jeweils mit getrennten Fragen oder Items anzusprechen, und in den Analysen oder Auswertungen wären sie gesondert auszuweisen.

### **1. "Zuwendung"**

Die Dimension der Zuwendung umfaßt vielfältige Facetten: Freundlichkeit, achtet auf Studierende, freundlich und aufgeschlossen sein, diskussionsbereit, Studierende achten/ nicht herabsetzen, Sensibilität für Reaktionen, demokratisch sein, Toleranz für einen anderen Blickwinkel, Zugänglichkeit für Kritik, emotionales Klima im Kurs, Respekt, macht Mut, gleichberechtigter Austausch, Erreichbarkeit.

In diese Dimension wurden alle Faktoren eingeordnet, die eher auf die "menschliche" Ebene der Beziehungen zwischen Lehrenden und Studierenden in Veranstaltungen bezogen werden können.

Eine Zwitterrolle nimmt das Element "Diskussionsbereitschaft" ein, da es je nach Itemzusammensetzung und Formulierung eher als Merkmal positiven Verhaltens den Studierenden gegenüber oder aber als didaktisches Mittel angesehen werden kann.

### **2. "Fairneß von Prüfungen und Benotungen"**

Diese Dimension könnte im allgemeinen schon in der ersten miteinbegriffen werden. Da jedoch Examen und Noten in der Hochschule eine große Rolle spielen, gibt es genügend Rechtfertigungen, dies als eigene Dimension zu betrachten.

Unter Umständen können Fragen zur Prüfungsrelevanz des Stoffes in ein Evaluationsinstrument mit aufgenommen werden, wenn dies als ein Maßstab für die Fairneß von Prüfungen definiert ist.

### **3. "Kommunikationsfähigkeit"**

Abstrakte und komplexe Ideen und Theorien darstellen können, verständlich darstellen können, gut erklären können, Tempo der Wissensvermittlung, Voraussetzungsadäquatheit, Erklärungsaktivität, Prägnanz, Verwendung von Hilfsmitteln, akustische Verständlichkeit und Zusammenfassungen geben.

Die genannten Aspekte machen eine Dimension deutlich, die vor allem in der Fähigkeit der Stoffvermittlung und ihrer Voraussetzungen (z.B. Umgang mit Hilfsmitteln) liegt.

### **4. "Kurs- bzw. Stofforganisation"**

Schwerpunktsetzung, Systematik, Zeitökonomie, Koordination und Vororientierung. Diese gut abzugrenzende Dimension ist in den meisten Fragebogen zu finden.

Mit der letzten, vierten Dimension wären die tendenziell leicht abzubegrenzenden Bereiche genannt. Dabei dürfen gerade für die Dimensionen drei und vier die notwendigen Voraussetzungen weder vergessen, noch als eigenständige Faktoren gesehen werden.

Diese Voraussetzungen sind: Fachwissen und Vorbereitung. Ohne ein elementares Fachwissen und eine grundlegende Vorbereitung werden keine guten Ergebnisse in den letzten beiden Dimensionen möglich sein. Die genannten Grundlagen sind aber nicht direkt beobachtbar und können nur abgeleitet werden. Sie sollten daher auch nicht abgefragt werden, wenn es um die Beurteilung der tatsächlichen Lehrleistung in Veranstaltungen geht (obwohl dies oft geschieht).

Die folgenden Dimensionen (fünf bis acht) erweisen sich als schwerer abgrenzbar, weil sie zum Teil ineinander übergreifen.

### **5. "Stimulierung"**

Stimulation und Bestärkung eigenständigen Denkens, zufriedenstellende Stimulation vs. enttäuschte Stagnation, Motivierung durch Interesse, Entwicklung von Eigeninitiative vs. Desinteresse und begeistern können.

Gerade in dieser Dimension wären weitere Forschungen notwendig. In der pädagogischen Psychologie gibt es z.B. Theorien zum sog. "Entdeckenden Lernen", daß u.a. die intellektuelle Neugier der Kinder stimulieren soll. Diese Theorien beziehen sich ausschließlich auf den Schulunterricht. Ähnliche Analysen wären auch für den Hochschulbereich notwendig.

## **6. "Variabilität vs. Monotonie"**

Dies ist ein Faktor, der weitreichendere Punkte beinhaltet als die Kurs- und Stofforganisation. Vermutlich ist es ein wichtiger Aspekt der Informationsaufnahme (Schott, 1973), da er gemeinsam mit anderen die Voraussetzung für das "Aufmerksam bleiben können" bildet und damit wiederum die Voraussetzung für die kontinuierliche Informationsaufnahme über einen längeren Zeitraum .

## **7. "Enthusiasmus"**

Darunter fällt auch Dynamik/ Charisma und Anziehungskraft des Vortragenden. Diese Dimension hat großen Einfluß auf die vorherigen beiden Dimensionen. Enthusiasmus für ein Fach ist mit Sicherheit eine gute Grundlage, um Studierende zu begeistern und intellektuelle Neugier zu stimulieren. Auf der anderen Seite wird es ein charismatischer Lehrer und eine dynamische Dozentin weitaus leichter haben, die Aufmerksamkeit wachzuhalten und weniger Monotonie aufkommen zu lassen.

Wie die „Dr. Fox-Studien“ zum Verhältnis von „Inhaltsdichte“ und „Unterhaltungswert“ nachweisen, behalten Studierende von einem enthusiastischen Vortrag auch mehr Informationen (vgl. Kapitel 4). Problematisch ist an dieser Dimension, daß sie von allen bisher genannten am ehesten auf eine Persönlichkeitseigenschaft deutet, welche für den/ die Lehrende weniger veränderbar oder trainierbar ist. Dennoch spielt dieser Faktor in der Lehre eine große Rolle (vgl. Murray, 1983).

## **8. "Kurswert"**

Wert des Kurses für die Studierenden, Relevanz des Kursmaterials, Prüfungsrelevanz und Theorie-Praxis-Verhältnis. Dieser Faktor beurteilt am stärksten den Kurs aus einer externen Perspektive, und ist somit auch ein erster Ansatz zu einer Kursreform. Stärker als bei den anderen Faktoren wird hier weniger die Qualität des/ der Lehrenden in den Mittelpunkt gestellt, sondern eher die Qualität und Relevanz des angebotenen Stoffes.

In diesen Bereich fallen grob zudem "Kursschwierigkeit" und "Arbeitsbelastung im Kurs". Da festgestellt wurde, daß diese beiden Aspekte kaum mit den Lehrbewertungen korrelieren, sollte überlegt werden, ob sie nicht als einfache Einschätzungen aus Feedbackgründen abgefragt werden können.

Möglich wäre aber auch, die entsprechenden Fragen oder Items so zu formulieren, daß sie zu anderen Dimensionen passen; z.B. kann die Einschätzung: "Der Kurs ist zu schwer" entweder der Kursorganisation oder der Kommunikationsfä-

higkeit des Lehrenden/der Lehrenden zugeordnet werden (bei entsprechender Umformulierung).

## **Schlußfolgerungen**

*Lehre besteht aus verschiedenen Dimensionen. Es kann als erwiesen gelten, daß Studierende unterschiedliche Lehraspekte differenzieren können. Daß in einigen Studien wenige Faktoren gefunden werden, liegt vermutlich häufig an der zu geringen Itemzahl oder an zu wenig differenzierten Formulierungen (z.B. Mazer, 1977); oder aber an ungenügender Kenntnis der Testtheorie, die davon ausgeht, daß ein Merkmal sich aus verschiedenen Merkmalsfacetten zusammensetzt, und diese mit Hilfe von Items erfaßt werden, anstatt zu versuchen, eine Dimension mit einem Item zu erfassen, (wie z.B. Terry und McIntosh, 1988)*

*Fragebogen müssen somit so konstruiert sein, daß sie unterschiedliche Faktoren der Lehrqualität erfassen, da bei globalen Items eine weitaus höhere Anfälligkeit für externe Einflüsse besteht. Wie kann man die möglichen Faktoren ermitteln? Eine „Theorie des Unterrichts“ an den Hochschulen existiert bestenfalls in Teilen. Daher ist ein stringent deduktives Vorgehen gegenwärtig nicht möglich, ein induktives Verfahren somit notwendig.*

*Nach Durchsicht verschiedener empirischer Studien kristallisieren sich acht mögliche Faktoren heraus: Zuwendung, Fairneß von Prüfungen und Benotungen, Kommunikationsfähigkeit, Kurs- bzw. Stofforganisation, Stimulierung, Variabilität vs. Monotonie, Enthusiasmus und Kurswert.*

*Diese acht Faktoren sind grundlegende und in den verschiedensten Fragebogen immer wieder zu findende Dimensionen. Die aufgeführten Dimensionen sollten jedoch nicht als gegeben hingenommen werden. Weitere Untersuchungen in diesem Bereich -und vor allem eine Theoriebildung- wären sinnvoll.*

*Deutlich wird an den genannten Dimensionen, daß sie sich ganz überwiegend nur auf einzelne Kurse und deren Aspekte beziehen. Das ist vor allem darauf zurückzuführen, daß die entsprechenden Studien zumeist aus den Vereinigten Staaten stammen, wo die Bewertung einzelner Lehrender und ihrer Veranstaltungen im Vordergrund steht.*

*Notwendig wäre aber darüber hinaus eine Dimension "Fachbereichsbewertung" in die Evaluationsinstrumente einzuführen, bei der Aspekte der Lehrorganisation, die den ganzen Fachbereich betreffen, bewertet würden. Günstiger noch erscheint, spezielle Fragebogen für die "Evaluation der Lehre eines Fachbereichs" zu entwickeln (vgl. Webler u.a., 1993).*

## 6.4 Vergleichsmaßstab der Beurteilenden

Die Evaluation der Lehre braucht, wie jede andere Bewertung, einen Vergleichsmaßstab, eine *Bezugsnorm*. Ohne vorgegebenen Vergleichsmaßstab verwenden Studierende einen eigenen Maßstab, der unterschiedlich ausgerichtet sein kann. Die Frage nach der "optimalen" Art der Bezugsnorm soll an dieser Stelle diskutiert werden. Ihre Beantwortung ist relevant, da die Formulierung von Items grundsätzlich auch bestimmte Vergleichsoptionen beinhaltet. Diese Optionen verschieben sich je nach gewählter Formulierung.

### Idealnorm oder sozialer Vergleich

Bei der Instrumententwicklung wäre als erstes festzulegen, ob mit den Items Beobachtungen der Studierenden oder Wertungen erfaßt werden sollen. Items, die beobachtbare Elemente der Lehre erfassen sollen, führen insgesamt zu höheren Gütewerten des Instruments.

Dennoch kann es an einigen Stellen sinnvoll sein, Items miteinzubeziehen, die wertende Urteile erfordern, entweder weil der erwünschte Sachverhalt nur so erfaßt werden kann (z.B. die Kursschwierigkeit) oder weil ein Interesse daran besteht, die Wirkung bestimmter Lehrelemente auf die Studierenden zu erfassen (z.B. inwieweit der Kurs bei der Vorbereitung auf die Prüfung nützlich war). Studierende würden in diesen Fällen, d.h. bei wertenden Beurteilungen, die *Selbstreferenz* als Vergleichsmaßstab anlegen.

Bei Items, die vor allem die Einschätzung von beobachteten Sachverhalten zum Ziel haben, sind zwei unterschiedliche Vergleiche möglich: der Vergleich mit einer Idealnorm und/ oder ein sozialen Vergleich.

Die *Idealnorm* kann eine existierende Person sein, deren Unterricht als "ideal" empfunden wird oder aber, was wahrscheinlicher ist, die persönliche Vorstellung von einer optimalen Lehre. Der *soziale Vergleich* dagegen erfolgt durch den Bezug auf andere Lehrende der Hochschule oder des Fachbereichs. Durch entsprechende Formulierung kann sichergestellt werden, welcher Vergleich von den Studierenden durchgeführt wird.

Beispiel I: Bitte bewerten Sie den Dozenten A bei den entsprechenden Items auf der Antwortskala von 1 wie sehr gut bis 5 wie ungenügend. Bei dieser Formulierung ist nicht klargestellt, ob Studierende den Dozenten im Vergleich zu anderen Lehrenden bewerten oder ihn mit ihrer Idealvorstellung vergleichen.

Beispiel II: Bitte bewerten Sie die Dozentin B bei den entsprechenden Items im Vergleich zu anderen Lehrenden auf der Antwortskala von 1 wie weit überdurchschnittlich bis 5 wie weit unterdurchschnittlich.

Beispiel III: Bitte bewerten Sie das Tafelbild des Dozenten A. Bei diesem Beispiel ist nicht näher spezifiziert, welche Elemente des Tafelbilds wichtig sind und bewertet werden sollen. Daher ist zu vermuten, daß Studierende das Tafelbild des Dozenten A nach ihren eigenen Vorstellungen von wichtigem und unwichtigem Aspekten beurteilen werden, da "Tafelbild" eine relativ breite Kategorie ist, die viel Interpretationsspielraum zuläßt.

Beispiel IV: Bitte bewerten Sie inwieweit das Tafelbild der Dozentin B übersichtlich ist. In diesem Fall ist bereits vorher bestimmt, daß die Übersichtlichkeit des Tafelbildes ein wichtiger oder entscheidender Punkt ist, den die Studierenden beurteilen sollen. Bei diesem Beispiel ist gut vorstellbar, daß Studierende die gleichen Kriterien an ein Tafelbild anlegen wie die Gestaltenden des Fragebogens, und auch, daß die Studierenden untereinander gleiche oder ähnliche Kriterien verwenden, da das zu Beurteilende relativ präzise gefaßt ist.

*Levinthal, Lansky und Andrews (1971)* gehen der Vermutung nach, daß Studierende ihre "maximalen Unterscheidungsabstände" strecken, um sie mit den Evaluationsschemata in Einklang zu bringen. Der Grund jedoch, daß Studierende ihr eigenes "Wertesystem" benutzen, liegt vor allem an der Formulierung der Items, wie im weiteren deutlich wird.

263 Studierende der Einführung in die Psychologie gaben in der Studie eine Gesamtbewertung ab und beantworteten einen Fragebogen. Ergebnis des Vergleichs zwischen den Bewertungen und den von den Studierenden genannten Ideal-(Soll-) Normen waren deutliche Unterschiede in der Wichtigkeitseinstufung der Items, z.B. wurde nur ein einziges Item einstimmig möglichst positiv gesehen, d.h. eine möglichst hohe Ausprägung als ideal angenommen; bei drei anderen Items wurde dagegen eine mittlere Ausprägung für ideal befunden, z.B. auch für das folgende Beispielitem: "Er war nachgiebig und flexibel."

Problematisch ist, daß die Autoren (bzw. die Fragebogenkonstrukteure) sich nicht bewußt gemacht haben, daß eine solche Ausprägung nicht an ihrem äußersten Ende, sondern eher im Mittelbereich als optimal betrachtet wird. Wenn weder das Wertesystem der Studierenden ermittelt, noch das Item aus einer Theorie abgeleitet wird, kommt es häufig zu einer nicht akkuraten Interpretation durch Anwenden eines solchen Fragebogens, da in kurzen und globalen Fragebogen die Ergebnisse summiert werden und somit implizit davon ausgegangen wird, daß ein hoher Wert auf einem Item gleichzeitig auch eine außergewöhnlich gute Lehrleistung bedeutet.

### **Einbeziehung von Soll- und Relevanzurteilen**

Sommer und Petermann (1978) haben untersucht, inwieweit präskriptive Urteile den Informationsgehalt des Fragebogens nachweisbar erhöhen. Sommer und Pe-

termann zielen einen Fragebogen an, der ein Modell zur Verknüpfung von Ist-, Soll- und Relevanzurteilen beinhaltet. Das Ist-Urteil bezieht sich dabei auf die Bewertung des momentanen Zustandes. Das Soll-Urteil hat die Funktion, die individuelle Bezugsnorm deutlich zu machen. Das Relevanzurteil entspricht einem Gewichtungsfaktor, mit dem die Ist-Urteile multipliziert wurden. Je geringer die Relevanz eingeschätzt wird, desto stärker nähern sich die gewichteten Urteile an ihr arithmetisches Mittel Null. Dadurch schränkt sich die mögliche Varianz ein und wird korrelationsstatistisch unwirksam.

114 Studierende beantworteten den Fragebogen; vier Image-Analysen wurden durchgeführt: 1) Betrachtet man nur das Ist-Urteil, führt dies zu zwei bipolaren und eindeutig interpretierbaren Faktoren mit 41 Prozent aufgeklärter Varianz. 2) Das Ist-Urteil gewichtet mit dem Relevanzurteil kommt zu drei Faktoren und 38 Prozent aufgeklärter Varianz. Die Faktoren sind jedoch unipolar und mit Ausnahme des ersten kaum zu benennen. 3) Nimmt man die Ist-Soll-Differenz, so ergibt sich ein starker Generalfaktor (38 Prozent der Varianz) und drei weitere kaum interpretierbare Faktoren. Insgesamt kommt es zu 59 Prozent aufgeklärter Varianz. Jedoch ist eine Vergrößerung der Struktur zu bemerken. 4) Gewichtet man die Ist-Soll-Differenz mit dem Relevanzurteil, ergeben sich 4 Faktoren und 53 Prozent aufgeklärter Varianz. Die Faktorenstruktur ist gut interpretierbar und differenzierter als in den vorherigen Analysen.

Die Schlußfolgerung der Autoren lautet: Präskriptive Urteile beinhalten wichtige Zusatzinformationen, daher sollten Soll- wie auch Relevanzurteile in die Lehrbewertung miteinbezogen werden.

Leider findet sich im Artikel von Sommer und Petermann nicht ein einziges Beispiel aus ihrem Fragebogen. Wenn man davon ausgeht, daß auch sie Items benutzen, deren Vergleichsnormen nicht eindeutig geklärt und deren optimale Ausprägungen nicht aufgeschlüsselt wurden, dann ist einsichtig, daß die Faktorenstruktur durch Einbezug von Soll- und Relevanzurteilen, somit durch die Klärung der Vergleichsnorm, differenzierter und deutlicher wird. Es ist dadurch möglich, einen höheren Informationsgehalt zur Ansicht der Studierenden über die Lehre zu erzielen.

Dieses Verfahren erscheint jedoch für die Lehrbewertung sehr aufwendig und sollte daher auf die Konstruktion entsprechender Instrumente beschränkt bleiben. Bei der Entwicklung von Instrumenten zur Lehrevaluation ist es notwendig, sowohl Studierende als auch Lehrende nach den für sie relevanten Aspekten zu befragen. Im späteren Einsatz bei der Lehrbewertung erscheint dieser Aufwand nicht mehr nötig.

Es ist daher unbedingt darauf zu achten, daß bei den Itemformulierungen allgemein akzeptierte Normen von Lehraspekten vorgegeben und in den Formulierun-

gen zu erkennen sind. Wenn das Ziel der Fragebogenkonstruktion sein soll, einen Fragebogen mit möglichst hohen Gütekriterien (im Sinne der Testtheorie) zu erhalten, so kann dies nur durch eine möglichst niedrige Subjektivität der Beurteilungen erreicht werden. In einem solchen Falle sollten die Items, die von den Studierenden keine Beobachtungen, sondern Wertungen erheben (d.h. als Bezugspunkt die Selbstreferenz haben), gesondert ausgewertet werden

Ein noch offen gebliebener Punkt ist die Frage nach dem Vorzug eines "idealen" oder eines "sozialen" Vergleichssystems. Dies läßt sich mit einem Zitat des Schulforschers Weiß beantworten: "Für eine Lehrveranstaltungskritik gib es gegenwärtig keine lehrzielorientierten Bezugsnormen. Jeder Studierende hat aber Erfahrungen mit verschiedenen Lehrern. Die Mängel des sozialen Maßstabs erscheinen mir bei der Beurteilung von Lehrleistungen geringer. Ich halte es für vertretbar, soziale Maßstäbe anzuwenden, etwa 'im Vergleich zu anderen Lehrveranstaltungen überdurchschnittlich - durchschnittlich - unterdurchschnittlich'" (Weiß, 1989, S. 39).

Da im Moment für den Hochschulunterricht keine andere Bezugsnorm vorhanden ist, d.h. kein Kategoriensystem für gute Lehre existiert, erscheint es angebracht, die Studierenden aufzufordern, einen Vergleich mit anderen Lehrenden zu ziehen.

## **Schlußfolgerungen**

*Es erscheint sinnvoll, bei der Konstruktion eines Fragebogens die Ansichten der Betroffenen zu erheben (subjektive Norm). Damit kann gewährleistet werden, daß zum einen die relevanten Aspekte der Lehrqualität im Instrument enthalten sind und zum anderen ihre unterschiedliche Gewichtung deutlich wird.*

*Bei der Datenerhebung jedoch sollte von den einzelnen Meinungen abstrahiert und ein möglichst „objektives Instrument“ angestrebt werden (objektive Norm). Dies ist vor allem dann erforderlich, wenn Lehrende miteinander verglichen werden sollen, da für eine angestrebte hohe Meßgenauigkeit eine geringe Antwortvarianz Voraussetzung ist.*

*Sollen die Studierenden wertende Aussagen abgeben, so sind diese getrennt von den Fragestellungen zur Lehrveranstaltung und ihrem Verlauf zu erheben und auszuweisen.*

*Zusammenfassend läßt sich sagen, daß es vorteilhaft ist, die angestrebten Vergleichsmöglichkeiten (Bezugsnorm) konkret vorzugeben. Ein sozialer Vergleich und genauer gefaßte Beschreibungskategorien sind dabei zweckmäßig.*

## **7 Kriterien und Empfehlungen zur Gestaltung von Fragebogen**

### **7.1 Formulierung der Fragen und Items**

Das wichtigste an einem Fragebogen sind die Fragen und Vorgaben, die sog. Items. An ihrer Qualität entscheidet sich die Validität der Befunde. Auf ihre Formulierung ist daher besondere Aufmerksamkeit zu verwenden.

An dieser Stelle sollen nicht allgemeine Kriterien der Itemformulierung wie z.B. Verständlichkeit oder Vermeidung von Antworttendenzen aufgeführt werden, wie sie in entsprechenden Lehrbüchern über empirische Methoden und die Konstruktion von Fragebogen zu finden sind. Vielmehr soll auf jene wichtigen Punkte verwiesen werden, die speziell bei der Formulierung von Items (Fragen und Vorgaben) einer Lehrbewertung zu beachten sind.

#### **Fragen und Vorgaben sollen von Beobachtbarem ausgehen**

Häufig sind in Fragebogen zur Lehrbewertung Items zu finden, die nach nicht beobachtbaren Elementen fragen. Dieses Problem ist, trotz vieler „Spontanerhebungen“ zur Veranstaltungskritik, in Deutschland nicht so häufig anzutreffen wie in den U.S.A. Dennoch schleichen sich in viele hiesige Fragebogen ebenfalls solche Items ein.

Alle folgenden Beispiele sind einer Dokumentation zur Evaluation der Lehre (Reissert, 1992, Teil 1) entnommen: "Wie gut ist der Professor vorbereitet?" und "Wie ist die fachliche Kompetenz des Dozenten?" sind zwei typische Fragen. Beide angesprochenen Sachverhalte sind jedoch für die Studierenden nicht direkt wahrnehmbar, sondern erschließen sich ihnen durch verschiedene im Kurs wahrgenommene Begebenheiten. Bei der Frage nach der Güte der Vorbereitung der Lehrenden z.B. kann aus der Beantwortung von Fragen zur Strukturierung der Vorlesung, dem berühmten "roten Faden", oder ähnlichem auf die vermutete Vorbereitung geschlossen werden.

Die Abfrage solcher „Vermutungen“ erscheint an zwei Punkten problematisch, wenn die „Lehrleistung“ erfaßt werden soll. Zum einen kann es vorkommen, daß sich Lehrende gut vorbereiten, dies aber nicht vermitteln können. Zum anderen verleitet eine solche Fragestellung jeden Studierenden zu seiner eigenen Interpretation von "gut vorbereitet". Thorndike und Hagen (1969) geben zu bedenken: "Bewertungen haben die größte Chance genau zu sein für die Aspekte, die nach außen hin gezeigt werden .... Die Erfahrung hat gezeigt, daß diese valider bewertet werden" (S. 428).

Die Einsicht, daß die Antworten der Studierenden subjektiv sein könnten, ist auch bei vielen Fragebogengestaltenden vorhanden. Um das Problem zu umgehen und Lehrenden nichts Falsches zu unterstellen, sind daher immer wieder Items der folgenden Art zu finden: "Wie wichtig, *glaubst Du*, ist dem Dozent der Lehrauftrag?", "Wie *beurteilst Du* das Engagement des Professors im Seminar?", "Zeigt der Professor starken persönlichen Einsatz, oder *wirkt er* eher lustlos?", "Mein Eindruck ist, daß die Einstellung des Dozenten zu den Studenten positiv ist.", "Der Dozent aktualisiert *meiner Meinung* nach sein Fachwissen." (Hervorhebung durch die Autorin.)

Studentische Veranstaltungskritik sollte keine "Meinungsumfrage" sein, sondern ein Beurteilungsinstrument im Sinne einer einschätzenden Beobachtung. Der direkte Nutzen einer Information sowohl für den administrativen Gebrauch als auch für Verbesserungsvorschläge zur Lehre ist äußerst gering, wenn die Frage z.B. nach dem "Engagement im Seminar" nicht näher definiert wird.

### **Wie eng sollen Items gefaßt werden?**

Bei der Formulierung von Items zu Lehrbewertung stellt sich die Frage, wie nah an Beobachtbarem diese abgefaßt werden sollen. Lehrbewertungsbogen unterscheiden sich maßgeblich von üblichen Beobachtungsmethoden durch die Breite der Beobachtungseinheiten. Dennoch ist es für die Güte eines Fragebogens zur studentischen Veranstaltungskritik wichtig, *Items möglichst nahe an Beobachtbarem* zu verfassen. Wie genau also müssen Itemeinheiten formuliert werden?

Ein trefflicher Hinweis ist bei Faßnacht (1979) zu finden: "Die Frage nach der Auflösung ist eine Frage nach dem Zweck ... Wir können ja versuchen, einen bayerischen Hackbraten ebenso fein aufzuschneiden wie Bündner Fleisch..." (S. 75). Man kann davon ausgehen, daß für eine Lehrevaluation relativ grobe Kategorien ausreichen würden. Allerdings ist dabei zu beachten, daß je ungenauer die Einheiten gewählt werden, desto eher dies zu niedrigeren Gütekriterien führen kann und praktische Folgerungen erschwert werden.

Ein *optimierter Mittelweg* muß somit zwischen der Forderung nach möglichst genauer Verhaltensbeschreibung und dem begrenzten Umfang eines Beurteilungsinstruments gefunden werden. Dieser muß jeweils für den einzelnen Fragebogen neu "bestimmt" werden.

### **Verantwortlichkeit des oder der Lehrenden**

Es müßte selbstverständlich sein, daß nur Aspekte in eine Lehrbewertung einfließen, die von den Lehrenden tatsächlich beeinflusst werden können. Ein in überraschend vielen Fragebogen auftauchendes Item ist folgendes: "Konnte Dich

der Dozent für den Stoff motivieren?" Alle Lehrenden werden die Erfahrung gemacht haben, daß es nicht einzig und allein von ihnen abhängt, ob einzelne Studierende sich für bestimmte Themen motivieren lassen.

Dieses Problem wird bereits in der vermutlich ersten deutschen Untersuchung zur studentischen Veranstaltungskritik von *Müller-Wolf und Fittkau* (1971) deutlich: "Die 'Entwicklung von Eigeninitiative im behandelten Fachgebiet' (...) hängt hingegen nur in schwächerem Maße mit 'didaktisch effektiver Aktivität' (...) und 'positivem demokratischem Lehrverhalten' ( $r = .44$  u.  $r = .41$ ) des Hochschullehrers zusammen. Hier zeigen sich die Grenzen der Auswirkungen des Lehrverhaltens und der Einfluß eigenständiger Entschlüsse der Studierenden" (S. 177).

Ähnlich liegt die Problematik bei der Frage nach einer ausreichenden Raumgröße. Oder aber wenn Assistenten/ Assistentinnen für die Konzeption der Vorlesung bewertet werden, für die sie nicht verantwortlich sind.

Sinnvoll können solche Fragen nur erhoben werden, wenn sie als zusätzliche Variablen verwandt werden, aber nicht in eine Wertung für die entsprechenden Lehrenden miteinbezogen werden, die zum Vergleich mit anderen Lehrenden dienen soll.

### **Relevanz von Items für die Lehrbewertung**

Lehre hat verschiedenste Aspekte, die man mit einem Fragebogen erfassen kann. Jedoch unterscheiden nicht alle Aspekte die Stärken und Schwächen der Lehrenden voneinander. Ein gutes Beispiel dafür ist die Kursschwierigkeit, die in verschiedenen Untersuchungen wenig bis kaum mit anderen Dimensionen der Lehrbewertung korreliert und wenig die Gesamtbewertung der Lehrenden beeinflusst. Ein solches Item ist somit für die Lehrbewertung nicht maßgeblich. Nur Items, die zwischen den Gruppen besserer und schlechterer Lehrender unterscheiden, sollten in die vergleichende Bewertung eingehen.

Dabei erhebt sich auch die Frage, wie mit Items umgegangen wird, die in einigen Kursen wichtige, in anderen dagegen keine Rolle spielen. Eine Lösungsmöglichkeit stellt das *Cafeteria-Prinzip* dar. Dieses Prinzip (vgl. Webler, 1992, S. 150 ff.) bietet die günstige Möglichkeit, ohne größere Schwierigkeiten Fragebogen abzuwandeln, um sie für spezielle Kurse passend zu gestalten. Vorteilhaft dabei wäre, daß ein Pool bereits formulierter und evaluierter Items existieren würde, die einen geringeren Aufwand bei der Fragebogenzusammenstellung ermöglichen.

Allerdings muß, um die Vergleichbarkeit zu gewährleisten, ein bestimmter Teil an Items identisch sein. Möglich wäre, "*Itempakete*" für verschiedene Kursarten, Fächer u.ä. zusammenzustellen. Webler (1992) schlägt ein 'Zwiebelmodell' vor,

mit einem Standardteil, zusätzlichen fach- bzw. lokalspezifischen Fragen und wenigen offenen Fragen. Eher aus "pädagogischen Gründen" wäre es sinnvoll, zusätzlich einige Items von Lehrenden der jeweiligen Kurse nennen zu lassen.

Letzterer Punkt ist relativ wichtig, da immer wieder Lehrende vermuten, daß sie besser bewertet würden, wenn sie selbst die Items formulieren oder zumindest zusammenstellen könnten. Mishra (1979) untersuchte diese Vermutung an 76 Kursen (mit 1650 Studierenden) in denen 50 Lehrende bewertet wurden. Der dabei benutzte Itempool enthielt 227 Items zu verschiedenen Aspekten der Lehrbewertung. Die 27 Items aus diesem Pool, die sowohl von Studierenden als auch von Lehrenden hoch bewertet wurden, stellte Mishra zu einem Fragebogen zusammen. Die Dozenten und Dozentinnen wurden aufgefordert, vor der Verteilung der Fragebogen diese Items aufmerksam durchzulesen und die wichtigsten Items für ihren Kurs zu identifizieren.

Mishra stellte fest, daß Lehrende signifikant besser auf den von ihnen genannten Items bewertet wurden. Korrelationen zwischen dem, was ein Lehrender wichtig findet, und dem, wo er gut bewertet wird, sind logischerweise hoch. Dies kann jedoch entweder dissonanztheoretisch erklärt werden: In den Lehraspekten, die man als wichtig empfindet, wird man sich auch um gute Leistungen bemühen, um eine Dissonanz zu vermeiden. Oder es ist als selbstwertdienliche Kausalbeziehung zu sehen, d.h. Lehrdimensionen, die gut beherrscht werden, werden auch als wichtig empfunden.

Eine andere Sache dagegen ist es, wenn Items nicht von einzelnen Lehrenden als "wichtiger" betrachtet werden, sondern entweder eine allgemeine Übereinstimmung besteht oder eine theoretische Begründung vorhanden ist, warum manche Aspekte der Lehre wichtiger sind als andere und somit stärker gewichtet werden sollten.

## **Schlußfolgerungen**

*Instrumente der studentischen Veranstaltungskritik sollten sich aus Items zusammensetzen, die möglichst von Beobachtbarem ausgehen, damit die Beurteilungen der Studierenden eindeutig sein können. Dabei muß ein akzeptabler Mittelweg zwischen der Genauigkeit der Itemformulierung und der Fragebogenlänge gefunden werden.*

*Die Fragen müssen des weiteren so gestellt sein, daß es für die Studierenden deutlich ist, ob sie Lehrende an ihren Idealvorstellungen oder an anderen Lehrenden vergleichen sollen. Es muß für die Studierenden fest umrissen sein, welchen Vergleichsmaßstab (Anker) sie bei der Beantwortung derselben anlegen sollen.*

*Weiter sollten Lehrende nur in Dimensionen bewertet werden, für die sie verantwortlich sind und die gute und weniger gute Lehrende tatsächlich unterscheiden.*

*Es sollte die Möglichkeit bestehen, allgemein interessierende Fragen, wie nach der Grundmotivation für das Fach oder der Überfüllung der Veranstaltung, zu stellen. Jedoch sollten diese Fragen klar abgegrenzt von den Lehrbeurteilungen sein und nicht in Maße zur Lehrleistung oder Lehrqualität einbezogen werden.*

*Weiter ist vorab zu klären, ob alle Aspekte gleich wichtig für die Beurteilung der Qualität einer Veranstaltung oder der Lehrleistung eines/ einer Lehrenden sind, und somit gleichbedeutend in die Bewertung einbezogen oder verschieden gewichtet werden.*

## **7.2 Zur Gestaltung des Gesamtbogens**

An dieser Stelle sollen Fragen der Instrumentengestaltung, die nicht einzelne Items oder Dimensionen, sondern den gesamten Fragebogen betreffen, behandelt werden.

### **Länge des Fragebogens**

Die einzige vorliegende Untersuchung zu dem nicht unwichtigen Punkt der Fragebogenlänge wurde von Meredith (1981b) erstellt. Im Anschluß an einen Fragebogen beantworteten 4.877 Studierende eine gestellte Zusatzfrage nach der bevorzugten Fragebogenlänge. Genannte Antworten variierten zwischen 1-75 Items; 50 Prozent der Studierenden nannten 20-40 Items. Durchschnittlich wurden 25 Items genannt. Meredith schließt daraus: "Evaluationsskalen von 15 bis 35 Items... sind für Studierende akzeptabel" (S. 490).

Um zu beantworten, wie lang ein Fragebogen zur Lehrbewertung sein darf, müssen verschiedene Aspekte beachtet werden. Natürlich ist allein schon nach Testgütekriterien ein längerer Fragebogen vorteilhaft. Man ist jedoch bei der Beantwortung darauf angewiesen, daß Studierende ihn möglichst sorgfältig ausfüllen. In welchem Maße dies geschieht, hängt von der Motivation der Studenten und Studentinnen ab. Und: Motivation ist beeinflussbar!

Für die praktische Anwendung wäre es denkbar, daß am Beginn einer institutionalisierten Evaluation mit sorgfältigen Erklärungen und Einführungen relativ umfangreiche Fragebogen einsetzbar wären. Für regelmäßige Durchführungen müßten sie dann vermutlich gekürzt werden, um ein Routineinstrument zu erhalten. Jedoch wären die Kürzungen so vorzunehmen, daß die notwendige Multidimensionalität des Instrumentes nicht darunter zu leiden hat.

## **Anzahl der Antwortpunkte (Ratingpunkte)**

Auch zur Frage der Zahl der Antwortpunkte (Ratingpunkte) liegt nur eine Studie vor. Lily Neumann & Neumann (1981) werteten einen von 500 Studierenden ausgefüllten Fragebogen aus acht geisteswissenschaftlichen Kursen aus. Dieser enthielt Items zur "Zufriedenheit mit verschiedenen Kursaspekten". Die Studierenden beantworteten den Fragebogen in sechs randomisiert ausgesuchten Gruppen. Variiert wurde dabei einzig die Länge der Ratingskala, die 2,3,4,5,7 oder 10 Punkte enthielt.

Das Wissenschaftspaar konnte feststellen, daß alle Studierenden die komplette Skala bei den wenigen Punkten (2/3) nutzten, 80 Prozent dies bei den mittleren (4/5) taten und nur 60 Prozent machten von dieser Möglichkeit bei den beiden langen Skalen (7/10) Gebrauch. Lily Neumann und Neumann kommen zu der Schlußfolgerung, daß die beiden kurzen Skalen zu besseren Ergebnissen führen, jedoch zu wenig Differenzierungsmöglichkeiten bieten. Die beiden langen Varianten weisen eine Vermeidung der Extreme auf und erscheinen ihnen zu verschwommen. Sie vermuten einige Schwierigkeiten bei der Abgrenzung der Beurteilungen und argumentieren schließlich für eine 5-Punkt-Skala.

Die Anzahl der Punkte auf der Rating-Skala ist auch im Hinblick auf die Bezugsnorm der Lehrbewertung bedeutsam. Würde man mit "Idealnormen" arbeiten, so müßte definiert werden, was die einzelnen Ratingpunkten zu bedeuten hätten: z.B. 1 - optimal, 2 - nur in Einzelpunkten verbesserungsbedürftig, 3 - verbesserungsbedürftig, usw., damit die Befragten nicht den einzelnen Ratingpunkt anders interpretieren. Benutzt man dagegen eine "Sozialnorm", so müßten die Ratingpunkte als Vergleiche formuliert werden, z.B. im Vergleich zu anderen Lehrenden: 1 - weit überdurchschnittlich, 2 - überdurchschnittlich, 3 - durchschnittlich, usw..

Nach der Untersuchung von Lily Neumann & Neumann (1981) ist zu bezweifeln, daß bei langen Skalen die einzelnen Ratingpunkte unterschieden werden können, was aber auch von der Prägnanz der Formulierung abhängig ist. Gerade bei der Sozialnorm ist es daher unwahrscheinlich, daß mehr als fünf Ratingpunkte unterscheidbar sind.

Auch das Skalenniveau, d.h. das Format der Antwortvorgaben und damit die Art der Meßdaten, sollte nicht vergessen werden. Wie solche Instrumente der Beurteilung im allgemeinen verfügen Fragebogen zur Lehrevaluation vermutlich in den seltensten Fällen über Intervallskalenniveau. Auf Basis des Ordinalskalenniveaus ist es zwar möglich zu ermitteln, daß Lehrende z.B. eine weniger gut aufgebaute Kursstruktur haben, aber nicht, um wieviel schlechter diese im Vergleich zu anderen ist. Auch aus diesem Grund erscheint es sinnvoll, die Antwortstufen nicht allzu sehr auszudifferenzieren.

### **Instruktion: Angaben zum Verwendungszweck**

Lehrbewertungen können mit ganz unterschiedlichen Zielsetzungen durchgeführt werden. Im weiteren wird der Frage nachgegangen, ob der spätere Verwendungszweck, wenn er den Studierenden bekannt ist, ihre Bewertung beeinflusst.

Orpen (1980) untersuchte, ob 326 Studierende der Einführung in die Wirtschaftswissenschaften sich von der Art der Einleitung beeinflussen ließen. Ihnen wurde entweder mitgeteilt, daß der Fragebogen nur Feedbackzwecken dienen sollte oder daß er Auswirkungen auf Beförderung, Festanstellung und Gehalt haben würde. Die Ergebnisse zeigen signifikant bessere Bewertungen (1%-Niveau), wenn die Auswertung der Fragebogen administrativen Zwecken dient.

Pasen et al. (1978) untersuchten zwei Kurse mit 85 Studierenden, die Kurse der Einführung in der Psychologie bewerteten. Das benutzte Instrument, die "Endeavour Instructional Rating Card", weist zwei Faktoren auf: "Fach-/ Lehrkenntnis" und "Verhältnis zu den Studierenden". Verglichen wurde die Bewertung zweier Dozenten A und B bei variiertem Instruktion (Feedbacknutzen und administrativer Nutzen). Ergebnis: Ein signifikanter Einfluß konnte nur auf den Faktor "Verhältnis zu den Studierenden" ermittelt werden, lag somit auf der "affektiven Ebene".

Laura A. Driscoll und Goodwin (1979) kamen bei einer Studie an 92 Kursen zu folgendem Ergebnis: Lehrende erhielten signifikant höhere Werte bei den Instruktionen: "Administrative Zwecke" und "Kursverbesserung". Es konnten Anzeichen ermittelt werden, daß von der Instruktion eher kleine und graduierte Kurse betroffen waren.

In der Studie von Blount et al. (1978) analysierten die Autoren Daten von 40 Dozenten und Dozentinnen, die mit dem "Illinois Course Evaluation Questionnaire" (CEQ) bewertet wurden. Dabei wurden in der Instruktion folgende Nutzungszwecke genannt: Feedback, Administrative Zwecke und Informationsmöglichkeit Studierender in den nächsten Semestern. An die Studierenden wurde randomisiert der CEQ mit den variierten Instruktionen verteilt. Lehrende erhielten signifikant unterschiedliche Werte entsprechend den verschiedenen Instruktionen.

Aleamoni und Pamela Z. Hexner (1980) analysierten zwei vergleichbare Gruppen eines Geschichtskurses beim gleichen Dozenten. In zwei Durchläufen bewerteten ihn 164 bzw. 80 Studierende mit dem Course Evaluation Questionnaire (CEQ/ 50 Items). Im ersten Durchlauf wurde der CEQ nur mit einigen Standardformulierungen ausgeteilt, im zweiten Durchlauf wurden die Zwecke des CEQ (Bestimmung von Gehaltserhöhung, Beförderung etc.) zusätzlich verbal erläutert. Es konnte festgestellt werden, daß beim zweiten Durchlauf die Bewertung auf allen sechs Subskalen signifikant höher war.

Insgesamt kann als gut abgesichert gelten, daß die Instruktion bzw. der dort genannte Verwendungszweck die Lehrbewertung signifikant beeinflusst. Um eine Vergleichbarkeit der Ergebnisse zu ermöglichen, ist die Empfehlung von Laura Driscoll und Goodwinn (1979) zu berücksichtigen, die zu einer Standardisierung raten.

### **Auftraggeber**

Mit dem Problem unterschiedlicher Instruktionen verwandt ist die Frage nach dem Einfluß verschiedener Auftraggeber. Das heißt: Beeinflußt es Studierende, ob eine Lehrevaluation von Kursleitern und -leiterinnen, der Fachschaft oder z.B. der Verwaltung durchgeführt wird?

Bedauerlicherweise gibt es dazu nur eine einzige Untersuchung. Orpen (1980) kam zu der Feststellung, daß die Ergebnisse in den Lehrbefragungen deutlich besser sind (1%-Niveau), wenn die Befragung als von Lehrenden initiiert dargestellt wurde, im Unterschied zu einer Einführung, in der das Studierendenparlament als Auftraggeber genannt wurde.

Es ist anzunehmen, daß sich Studierende in den U.S.A. und Deutschland nicht so grundlegend unterscheiden und daher auch hier ähnliche Effekte bei unterschiedlichen Instruktionen möglich sind. Es ist allerdings zu vermuten, daß die Rangreihe der Lehrenden in beiden Fällen die gleiche sein dürfte, falls vergleichbare Dimensionen verwandt werden.

### **Anonymität**

Erstaunlich wenige Untersuchungen wurden zur Anonymität von Befragungen und den Folgen für die Lehrbewertung durchgeführt. In der Umfrage von Penfield (1978) wurde ermittelt, daß eine zufriedenstellende Anonymität offensichtlich nicht immer erreicht werden konnte. Nur ein Drittel der Studierenden empfand die Anonymität als meistens gesichert, 54 Prozent waren der Ansicht, daß dies nur manchmal der Fall sei.

Dabei kommt erschwerend hinzu, daß es sich um Lehrbewertungen handelt, bei denen die Namen nicht angegeben wurden, d.h. die Studierenden nur Angst haben mußten, daß der/ die Lehrende z.B. ihre Schrift erkennen könnte (siehe auch Laura Driscoll und Goodwinn, 1979).

Orpen (1980) konnte jedoch in seiner Untersuchung keine Bewertungsunterschiede zwischen namentlich und anonym beantworteten Fragebogen finden.

## Schlußfolgerungen

*Als groben Richtwert für die Fragebogenlänge können nach Meredith (1981b) 15-35 Items gelten. Ansonsten ist die Fragebogenlänge vom Verwendungszweck abhängig. Ein häufig eingesetztes Routineinstrument muß natürlich entsprechend kürzer sein als eine umfassende, einmalige Grund-Datensammlung.*

*Nach Lily Neumann und Neumann (1981) ist eine 5-Punkte Rating-Skala zu empfehlen. Allerdings sind auch Instrumente mit Antwortmöglichkeiten von einer groben 3-Punkte Skala bis zu einer typischen 6-Punkte Skala (entsprechend des deutschen Notensystems) oder bis hin zu einer 7-Punkte-Skala verwendbar.*

*Will man vergleichbare Ergebnisse erreichen, so erscheint es angebracht, standardisierte Instruktionen zu verwenden. Für die Verantwortlichkeit der Studierenden spricht, daß Ergebnisse der Veranstaltungskritik bei schwerwiegenderen Konsequenzen positiver ausfallen. Entsprechend sollte man auch beim Vergleich von Ergebnissen unterschiedlicher Auftraggeber vorsichtig sein.*

*Viele Faktoren sprechen für eine möglichst anonyme Datenerhebung. Anzunehmen ist jedoch, daß sich dies nicht auf die Ergebnisse auswirkt. Diese Diskussion scheint in Deutschland aber schon abgeschlossen, noch bevor sie geführt wurde: Es werden nur anonyme Daten verwendet.*

### 7.3 Vergleichstabellen nach Veranstaltungsart und Fächern

Lehrende sollten vergleichen können, wie sie in der Gegenüberstellung mit ihren Kolleginnen und Kollegen bewertet werden. Will man diesen Vergleich nicht nur am jeweiligen Fachbereich oder an der jeweiligen Hochschule belassen, sind langfristig Vergleichstabellen für das entsprechende Instrument notwendig, etwa für verschiedene Kursarten (Vorlesung, Seminar u.a.) oder verschiedene Fächer.

Patricia A. Cranton und Schmith (1986) sprechen sich für ein Normsystem aus, das die unterschiedlichen Kurscharakteristika miteinbezieht. Diese vorgeschlagene unterschiedliche Eichung bezieht sich vor allem auf die bereits erwähnten unterschiedlichen Bewertungen, z.B. für große vs. kleine Klassen, für verschiedene Veranstaltungsformen und für alle weiteren externen Kriterien, die nachweislich einen Einfluß auf die Lehrbewertungen haben.

Jedoch ist fraglich, ob man sehr unterschiedliche Fächer einheitlich in Bezug auf die Qualität der Lehre vergleichen kann. Im folgenden wird der Frage nachgegangen, ob die Wichtigkeit unterschiedlicher Lehraspekte über verschiedene Fächer divergiert.

In der Untersuchung "Studenten und Hochschullehrer" des Ministeriums für Wissenschaft und Kunst (Baden-Württemberg) werden in verschiedenen Fächern unterschiedlich eingeschätzt: Wichtigkeit des Sprechstundenbesuchs, Inanspruchnahme und Einschätzung der Nützlichkeit von Studien- und Fachschaftsberatung, unterschiedliche Akzeptanz studentischer Kritik und unterschiedliche Studienmotivationen. Schon allein zahlenmäßig ist das Betreuungsverhältnis unterschiedlich (Gleich et al., 1982).

Ähnliches wird in dem Bericht: "Studiensituation und studentische Orientierung an Universitäten und Fachhochschulen" des Bundesministerium für Bildung und Wissenschaft (1988) deutlich. Studierende verschiedener Fächer unterscheiden sich in Bezug auf Geschlechterverteilung, soziale Herkunft und Berufserfahrung (Peisert et al., 1988), Prozentsatz an Zweitstudierenden, in der Schule belegte Leistungskurskombinationen, d.h. in Bezug auf unterschiedliche schulische Vorerfahrung und Motive der Studienfachwahl.

Außer den Studierendenpopulationen unterscheiden sich auch die Fächer voneinander bzw. werden als unterschiedlich empfunden in Bezug auf: die unterschiedliche Beteiligung an Forschungsprojekten und unterschiedlichen Praxisbezug bzw. dem Wunsch danach, Regelungsdichte, Studieraufwand, das empfundene soziale Klima, die Notenstandards, die Berufsaussichten, das politische Interesse und die alternative oder konventionelle Orientierung. Beim letzteren wurde auch die Auswirkung auf das Studium überprüft: Konventionell Orientierte legen deutlich mehr Wert auf eine hohe soziale Position und ein gutes Einkommen und auf eine leistungsbezogene Studienhaltung.

Obwohl Befunde einer Erhebung aus den Jahren 1984/ 85 benutzt wurden, kann man davon ausgehen, daß sich in den letzten zehn Jahren die grundlegenden Tendenzen, vor allem in Bezug auf die unterschiedlichen Anforderungen in den Fächern, nicht geändert haben (vgl. Bargel, Multrus und Ramm, 1996; Multrus, 1995).

Steffani Englert (1991) ermittelte Unterschiede bei Studierenden der Fächer Erziehungswissenschaften, Rechtswissenschaften, Elektrotechnik und Maschinenbau, vor allem bezüglich Zeitbudget und Lebensstil. Hinweise für die Fächerunterschiede sind auch in der Untersuchung von Andrea Frank (1991) zu finden.

Marquez, David und Dorfman (1979) hatten bei ihrer Studie eine ähnliche Zielsetzung. Mögliche Differenzen in der Bewertung von Studierenden verschiedener akademischer Disziplinen sollten gefunden werden. Die Studierenden wurden auch nach ihrem Konzept einer 'effektiven Lehre' befragt. Interviewt wurden vierzig Studenten und vierzig Hochschullehrer, gleichermaßen verteilt auf insgesamt vier verschiedene Fakultäten. Von den befragten 80 Personen wurden nur bei acht systematische Variationen gefunden. Diese waren abhängig von der Fa-

kultätzugehörigkeit, die bei der Beschreibung des fiktiven Lehrenden genannt wurde.

Vorliegende Untersuchungen lassen die Vermutung zu, daß Studierende verschiedener Fächer ihre Lehrenden nach unterschiedlichen Ansprüchen und Kriterien bewerten. Leider ist es nicht möglich, eindeutig zu sagen, ob sich die Zugehörigkeit zu einem Fach in der Beurteilung von Lehrenden niederschlägt. Man kann nur vermuten, daß dies bei der studentischen Veranstaltungskritik einen möglichen Einflußfaktor darstellt, obwohl die Studie von Marquez, David und Dorfman (1979) dagegen spricht.

### **Schlußfolgerungen**

*Insgesamt scheinen in diesem Bereich der mit der Fachzugehörigkeit zusammenhängenden möglichen Einflußfaktoren auf die Lehrbewertung und deren Berücksichtigung in „Vergleichstabellen“ weitere Untersuchungen erforderlich. Es kann für mögliche Vergleichstabellen jedoch festgestellt werden, daß sie vermutlich für verschiedene Veranstaltungsarten, u.U. auch für verschiedene Semester oder Kursgrößen, unterschiedlich sein müßten.*

*Kursarten- und Fächerunterschiede könnten ohne größere Bedenken mit Hilfe unterschiedlicher Itemzusammenstellung (Cafeteria-Prinzip) aufgefangen werden. Ob Vergleichstabellen dennoch nach Fächern angelegt werden sollten, müßte noch genauer geprüft werden.*

*Bisher scheint nur von der Projektgruppe um Prof. Diehl in Gießen geplant, eine Datenbank aufzubauen (vgl. Kapitel 8). Diese soll sich aus den Evaluationsdaten speisen, die mit den von ihm entwickelten Instrumenten (Fragebogen zur Beurteilung von Seminaren mit Referaten [VBREF] und Fragebogen zur Beurteilung von Vorlesungen [VBVOR]) gewonnen werden.*

### **7.4 Ergebnisdarstellung und ihre Vorbedingungen**

Einige Vorbedingungen für verwendbare Lehrevaluationen, wie z.B. ein reliabler und valider Fragebogen, wurden schon behandelt. Um aber verwertbare Ergebnisse erreichen zu können, müssen zusätzlich gewisse Standards der Beteiligung und des Rücklaufs beachtet werden.

Dabei sind vor allem zwei Prozentzahlen wichtig: Zum einen der Anteil Studierender, der im Laufe des Semesters den entsprechenden Kurs verließ, die sogenannte *Drop-out-Quote*; zum anderen der Anteil Studierender, der den Fragebogen ausgefüllt zurückgegeben hat, nachdem er im Kurs verteilt wurde, die sog.

*Rücklaufquote.* Außer diesen Vorbedingungen sollen an dieser Stelle außerdem einige Kriterien der Ergebnisdarstellung genannt und die Frequenz der Datenerhebung besprochen werden.

### **Abnehmen der Teilnehmerzahl (Drop-out-Quote)**

Im Vergleich zur Rücklaufquote gibt es zu diesem Thema keinerlei Untersuchungen im Bereich der Lehrevaluation. Es konnten nicht einmal Überlegungen oder Ansätze zu Hypothesen gefunden werden. Daher sollen im weiteren nur Fragen, die auch bei allen Längsschnittstudien eine Rolle spielen, aufgeworfen und keine Konklusionen vorgestellt werden.

Das Phänomen des „Kursschwundes“ ist allen Mitgliedern und Beobachtenden von Hochschulen bekannt und in nahezu allen Fachbereichen zu finden: Am Beginn des Semesters sind die Veranstaltungen oft brechend voll, je weiter das Semester voranschreitet, um so weniger Teilnehmer und Teilnehmerinnen sind in den Kursen zu finden. Ohne sich allgemein mit den Gründen zu beschäftigen, die zu diesem Phänomen führen, stellt sich die Frage, ob sich dies auf die Qualität der Lehrbewertung auswirkt: Sind die Bewertungen der Studierenden, die nach und nach aus dem Kurs "abwandern" mit denen, die bleiben, vergleichbar?

Wenn es Unterschiede gibt, sind bestimmte Gesetzmäßigkeiten zu finden? Sind diese Gesetzmäßigkeiten auf die Qualität der Lehre zu beziehen? Sind die gefundenen Muster selbst schon Ausdruck von Mängeln der Lehrqualität? Findet man Unterschiede zwischen den "Anwesenden" und den "Dropouts", wie soll man damit umgehen?

Es wäre nützlich, entweder vorhandene Erhebungen zur Lehrevaluation (soweit möglich) heranzuziehen oder gesonderte Untersuchungen vorzunehmen, um diese Fragen empirisch fundiert beantworten zu können.

### **Beteiligung an der Befragung (Rücklaufquote)**

Die Rücklaufquote ist ein Merkmal, das in den einzelnen Studien häufiger genannt wird. Dennoch liegt nur eine einzige Untersuchung zur Auswirkung unterschiedlicher Rücklaufquoten vor.

McBean und Lennox (1985) berichten von großen Anstrengungen im Herbstsemester 1982 an der Universität von Waterloo, eine überdurchschnittlich hohe Rücklaufquote zu erzielen. Untersucht wurde, welche Auswirkung unterschiedliche Rücklaufquoten für die Ergebnisse haben. Randomisierte Untergruppen wurden gebildet, die 80 Prozent, 68 Prozent, 50 Prozent und 25 Prozent Rücklauf simulierten.

Die wichtigsten Ergebnisse waren: Ist die Stichprobe viel kleiner (geringer als 25 Prozent) als die Ursprungsgruppe, so besteht Potential für eine erhebliche Datengenauigkeit, jedoch nur in Gruppen, die weniger als 30 Mitglieder haben. Ist die Gruppe dagegen größer als 30, so ist der Fehler bis zu einer Rücklaufquote von 50 Prozent nicht sehr groß. Die Autoren schließen daraus, daß in großen Gruppen (>30) eine Antwortrate von 50 Prozent eine genügend präzise Auskunft der Lehrbewertung gibt, während in kleinen Gruppen (<30) ungefähr 80 Prozent ihre Fragebogen abgeben müssen, um den gleichen Vertrauensgrad zu erreichen.

Problematisch an dieser Feststellung ist die Ausgangsstichprobe von randomisierten Subgruppen. Es müßte ebenfalls untersucht werden, ob die Ergebnisse von Studierenden, die Lehrbeurteilungen abgeben, anders ausfallen, als die von Studierenden, die sie nicht abgeben.

Ergänzend ist darauf hinzuweisen, daß Rindermann im Zuge der Entwicklung und Erprobung des „Heidelberger Inventars zur Lehrveranstaltungsevaluation (HILVE)“ bereits eine relativ hohe Genauigkeit und Stabilität von Lehrbewertungen bei einer Mindestzahl von zehn antwortenden Studierenden ermitteln konnte (vgl. Kapitel 8).

### **Ergebnisdarstellung: Profile und Verteilungen**

Aus der Forderung nach multidimensional gestalteten Fragebogen zur Lehrevaluation ergibt sich die Konsequenz einer differenzierten Darstellung unterschiedlicher Dimensionswerte. Diese einzelnen Werte sollten in keinem Fall zu einem bloßen Gesamtwert zusammengezogen werden, der für Vergleiche dienen soll. Bei einem solchen Verfahren bleibt nicht nur eine Unmenge an Informationen ungenutzt und gleiche Gesamtwerte können sich aus völlig unterschiedlichen Einzelergebnissen zusammensetzen, sondern es ist auch methodisch unzulässig, Dimensionen, die möglichst niedrig miteinander korrelieren, d.h. verschiedene Sachverhalte erfassen, einfach zu summieren.

Die Ergebnisse der Lehrbewertung sollten nach ihren einzelnen Dimensionen getrennt aufgeführt und als *Profile* dargestellt werden. Die einzelnen Markierungen, zwischen den Extrempolen abgetragen, werden miteinander verbunden und zeigen anschaulich positive und negative „Ausschläge“ in der Lehrevaluation einzelner Lehrender.

Außer methodischen Gründen sprechen noch Gründe praktischer Art für diese Form der Darstellung. Erstens muß dem Lehrenden oder der Lehrenden deutlich werden, wo genau seine/ihre „Schwächen“ liegen, um diese verbessern zu können. Zweitens erscheint es für den einzelnen/ die einzelnen Lehrenden motivierender, anstelle eines möglichen schlechten Gesamtergebnisses ein Profil zu er-

halten, das vermutlich nur in den seltensten Fällen keine positiven „Ausschläge“ enthält und damit auch auf „Stärken“ verweist.

Der Zweck der Datenerhebung hat nicht nur Auswirkungen auf die Wahl eines Instruments zur Lehrbewertung, sondern auch auf die Art der Ergebnisdarstellung. Wird angestrebt, Lehrenden eine möglichst detaillierte Rückmeldung über ihre Veranstaltung zu geben, so sind differenziertere und individuellere Auswertungen notwendig, als wenn ein allgemeiner Vergleich über Lehrende und Lehrveranstaltungen angestrebt wird.

### **Schlußfolgerungen**

*Es wäre an der Zeit, eine Untersuchungen unter Studierenden durchzuführen, die jene Gründe und Motive herauszufinden versucht, die dazu führen, eine Veranstaltung nicht weiter zu besuchen. Zeitmangel, verschobene Prioritätensetzung, Kurs wurde kurzfristig aus Neugier belegt, viele Antworten scheinen denkbar; darunter auch Urteile, die u.U. dazu führen müßten einen Kurs neu zu gestalten, weil sie sich auf Mängel des Kurses beziehen.*

*Bei der Auswertung von Fragebogen sollte eine akzeptable Rücklaufquote vorliegen. Dies wären nach Rindermann mindestens 10 Studierende. Man kann sich aber auch an den Ergebnissen von McBean und Lennox (1985) orientieren, die bei großen Kursen (>30) eine notwendige Rücklaufquote von 50 Prozent ermittelten, um keine Ergebnisverzerrung zu riskieren, bei kleinen Kursen (<30) sollte dagegen eine Rücklaufquote von 80 Prozent erreicht werden.*

*Nachdem ein sorgsam konstruiertes Instrument zur Datenerhebung verwandt wurde, sollte auf die Ergebnissdarstellung ähnlich viel Wert gelegt werden. Nur eine differenzierte Präsentation (z.B. in Profilform und mit Verteilungen) kann zu einer optimalen Nutzung der Daten führen. Allerdings hängt die Form und Detailliertheit der Datenpräsentation nicht zuletzt vom Zweck der Datenerhebung ab.*

### **7.5 Zeitpunkt und Frequenz der Beurteilung**

Der Zeitpunkt und die Häufigkeit studentischer Lehrbeurteilungen sind maßgeblich durch den Verwendungszweck, dem sie dienen sollen, zu bestimmen. Über „optimale“ Zeitpunkte und Frequenzen liegen zwar keine Untersuchungen vor, aber einige Überlegungen und Anhaltspunkte zu dieser Frage sollen hier genannt werden.

## **Zeitpunkt der Befragung**

Bevor nicht geklärt ist, ob Studierende, die bis zum Kursende im Unterricht verbleiben, gleiche Bewertungsmuster wie Kursabrecher und -abrecherinnen abgeben, erscheint es unzureichend, nur am Ende eines Kurses Daten zu erheben. Es sollte mindestens ein weiterer Erhebungszeitpunkt festgelegt werden.

Um einen günstigen Zeitpunkt zu finden, müßte geklärt werden, ab welcher Semesterwoche sich Studierende ein Bild von den Lehrleistungen ihres Dozenten/ihrer Dozentin bilden können und ob an diesem Zeitpunkt die Anwesenheitsquote noch hoch genug ist.

## **Frequenz der Befragung**

Ein weiterer Grund für Beurteilungen in der Semestermitte sind mögliche Verbesserungen für den weiteren Kursverlauf, da der Lehrende/die Lehrende das Feedback nutzen kann, um Veränderungen noch vorzunehmen.

Sturm (1991) unterscheidet dabei zwischen Prozeß- und Gesamtevaluation. Bei der Prozeßevaluation für Feedbackzwecke kann eine möglichst hohe Beurteilungsfrequenz sinnvoll sein, die z.B. auch durch am Ende einer Stunde abzugebende Kurzkomentare zu erreichen ist (siehe Ritter, 1992). Für eine Gesamtevaluation, z.B. für administrative Zwecke, können jährliche bzw. zweijährliche Bewertungen, die sich grundsätzlich aus mehreren Kursen zusammensetzen sollten, ausreichen, da davon auszugehen ist, daß Lehrfähigkeit eine relativ stabile Fertigkeit darstellt.

Betrachtet man die Gesamtevaluation über einen gewissen Zeitraum für einen Kurs, so ergibt sich wiederum eine Prozeßevaluation, die als Grundlage zu einer Neustrukturierung dienen kann.

## **Schlußfolgerungen**

*Insgesamt ist die vorzusehende Beurteilungsfrequenz letztlich vom Zweck der Veranstaltungskritik her zu bestimmen. So wäre es sinnvoll, bei neu konzipierten Kursen oder Veranstaltungen, die ein Lehrender/ eine Lehrende erstmals anbietet, bereits während des Kursverlaufs ein Feedback zu ermöglichen.*

*Diese Form von Feedback könnte z.B. im Sinne einer Prozeßevaluation einfache Instrumente verwenden, wie z.B. die Kurzkomentare, die Ritter (1992) vorschlägt. Bei der Einführung einer solchen Prozeßevaluation ist sowohl die Kapazität der Hochschule und ihre Änderungsbereitschaft als auch die Motivation der Studierenden zu beachten.*

*Bei bereits etablierten Kursen und erfahrenen Lehrenden erscheint eine abschließende Gesamtevaluation am Ende jedes Kurses ausreichend. Diese Rückmeldung sollte zwar relativ kurz und leicht auszuwerten sein, aber doch so gestaltet werden, daß die Lehrenden entsprechende Umgestaltungen ihrer Kurse vornehmen können.*

*Weiter sollte eine umfassendere Erhebung alle zwei bis drei Jahre erfolgen, die dann zusätzlich Elemente der Lehrorganisation umfassen müßte. Dies kann z.B. in Form von Lehrberichten geschehen und in diesem Rahmen veröffentlicht werden. Die jährliche Erstellung von Lehrberichten mit umfassender Evaluation könnte eine Überforderung darstellen (vgl. auch Kapitel 3).*

## **7.6 Zusammenfassende Übersicht: Empfehlungen**

Die vorgestellten Punkte sind gedacht als Ansatz zur Formulierung verbindlicher Kriterien für die Erstellung und Verwendung von Fragebogen der studentischen Veranstaltungskritik. Gerade im Bereich der Itemformulierung, der möglicherweise unterschiedlichen Vergleichstabellen und der Frequenz von Datenerhebungen sollten allerdings weitere Untersuchungen durchgeführt bzw. in der Praxis erprobt werden, um die Erkenntnislücken zu schließen.

Dennoch lassen sich aufgrund empirischer Untersuchungen und praktischer Erprobungen eine Reihe von Empfehlungen zusammenstellen, an denen sich die Entwicklung, der Einsatz und die Verwendung von Instrumenten zur Bewertung der Lehre richten sollten; sie sind in der nachfolgenden Übersicht zusammengestellt.

Die aufgeführten Kriterien und Empfehlungen beziehen sich auf Fragebogen zur Bewertung von Veranstaltungen und Veranstaltungsreihen, die so konstruiert sein sollen, daß sie einen Vergleich zwischen den Lehrenden ermöglichen. An sie sind höhere Ansprüche hinsichtlich Validität und Meßgenauigkeit zu stellen. Für eine direkte Rückmeldung an den Veranstaltungsleiter/ die Veranstaltungsleiterin am Ende eines Kurses reichen weitaus einfachere Instrumente aus.

Soll neben der Bewertung einzelner Veranstaltungen und Lehrenden auch eine Beurteilung der Lehre im gesamten Fachbereich vorgenommen werden, so sind zusätzliche Instrumente bzw. zusätzliche Fragen nötig, die stärker die Beurteilung der Lehrorganisation fokussieren.

Vor der Erhebung von Daten sollte stets deren Nutzung und Verwendung feststehen. Erst nachdem eine Entscheidung über die Richtung und den Zweck der Lehrevaluation getroffen ist, sollte die Festlegung auf ein Instrument oder eine Kombination verschiedener Verfahren getroffen werden.

## **Empfehlungen zur Entwicklung und Gestaltung von Instrumenten der Lehrevaluation durch Studierende**

### **(1) Abgrenzung verschiedener Dimensionen**

- ⇒ Fragebogen sind mehrdimensional zu konstruieren. Jede Dimension sollte von mehreren Items erfaßt werden.
- ⇒ Wichtige Dimensionen, die nicht ausgelassen werden sollten: Zuwendung, Fairneß von Prüfungen und Benotungen, Kommunikationsfähigkeit, Kurs- bzw. Stofforganisation, Anregung, Variabilität vs. Monotonie, Enthusiasmus und Kurswert.
- ⇒ Zusätzliche Aspekte erheben, die Lehrende oder Studierende interessieren (z.B. Kursschwierigkeit) oder für weitere wissenschaftliche Untersuchungen interessant sind (z.B. Geschlecht).
- ⇒ Faktoren erfassen, die einen möglichen Einfluß auf die Lehrbewertung haben, um sie in der Darstellung kontrollieren bzw. ausweisen zu können.

### **(2) Formulierung der Fragen und Items**

- ⇒ Möglichst von Beobachtbaren ausgehen. Soweit wie möglich konkrete Items verwenden und globale vermeiden.
- ⇒ Vergleichsmaßstab auch bei der Itemformulierung beachten. Die optimale Ausprägung eines Lehraspektes sollte am Anfang oder Ende der Ratingskala liegen.
- ⇒ In die Wertung nur Merkmale einbeziehen, für die Lehrende verantwortlich sind.
- ⇒ Vorab entscheiden, ob die unterschiedlichen Items unterschiedlich gewichtet werden sollen

### **(3) Gesamtbogen**

- ⇒ Itemzahl von Verwendungszweck abhängig. 15-35 Items empfehlenswert.
- ⇒ Bei der Beurteilungsskala 5er Punktsystem sinnvoll, aber auch 3-7 Punkte sind je nach Zweck vertretbar.

### **(4) Anweisung (Instruktion) und Durchführung**

- ⇒ Anweisung (Instruktion) standardisieren, sowohl hinsichtlich der Zwecksetzung als auch des Bewertungsbezuges (Anker).
- ⇒ Studentische Veranstaltungskritiken anonym durchführen.

### **(5) Auswertung**

- ⇒ Aspekte, die außerhalb der Lehrdimensionen liegen (z.B. Kursschwierigkeit), sollten nicht in die Lehrbewertung eingehen.
- ⇒ Faktoren, die einen Einfluß auf die Lehrbewertung haben, müssen entsprechend konkretisiert (ausgewiesen oder verrechnet) werden. Nach dem heutigen Kenntnisstand wäre dies hauptsächlich der Faktor „Motivation und Interesse am Kursthema“.

### **(6) Vergleichstabellen**

- ⇒ Vergleichstabellen existieren noch nicht. Ein erster Ansatz ist von Diehl (Gießen) geplant.
- ⇒ Bei der Erstellung ist es sinnvoll, nach verschiedenen Veranstaltungsarten, u.U. auch nach der Semesterzahl der Studierenden, nach Kursgröße, vielleicht auch nach Fächern zu unterteilen.

### **(7) Ergebnisdarstellung**

- ⇒ Drop-out-Quote der Veranstaltung beachten.
- ⇒ Eine grobe Richtlinie für den Rücklauf stellen mind. 10 Studierende dar. Ansonsten sollte die Rücklaufquote der ausgegebenen Bogen in kleinen Kursen (< 30) mindestens 80 Prozent betragen, in großen Kursen (> 30) mindestens 50 Prozent.
- ⇒ Daten nicht als simple Gesamtnoten darstellen, sondern differenziert (z.B. als Profile) ausweisen.
- ⇒ Nicht nur Mittelwerte, sondern auch Streuungen bzw. Verteilungen einbeziehen und ausweisen.

### **(8) Frequenz der Beurteilung**

- ⇒ Bei neuen Kursen/ Lehrenden sollten *Kurzrückmeldungen* bereits während des Kursverlaufs erfolgen (Prozessevaluation).
- ⇒ Bei etablierten Veranstaltungen ist eine Gesamtevaluation am Ende bzw. in der Mitte der Veranstaltung ausreichend.
- ⇒ Alle 2-3 Jahre sollte eine ausführliche Evaluation, die auch die Lehrorganisation beinhaltet, durchgeführt werden (z.B. in Form eines Lehrberichts).

Neben den genannten relativ konkreten Kriterien und Empfehlungen, die sich direkt auf die Entwicklung von Fragebogen der studentischen Lehrveranstaltungs-kritik beziehen, gibt es einige übergreifende Punkte, die wenigstens kurz erwähnt werden sollen.

So ist es für die Evaluation der Lehrqualität an einer Hochschule auf Dauer sinnvoll, einen Pool von Items zu sammeln, aus dem je nach Bedarf für einzelne Lehrveranstaltungen ein spezifischer Bogen für dessen Beurteilung zusammengestellt werden kann. Dieser Empfehlung liegt der Gedanke des Cafeteria-Prinzips zugrunde. Dabei werden für die Lehrbewertung einem standardisierten Kern, der für die Beurteilung aller Lehrveranstaltungen verwendet wird, fach- oder kursspezifische Fragen hinzugefügt. Dies hat den Vorteil, daß man sowohl ein erprobtes als auch ein flexibles Instrument zu Verfügung hat.

Des weiteren werden Verfahren eher akzeptiert, wenn die Lehrenden sie selbst wenigstens zum Teil den Bedürfnissen ihrer Veranstaltung und ihrem Informationsbedarf anpassen können.

Das Thema Akzeptanz der Lehrevaluation durch die Lehrenden spielt auch bei der Art und Weise der Datenauswertung und ihrer Präsentation eine Rolle. Der vorliegende Bericht geht vor allem auf die Konstruktionskriterien von Instrumenten studentischer Veranstaltungskritik ein. Dabei werden auch Punkte der Auswertung genannt. Nicht näher wird jedoch der Bereich der Umsetzung und Durchführung von Projekten der Lehrevaluation beschrieben. Aus den bisher gewonnen Erfahrungen erscheint es sinnvoll, bereits vor der Erhebung eine gewisse Akzeptanz der Verfahren zu erreichen. Noch wichtiger ist allerdings zu klären, welchem Zweck die Daten dienen, wie sie ausgewertet und veröffentlicht werden sollen.

Die Beurteilung von Lehrveranstaltungen kann auch Elemente einer Personalbeurteilung haben oder zumindest so aufgefaßt werden. Damit ist sie ein sensibles Instrument, mit dem in einer gewissen Rücksichtnahme verfahren werden sollte. Rücksicht ist zu nehmen auf die Beurteilten und ihrem Bedürfnis nach Fairness, manchmal auch nach Diskretion, aber auch auf die Studierenden und deren Bedürfnis nach Information und Transparenz sowie nach Veränderungen.

## 8 Instrumente zur Lehrevaluation in Deutschland: Beispiele

In Deutschland sind in den letzten Jahren einige Instrumente der studentischen Veranstaltungskritik konstruiert und erprobt worden, die einen allgemeineren und beispielhaften Anspruch haben. Sie befinden sich zum Teil bereits in einem Entwicklungsstadium, in dem es ohne weiteres möglich ist, sie in beliebigen Kursen einzusetzen. Der Empfehlung folgend, sich bei der Entwicklung von Fragebogen zur Lehrbewertung an methodisch geprüften und praktisch erprobten Instrumenten zu orientieren bzw. diese einzusetzen, werden im folgenden einige angeführt. Da es nicht möglich ist, sie im Detail genauer darzustellen, soll dieses Kapitel nur dazu dienen, einen Überblick über entsprechende Verfahren zu erlangen und weiterführende Literaturverweise anzubieten.

Professor Diehl in Gießen hat bereits in den siebziger Jahren einen Fragebogen zur studentischen Veranstaltungskritik erarbeitet (Diehl et al., 1977). Dieser wurde weiterentwickelt und liegt inzwischen als Version für Vorlesungen und als Version für Seminare mit Referaten vor (Diehl, 1994). Besonders interessant ist dabei, das sowohl das Anleitungshandbuch, als auch die Fragebogen auf Diskette vorliegen und beliebig vervielfältigt werden dürfen. Ebenfalls auf Diskette kann man das Auswertungsprogramm erhalten (vgl. Abschnitt 8.1 und 8.2).

Des weiteren plant Diehl eine Datenbank mit entsprechenden Evaluationsergebnissen zu erstellen, die eine spätere Vergleichbarkeit solcher Ergebnisse ermöglichen sollen.

Ein weiteres Evaluationsprojekt, das zu Beginn vor allem die Überprüfung möglicher unerwünschter Einflußfaktoren auf die studentischen Urteile zur Lehre zum Ziel hatte, ist das von Professor Kromrey an der Universität Bochum. Eine ganze Reihe von Veröffentlichungen, die sich mit den methodischen Problemen der Lehrevaluation auseinandersetzen, wurden von Kromrey in diesem Zusammenhang publiziert (vgl. Kromrey, 1995a;b, 1994 a;b, 1993).

Auf Grundlage der in Bochum gesammelten Daten wurde ein speziell für die studentische Veranstaltungskritik entwickelter Fragebogen gestaltet, der seit einigen Semestern an der FU Berlin erprobt wird (vgl. Abschnitt 8.3 und 8.4).

Rindermann und Amelang sind bisher die einzigen, die das von ihnen entwickelte Instrument, das „Heidelberger Inventar zur Lehrveranstaltungsevaluation“ (HILVE) als käufliches Verfahren publiziert haben (Rindermann & Amelang, 1994a;b). Seine Entwicklung hat sich am stärksten an testtheoretischen Gütekriterien orientiert (vgl. Abschnitt 8.5).

Rindermann (1996a) hat in diesem Zusammenhang auch Untersuchungen zur Brauchbarkeit und Gültigkeit vorgenommen. Wie die meisten neu veröffentlichten

ten Instrumente leidet der Fragebogen noch unter einigen „Kinderkrankheiten“ (für eine kritische Auseinandersetzung vgl. auch Kromrey, 1995c).

Am Interdisziplinären Zentrum für Hochschuldidaktik (IZHD) in Bielefeld wurde von Dr. Webler und seinen Mitarbeiterinnen und Mitarbeitern ein maschinenlesbarer Fragebogen zur studentischen Lehrbeurteilung entwickelt, der an der Universität Bielefeld nahezu flächendeckend eingesetzt wurde (vgl. Abschnitt 8.6). Dieser Fragebogen wird zum Teil auch im Rahmen der Datensammlung für die verschiedenen Projekte des IZHD zur Erstellung von Lehrberichten verwendet (Webler, 1993, 1992, Webler et al., 1993).

Das Evaluationsprojekt zur Lehrbeurteilung in Mannheim wurde aufgrund eines Antrags des Mannheimer AStAs Ende 1992 begonnen und vom Landesministerium Baden-Württemberg gefördert; seine Leitung lag in den Händen von Daniel. Das Projekt hatte zum Ziel, eine Vollerhebung im betriebswirtschaftlichen Grundstudium (BWL I und BWL II) durchzuführen (vgl. Daniel, 1994, 1995a-d, Daniel et al., 1995, Daniel und Hornbostel, 1993, ). Später erfolgten weitere Erhebungen zur Lehrbewertung in den Geisteswissenschaften.

Das Projekt wurde von einer Senatskommission für Lehre betreut, die die Endfassung der jeweiligen Fragebogen verabschiedete. Folge dieser Konstellation sind ausgesprochen umfangreiche Fragebogen, die vor allem zu wissenschaftlichen Zwecken oder zur Itemrequirierung geeignet erscheinen (vgl. Abschnitt 8.7).

Dies sind bei weitem nicht alle in Deutschland entwickelten Instrumente der studentischen Veranstaltungskritik. Beispielhaft wurden hier nur einige aufgeführt, um einen Eindruck darüber zu vermitteln, in welche Richtung man sich orientieren kann. Die Wissenschaftler, die durch entsprechende Veröffentlichungen auf sich aufmerksam gemacht haben, wurden gebeten, das von ihnen entwickelte Verfahren selbst vorzustellen.

Die Darstellung der Fragebogen beginnt mit einer kurzen Beschreibung des Instruments. Dabei soll deutlich werden, um welche Art von Verfahren es sich handelt und wie es konstruiert wurde. Weiter interessieren die zu messenden Dimensionen und die Zielgruppe, für die die Verwendung des Instruments gedacht ist. Außerdem werden Hinweise zur Durchführung und Auswertung der Fragebogen gegeben.

Um die Qualität des Fragebogens beurteilen zu können, sind neben der Art der Fragebogenerstellung auch die Kennzahlen der Gütekriterien aussagekräftig. Diese können mit einer ganzen Reihe mehr oder weniger aufwendiger Verfahren ermittelt werden. Es wurden nicht bei allen Verfahren die entsprechenden Gütekriterien ermittelt. Diese Art der Qualitätsbestimmung wird auch erst dann wirklich relevant, wenn es beim Vergleich zwischen Lehrenden um tatsächliche Auswirkungen gehen wird.

## **8.1 Fragebogen zur Beurteilung von Vorlesungen (VBVOR) von J. M. Diehl**

### **I. Kurze Beschreibung des Instruments**

Der VBVOR ist ein Fragebogen zur studentischen Beurteilung von Hochschulveranstaltungen, in denen die Vermittlung des Stoffes ausschließlich durch den Dozenten bzw. die Dozentin erfolgt (Veranstaltungen vom Typ Vorlesung).

Das aus 16 Items sowie einigen Zusatzfragen bestehende Instrument stellt eine neukonstruierte Kurzform des von Diehl & Kohr (1977) entwickelten VBPSYCH (40 Items) dar. Der ursprüngliche Itempool wurde aus einer Sammlung (freier) studentischer Aussagen zu Vorlesungen im Fach Psychologie gewonnen. Durch Faktorenanalysen und nachfolgende Itemanalysen entstanden dann die Lang- bzw. Kurzform des Instrumentes.

Mit den vier Skalen zu je vier Items (und vierstufigem Antwortformat) werden folgende Veranstaltungsaspekte erfaßt:

- [1] Relevanz und Nützlichkeit der Veranstaltungsinhalte
- [2] Verhalten des Dozenten (der Dozentin) gegenüber den Veranstaltungsteilnehmern und -teilnehmerinnen.
- [3] Angemessenheit von Schwierigkeit und Umfang der Veranstaltungsinhalte
- [4] Methodik und Aufbau der Veranstaltung.

Vom Fragebogen existiert eine männliche und eine weibliche Form, bei denen vom Dozenten bzw. der Dozentin gesprochen wird.

Der Fragebogen wird bereits an mehreren Psychologie-Fachbereichen der Bundesrepublik eingesetzt (teilweise von Seiten der Fachschaften). Er hat sich gleichermaßen in anderen Fachrichtungen bewährt (z.B. Wirtschaftswissenschaften, Pädagogik, Medizin).

### **II. Durchführung und Auswertung**

Die Beantwortung der 16 Items sowie der vorgegeben Zusatzfragen dauert 5-10 Minuten. Für die Eingabe und Auswertung des Daten existiert ein DOS-Programm, das aufgrund seiner einfachen Bedienung und Eingabemasken auch von EDV-Ungeübten schnell gehandhabt werden kann. Nach Einarbeitung liegt die Eingabedauer für einen Fragebogen bei 1-2 Minuten.

### III. Ergebnisdarstellung

Nach Dateneingabe liefert das Programm unmittelbar eine achtseitige Auswertung der Veranstaltung, die folgendes enthält:

- Itemtexte (skalenweise geordnet) mit Antwortverteilungen
- Durchschnittliche Skalenwerte
- Verteilung der Skalenwerte (grafisch)
- Interkorrelation der Skalen
- Verteilung der Antworten bei den Zusatzfragen.

Die Daten einer Veranstaltung werden zugleich als ASCII-File abgespeichert und können dadurch mit Statistik-Programmen weiter analysiert werden. Das Anlegen einer Datenbank mit Veranstaltungen erfolgte bisher (aus Zeitgründen) nicht.

### IV. Gütekriterien

Die Skalen besitzen faktorielle Gültigkeit. Die Item-Trennschärfen sowie die innere Konsistenz der Skalen können als gut bezeichnet werden. Die Werte von Cronbachs alpha sind im einzelnen: Skala 1 (.84), Skala 2 (.81), Skala 3 (.81) und Skala 4 (.89).

### V. Angestrebte Weiterentwicklung

Es ist geplant, das Auswertungsprogramm derart zu erweitern, daß es auch benutzer-formulierte Zusatzfragen analysieren kann.

### VI. Besondere Hinweise zur Verwendung des Verfahrens

Manual und Fragebogenexemplare sind als ASCII-Dateien auf einer Diskette erhältlich, die auch das Auswertungsprogramm enthält. Diese Diskette darf beliebig kopiert und weitergegeben werden. Anwender(innen) des VBVOR werden gebeten, sich beim Autor registrieren zu lassen. Dies soll die Sammlung von Evaluationsdaten für eine (spätere) VBVOR-Datenbank erleichtern.

## **8.2 Fragebogen zur Beurteilung von Seminaren mit Referaten (VBREF) von J. M. Diehl**

### **I. Kurze Beschreibung des Instruments**

Der VBREF ist eine Fragebogen zur studentischen Beurteilung von Hochschulveranstaltungen, in denen die Stoffvermittlung (überwiegend) durch studentische Referate erfolgt. Das aus 30 Items sowie Zusatzfragen bestehende Instrument stellt eine noch nicht veröffentlichte Neukonstruktion dar. Der ursprüngliche Itempool wurde aus einer Sammlung (freier) studentischer Aussagen zu Referatveranstaltungen im Fach Psychologie gewonnen. Durch Faktorenanalysen und nachfolgende Itemanalysen entstand anschließend die endgültige Form des Fragebogens.

Mit den sechs Skalen zu je fünf Items (und vierstufigem Antwortformat) werden folgende Veranstaltungsaspekte erfaßt:

- [1] Qualität der Referatsvorträge
- [2] Verhalten des Dozenten (der Dozentin) gegenüber den Veranstaltungsteilnehmern und -teilnehmerinnen.
- [3] Angemessenheit von Schwierigkeit und Umfang der Veranstaltungsinhalte
- [4] Relevanz und Nützlichkeit der Veranstaltungsinhalte
- [5] Umfang der Frage- und Diskussionsmöglichkeiten
- [6] Auswahl und Zusammenhang der Referatsthemen

Vom Fragebogen existiert eine männliche und eine weibliche Form, bei denen vom Veranstaltungsleiter bzw. der Veranstaltungsleiterin gesprochen wird. Der Fragebogen wird bereits an mehreren Psychologie-Fachbereichen der Bundesrepublik eingesetzt (teilweise von Seiten der Fachschaften). Er hat sich gleichermaßen in anderen Fachrichtungen bewährt. (z.B. Ernährungswissenschaften, Pädagogik).

### **II. Durchführung und Auswertung**

Die Beantwortung der 30 Items sowie der vorgegeben Zusatzfragen dauert etwa 10 Minuten. Für die Eingabe und Auswertung der Daten existiert ein DOS-Programm, das aufgrund seiner einfachen Bedienung und Eingabemasken auch von EDV-Ungeübten schnell gehandhabt werden kann. Nach Einarbeitung liegt die Eingabedauer für einen Fragebogen bei etwa zwei Minuten.

### III. Ergebnisdarstellung

Nach Dateneingabe liefert das Programm unmittelbar eine elfseitige Auswertung der Veranstaltung, die folgendes enthält:

- Itemtexte (skalenweise geordnet) mit Antwortverteilungen,
- Durchschnittliche Skalenwerte,
- Verteilung der Skalenwerte (grafisch),
- Interkorrelation der Skalen,
- Verteilung der Antworten bei den Zusatzfragen.

### IV. Gütekriterien

Die Skalen besitzen faktorielle Gültigkeit. Die Item-Trennschärfen sowie die innere Konsistenz der Skalen können als gut bezeichnet werden. Die Werte von Cronbachs alpha sind im einzelnen: Skala 1 (.84), Skala 2 (.81), Skala 3 (.88), Skala 4 (.74), Skala 5 (.85) und Skala 6 (.79).

### V. Angestrebte Weiterentwicklung

Es ist geplant, das Auswertungsprogramm derart zu erweitern, daß es auch benutzerformulierte Zusatzfragen analysieren kann.

### VI. Besondere Hinweise zur Verwendung des Verfahrens

Manual und Fragebogenexemplare sind als ASCII-Dateien auf einer Diskette erhältlich, die auch das Auswertungsprogramm enthält. Diese Diskette darf beliebig kopiert und weitergegeben werden. Anwender(innen) des VBVOR werden gebeten, sich beim Autor registrieren zu lassen. Dies soll die spätere Sammlung von Evaluationsdaten für eine VBREF-Datenbank erleichtern.

Diehl, J.M. & Kohr, H.U. (1977). Entwicklung eines Fragebogens zur Beurteilung von Hochschulveranstaltungen im Fach Psychologie. *Psychologie in Erziehung und Unterricht*, 24, 61-75.

Diehl, J.M. (1994). Fragebogen zur studentischen Evaluation von Hochschulveranstaltungen. Manual zum VBVOR und VBREF. Fachbereich 06 Psychologie (Abteilung Methodik), Universität Gießen.

### **8.3 Evaluation der Lehre an der Ruhr-Universität Bochum: Fragebogen für Vorlesungen (I. Studierende, II. Dozent(inn)en) von H. Kromrey**

#### **I. Kurze Beschreibung des Instruments**

Befragt wurden alle Hörer/innen in Vorlesungen; erhoben wurde in einer Stichtagswoche ("Normalwoche" in der Mitte des Semesters) in allen Vorlesungen einer Fakultät. Parallel füllte jede Lehrperson einen Dozent(inn)enFragebogen aus.

Der Studierendenbogen enthält Fragen zur Stellung der/des Studierenden im Studiengang, zu den Teilnahmegründen sowie zu den Erwartungen an die Vorlesung, zu studentischen Beurteilungen über Didaktik, Stoffmenge und Schwierigkeitsgrad, über Vorkenntnisse und Selbsteinschätzung des Lernerfolgs, den eigenen Arbeitsaufwand, Rahmenbedingungen der Vorlesung sowie das Auftreten und Verhalten der Lehrperson. Im Dozent(inn)enbogen sind ähnliche Themenbereiche aus der Perspektive der Lehrperson einzuschätzen.

Die Items wurden anhand der Durchsicht hochschuldidaktischer Literatur sowie bereits vorliegender Fragebogen zusammengestellt; Bewertungen werden auf einer 5-Punkte-Ratingskala abgefragt. Das endgültig eingesetzte Instrument ist das Ergebnis mehrerer Pretest-Durchgänge: Studierende unterschiedlicher Fakultäten wurden gebeten, den Bogen auszufüllen; anschließend wurde in einem mündlichen Nach-Interview Frage für Frage die Semantik aus dem Verständnis des Ausfüllenden sowie das Ausfüll-Verhalten erfaßt. Bei Verständnis-Differenzen wurden die Befragten um (für sie eindeutigere) Alternativ-Formulierungen gebeten.

Neben dem Vorlesungsbogen existieren spezielle Instrumente für Seminare sowie für Veranstaltungen mit Übungscharakter. Sie unterscheiden sich vom Vorlesungsbogen insbesondere im didaktischen Teil sowie bei den Fragen nach dem studentischen Teilnahmeverhalten.

#### **II. Hinweise zur Durchführung und Auswertung**

Der Studierendenbogen ist im Einvernehmen mit der jeweiligen Lehrperson ca. 20-30 Minuten vor Schluß der Vorlesung von einer studentischen Hilfskraft zu verteilen, von den Teilnehmern unmittelbar auszufüllen und anschließend von der Hilfskraft wieder einzusammeln. Die Lehrperson verläßt die Veranstaltung und füllt den eigenen Bogen in ihrem Büro aus. Dadurch soll die Anonymität der Befragung nicht nur gesichert, sondern auch sichtbar gemacht werden. Für die Gültigkeit der Befragungsergebnisse ist es unabdingbar, daß möglichst alle Teilnehmer den Bogen ausfüllen; ein Austeilen der Bogen, verbunden mit dem Appell, ihn ausgefüllt in der nächsten Woche wieder mitzubringen oder ihn in einem

"Evaluationsbüro" abzugeben, führt zu extrem selektivem Rücklauf und inhaltlich verzerrten Resultaten.

Ausgewertet wird der Studierendenbogen a) je Vorlesung als Rückmeldung an die jeweilige Lehrperson sowie b) kumuliert über alle Vorlesungen als zusammenfassende Information für die Fakultät. Der Dozent(inn)enbogen kann nur fakultätsweise ausgewertet werden.

### III. Ergebnisdarstellung

Die Daten werden schnellstmöglichst erfaßt, so daß eine Häufigkeitsauszählung noch vor Ende des Semesters der jeweiligen Lehrperson zur Verfügung gestellt werden kann. Die Auszählung orientiert sich an den Kategorien des Fragebogens und enthält neben den (absoluten und relativen) Häufigkeiten der betreffenden Veranstaltung als Vergleich die kumulierten Werte der Fakultät. Da sich gezeigt hat, daß die studentischen Urteile innerhalb der Vorlesungen sehr heterogen ausfallen, wird auf die Berechnung von Durchschnittswerten verzichtet, d.h. es wird die informationsreichere Dokumentation der Verteilung der Antworthäufigkeiten beibehalten.

### IV. Gütekriterien

In abschließenden Auswertungen wurde der Einfluß von Bias-Variablen überprüft. Ergebnis: Je globaler studentische Beurteilungen erfragt werden (Beispiel: Hat die Vorlesung bisher Ihre Erwartungen insgesamt erfüllt?), desto geringer ist ihr Bezug zum Lehrverhalten der jeweiligen Dozentin bzw. des jeweiligen Dozenten. Im Vordergrund stehende Einflußgrößen sind: das "Image" des zu vermittelnden Stoffs (beliebt/unbeliebt), der Grad der Freiwilligkeit des Veranstaltungsbesuchs (Pflichtveranstaltung mit Klausur versus Wahlveranstaltung), das "mitgebrachte" persönliche Interesse der/des Studierenden. Werden Durchschnittsbeurteilungen je Lehrveranstaltung gebildet, erklären diese - von der Lehrperson kaum beeinflussbaren - Faktoren den größten Teil der Variation zwischen den Veranstaltungen.

### V. Besondere Hinweise zur Verwendung des Verfahrens

Teilnehmerbeurteilungen (insbesondere globale Evaluations-Statements) eignen sich zwar als Indikatoren des Stimmungsbilds in der Veranstaltung, nicht jedoch als Maße der Qualität der Lehre und führen bei einem direkten Vergleich zwischen Lehrveranstaltungen zwangsläufig zu Fehlschlüssen. Teilnehmerbefragungen sind jedoch - bei großen Lehrveranstaltungen - ein durch andere Instrumente nicht zu ersetzendes Verfahren der Rückmeldung an die Lehrperson (nur

in dieser Funktion ist im übrigen die studentische Veranstaltungskritik in früheren Jahrzehnten diskutiert worden).

## VI. Angestrebte Weiterentwicklung

Aufbauend auf den Erfahrungen mit Hörer(innen)befragungen an der Ruhr-Universität Bochum (über 10.000 ausgewertete Bögen) sowie vergleichbaren Ansätzen in anderen Universitäten wird ein explizit auf die Funktion "studentische Rückmeldung" optimiertes Instrument entwickelt und seit drei Semestern teilweise am Soziologischen Institut der Freien Universität Berlin eingesetzt.

### **8.4 Fragebogen zur studentischen Rückmeldung in Lehrveranstaltungen von H. Kromrey**

Eingangsbefragung und Befragung zur abschließenden Evaluation

#### I. Kurze Beschreibung des Instruments

Ziel der Befragung ist a) die Bereitstellung von Informationen an die Lehrperson über die Teilnehmer in ihrer Veranstaltung, b) die Rückmeldung der Lehrbeurteilungen durch die Studierenden am Semesterschluß.

Entsprechend enthält der Bogen zur Eingangsbefragung ausschließlich Angaben über die ausfüllende Person: Stellung im Studium, zeitliche Belastung durch Erwerbstätigkeit und andere außeruniversitäre Verpflichtungen, Besuchsgründe, Erwartungen/Befürchtungen, Informationsquellen für die Entscheidung zum Besuch der Veranstaltung, subjektive Anforderungen an die Didaktik, Vorkenntnisse, beabsichtigte Teilnahmefrequenz, geplanter Arbeitsstil und Arbeitsaufwand für die Veranstaltung.

Der Bogen zur abschließenden Evaluation greift einen Teil dieser Fragen wieder auf (Auswirkungen der außeruniversitären Belastungen auf das Studium, erfüllte/nicht erfüllte Erwartungen, bestätigte/nicht bestätigte Befürchtungen, Beurteilung der Didaktik) und ergänzt die Items um weitere evaluative Komponenten: Beurteilung der Veranstaltungsstruktur sowie Arbeitsmaterialien, der Beratung/ Betreuung durch die Lehrperson, des Auftretens der Lehrperson, des Stoffumfangs und seines Schwierigkeitsgrads, des wahrgenommenen eigenen Lernerfolgs sowie der Abstimmung der Veranstaltung mit anderen Studieninhalten.

Die Items wurden anhand der Durchsicht der Didaktik-Literatur, eigener vorhergehender Lehrevaluations-Befragungen sowie bereits vorliegender Fragebogen

zusammengestellt; Bewertungen werden überwiegend auf 5-Punkte-Ratingskalen abgefragt.

Das endgültig eingesetzte Instrument ist das Ergebnis mehrerer Pretest-Durchgänge: Studierende wurden gebeten, den Bogen auszufüllen; anschließend wurde in einem mündlichen Nach-Interview Frage für Frage die Semantik aus dem Verständnis des Ausfüllenden sowie das Ausfüll-Verhalten erfaßt. Bei Verständnis-Differenzen wurden die Befragten um (für sie eindeutigere) Alternativ-Formulierungen gebeten.

## II. Hinweise zur Durchführung und Auswertung

Die Rückmeldung soll die Lehrperson in die Lage versetzen, ihre Veranstaltung nicht nur langfristig den Erfordernissen des Studiengangs anzupassen, sondern sie auch kurzfristig möglichst "zielgruppengerecht" durchzuführen. Zu diesem Zweck müssen die Informationen über die Teilnehmer möglichst schnell zur Verfügung stehen.

Die Eingangsbefragung wird daher bereits in der zweiten Veranstaltung des Semesters durchgeführt (Verteilen des Fragebogens ca. 20-30 Min. vor Schluß, sofortiges Ausfüllen und anschließendes Einsammeln), sofort für die EDV aufbereitet und ausgezählt. Spätestens in der vierten Semesterwoche kann die Lehrperson die Ergebnisse mit den Studierenden diskutieren und ggf. noch Anpassungen des Lehrprogramms vornehmen.

Die abschließende Evaluation erfolgt in der vorletzten Semesterwoche (Durchführung wie bei der Eingangsbefragung). Da die Studierenden gebeten werden, sowohl auf dem Eingangs- wie auf dem Abschlußbogen eine anonyme Personenkennung einzutragen (die beiden letzten Buchstaben des Vornamens der Mutter sowie des Vaters, die ersten beiden Ziffern des eigenen Geburtsdatums), können die Daten der Anfangs- und Abschlusserhebung personenweise zusammengeführt werden. So ist in der Gesamtauswertung für die Lehrveranstaltung nicht nur ein Vergleich der Häufigkeitsverteilungen (Anfang / Schluß) möglich, sondern auch die individuelle Veränderung des Meinungs-bildes nachvollziehbar.

## III. Ergebnisdarstellung

Die Daten werden veranstaltungsweise aufbereitet und ausgezählt sowie darüber hinaus auf Institutsebene kumuliert. Dadurch sind auch veranstaltungsübergreifende Auswertungen (z.B. Cluster- und Faktorenanalysen) möglich.

#### IV. Gütekriterien

Über die mehrfach pretest-gestützte Gestaltung des Fragebogens und der Items (vgl. Punkt I) hinausgehend wurden spezifische Gütekriterien-Analysen nicht durchgeführt. Bei dem Fragebogen handelt es sich nicht um ein "Meß"-Instrument (im engeren Sinne) zur Ermittlung von "Lehrqualität", sondern um einen breit angelegten - in der Diskussion mit den Lehrveranstaltungsteilnehmern zu validierenden - Informationsinput für die Qualitätsentwicklung und Qualitätssicherung in der Lehre.

#### V. Angestrebte Weiterentwicklung

Keine. Die Befragungen werden je nach Wunsch der Lehrperson als "Einmal-Erhebungen" mit möglichst individuellem Zuschnitt auf den eigenen Informationsbedarf durchgeführt oder in mehreren Semestern wiederholt.

#### VI. Besondere Hinweise zur Verwendung des Verfahrens

Die Akzeptanz dieser Form von studentischer Rückmeldung ist bisher sowohl auf seiten der Lehrenden als auch der Studierenden auf große Akzeptanz gestoßen. Mehr als 75 Prozent der ausfüllenden Studierenden plädieren für die Wiederholung solcher Befragungen. In Nachfolgediskussionen mit Lehrenden haben diese die Verwertbarkeit und den Nutzen der erhaltenen Informationen als hoch eingeschätzt. Allerdings ist der damit verbundene Arbeitsaufwand recht hoch. Für kleine Lehrveranstaltungen empfiehlt sich daher eher der Rat an die Lehrenden, das Thema Lehre und Didaktik unmittelbar mit den Teilnehmern (durchaus angeleitet durch die Fragen der Rückmeldebögen) zu diskutieren.

H. Kromrey: Studentische Vorlesungskritik. Empirische Daten und Konsequenzen für die Lehre. In: Soziologie, Heft 1/1993, S. 39-56.

H. Kromrey: Wie erkennt man "gute Lehre"? Was studentische Vorlesungsbefragungen (nicht) aussagen. In: Empirische Pädagogik, Heft 2/1994, S. 153-168.

H. Kromrey: Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In: P. Ph. Mohler (Hg.): Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung, Münster 1995 (2.), S. 105-128.

## 8.5 Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE) von H. Rindermann und M. Amelang

### I. Beschreibung des Instruments und der Konstruktion

Der HILVE stellt ein multidimensionales Instrument zur Evaluation von Lehrveranstaltungen dar. Auf 14 Skalen (37 vorgegebene Items) können Studierende differenziert zur Lehrveranstaltung und zur Lehre des Dozenten Stellung nehmen. Objekte der Beurteilung sind das *didaktische Lehrverhalten des Dozenten* (z.B. Die Dozentin/Der Dozent kann Kompliziertes verständlich machen), das *soziale Lehrverhalten* (z.B. Die Dozentin/Der Dozent ist kooperativ), *Veranstaltungsaspekte* (Anforderungen, Interessantheit der Veranstaltung), die *studentischen Beiträge* (Referate, Beteiligung), die *studentische Vor- und Nachbereitung* der Veranstaltung sowie das *Veranstaltungsthema*. In zwei Skalen (quantitativer und qualitativer *Lernumfang* sowie *Allgemeinbeurteilung*) können der Erfolg und summativ die Qualität der Veranstaltung eingeschätzt werden.

Die einzelnen Skalen umfassen jeweils zwei bis vier vorgegebene Items. Die Antwortskala variiert zwischen trifft nicht zu (1) und trifft zu (7). Fünf freie Items erlauben der Lehrkraft, veranstaltungsspezifische Fragen zu stellen. Diese können aus einer angebotenen Itemliste ausgewählt oder selbständig formuliert werden. Offene Fragen (was ist besonders positiv?, was weniger?, Verbesserungsvorschläge?, Anmerkungen zum Instrument) ermöglichen den Veranstaltungsteilnehmern, frei formulierte Kommentare zu schreiben. Diese offenen Fragen sollen helfen, durch Items des Instruments nicht abgedeckte, aber subjektiv bedeutsame Aspekte erfassen zu können. Zudem können den Verbesserungsvorschlägen Anregungen für die Seminargestaltung entnommen werden.

Frei wählbare Items und die fakultativen Blöcke *Referate*, *Beteiligung* und *Diskussion* erlauben ein veranstaltungsadaptives Vorgehen. Das Instrument ist für alle gängigen Veranstaltungsformen geeignet. Nicht geeignet ist es für Praktika (z.B. im Labor), Seminare mit überwiegender Gruppenarbeit oder reine Referatvorlesestunden.

Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation wurde bisher in sozialwissenschaftlichen (Psychologie, Soziologie, Pädagogik, Ethnologie etc.), sprachwissenschaftlichen (Romanistik) und medizinischen Studienfächern an drei verschiedenen Hochschulen eingesetzt (N=6.400 Studierende). Es kann sowohl von Studierenden als auch von geschulten Fremdgutachtern eingesetzt werden.

Das Inventar wurde anhand bewährter Testkonstruktionsverfahren erstellt. Ausgangspunkt der Itemauswahl waren Befragungen von Lehrenden und Studierenden, anhand welcher Kriterien sie die Lehre beurteilen würden und welche Aspekte aus ihrer Sicht gute Lehre charakterisieren würde. Darüber hinaus wurden englisch- und deutschsprachige Lehrinventare als weitere Itemgenerierungsquelle

herangezogen (z.B. Marshs SEEQ). Dieser Itempool wurde nach erhobenen Veranstaltungsbeurteilungen durch statistische Analysen (Objektivität, Reliabilität, Dimensionalität etc.) und unter Zuhilfenahme von inhaltlichen Kriterien (Feedbackrelevanz, Multiperspektivität der Lehre) auf 37 Items des HILVE reduziert. Zum Schluß wurde der HILVE Lehrenden und Fachschaften zur Überprüfung der studienfachspezifischen Geeignetheit vorgelegt. Das Konstruktionsverfahren kann als iterativ statistisch-inhaltlich bezeichnet werden. Ein Manual (Rindermann & Amelang, 1994) beschreibt Konstruktion und Gütekriterien und schlägt Anwendungsmöglichkeiten vor.

## II. Hinweise zur Durchführung und Auswertung

Lehrevaluationen sollten *innerhalb* der Veranstaltungen vorgenommen werden, weil bei veranstaltungsexternem Vorgehen die Teilnahmequote zu gering ist. Die Inventare sollten 15 bis 10 Minuten vor regulärem Veranstaltungsende ausgeteilt werden, damit genug Zeit für Instruktion, Austeilung, Bearbeitung (ca. 8 bis 10 Minuten) und Einsammlung der Bögen zur Verfügung steht. Sollte der Dozent die Bögen selbst einsammeln, sollte Anonymität durch eine bereitgestellte Kiste gewährleistet werden, in die die Studenten die Bögen einwerfen.

Um Auswirkungen auf laufende Veranstaltungen erzielen zu können, sollten Lehrevaluationen Mitte des Semesters vorgenommen werden. Spätestens nach zwei Wochen müßte das Feedback (Auswertung) den Lehrenden vorliegen. In einer veranstaltungsinternen Besprechung können dann die Resultate besprochen werden, Ursachen kritischer Beurteilungen eruiert und Verbesserungsvorschläge gemeinsam erarbeitet werden. Es empfiehlt sich eine zusätzliche Begleitung durch hochschuldidaktische Beratung.

Die Auswertungsdauer je Fragebogen hängt von der Auswertungsmethode ab: Der HILVE wird in Varianten angeboten, die sowohl eine maschinelle als auch eine manuelle Auswertung erlauben. Bei Vorhandensein eines Lesegerätes können in kurzer Zeit große Zahlen an Bögen gelesen und verarbeitet werden. Freie (handschriftliche) Antworten müssen durch Hilfskräfte oder den Dozenten selbst ausgewertet werden (Gruppierung, Auszählung, Feedbackformulierung). Bei universitätsweiten Erhebungen empfiehlt sich die Einrichtung einer zentralen Evaluations- und Hochschuldidaktikstelle, die professionell Auswertung und Beratung der Dozenten vornimmt.

## III. Ergebnisdarstellung

In der Handanweisung des HILVE ist eine Feedbackform beschrieben und abgebildet. Dozenten sollten unstandardisierte Rohwerte (Mittel und Streuungen, je nach Interesse auch Verteilungen) als auch normierte Item- und Dimensionswerte

rückgemeldet werden. Normen (s. Tabellen in der Handanweisung) bieten ein Bezugssystem zur relativen Interpretation der eigenen Resultate an. Die Normen wurden für Vorlesungen und Seminare sowie für Studienfächer separat berechnet.

Die Ergebnisse können graphisch (Histogramme, Liniendiagramm) aufbereitet werden. Hierfür empfehlen sich gängige Programme (Harvard Graphics, SAS, SPSS etc.).

In einer Datenbank werden Resultate gesammelt und aufbereitet, um für wissenschaftliche Reanalysen sowie für fachspezifische Normierungen ausreichende Stichprobengrößen erzielen zu können. Die Autoren bitten, neue HILVE-Daten an die unten genannte Adresse (Rindermann) zu schicken.

#### IV. Gütekriterien

Für den HILVE wurden die klassischen Gütekriterien ermittelt (ausführliche Darstellung s. Rindermann, 1996). Berichtet werden an dieser Stelle Dimensionswerte für Veranstaltungsmittel, jeweils für alle HILVE-Skalen zusammen und für die Dozenten- bzw. Lehreffektivitätsskalen (ohne studentische Skalen und Thema) getrennt in Klammern.

Die mittleren *Trennschärfen* der Items betragen  $r=.78$  (.80) und können als sehr gut bezeichnet werden.

Die *Interraterreliabilität* liegt bei 20 Studierenden in der Höhe von  $r=.86$  (.90), bei 10 Studierenden beträgt sie  $r=.77$  (.80). Die *innere Konsistenz* (Cronbachs- $\alpha$ ) beträgt  $r=.89$  (.90). Die *Retestreliaibilität* (Mitte-Ende des Semesters) liegt bei  $r=.71$  (.72). In einer Vorstudie ohne Aufforderung zur Feedbackbesprechung innerhalb der Veranstaltung liegt die Retestreliaibilität bei  $r=.90$  (.92). Die Meßgenauigkeit der Dimensionen kann als gut bezeichnet werden, die Stabilität erreichte aber in der Hauptstudie aufgrund der Rückmeldeprozedur (Besprechung) relativ geringe Werte. Studentische Rohwerte sind weniger zuverlässig als Veranstaltungsmittel.

Die *Urteilerübereinstimmungen* einzelner studentischer Rohwerte sind gering ( $r=.27$  bzw.  $.31$  in Dozenten- und Lehreffektivitätsskalen), sie liegen aber in der Höhe, wie sie von studentischen Beurteilerübereinstimmungen aus der internationalen Forschung und aus der Forschung zum Peer-Review-Verfahren bei der Beurteilung eingereichter Manuskripte durch Fachgutachter mitgeteilt werden (s. Rindermann, 1996). Veranstaltungsmittelwerte ( $N \geq 10$  Studenten) zeichnen sich jedoch durch eine hohe Interraterreliabilität aus, d.h. würde eine andere studentische Stichprobe die gleiche Veranstaltung beurteilen, würden deren Resultate zu  $r=.77$ -.90 mit den Werten der ersten Stichprobe korrelieren. Veranstaltungsbeur-

teilungen müssen sich deshalb auf Mindeststichprobenumfänge von  $N \geq 10$  Studierenden stützen. In diesem Falle können Evaluationsresultate als relativ meßgenau und repräsentativ bezeichnet werden.

In *Faktorenanalysen* von Beurteilungen durch Studierende, Dozenten und Fremdurteiler konnten vier Faktoren zweiter Ordnung extrahiert werden: Dozenten-, Lehreffektivitäts-, studentische Selbstbeurteilungs- und Anforderungsskalen können bei den drei Urteilergruppen faktorenanalytisch getrennt werden.

Die *Validität* wurde mittels des Vergleichs der studentischen Urteile mit denjenigen von geschulten Fremdurteilern überprüft. Die mittleren Korrelationen betragen  $r = .59$  (.49) und liegen damit in für Validitätskoeffizienten günstiger Höhe. In Analysen zur Untersuchung des Einflusses von Biasvariablen (Variablen, die das studentische Urteil verzerren ohne mit der realen Lehre zusammenzuhängen) konnten individuelle studentische Charakteristika als Biasvariablen ausgeschlossen werden. Leistungsniveau (Abiturnote, veranstaltungsspezifische Klausurresultate), Geschlecht, Alter, Studien-motivation, Studienzufriedenheit usw. korrelieren wenig oder nicht mit der Beurteilung der Lehre.

Zusammenhänge konnten jedoch beim Besuchsgrund und dem Vorinteresse (beide retrospektiv eingeschätzt) in der Höhe von  $r = .10$  bis  $.30$  beobachtet werden. Studenten, die eine Veranstaltung aus Interesse besuchen, beurteilen diese etwas besser als Kommilitonen, die diese wegen eines Scheines oder aus Pflichtgründen besuchen. Zum Thema (interessant, relevant) können hohe Korrelationen beobachtet werden, allerdings kann auch gute Lehre für Themen interessieren – die Stärke des jeweiligen Bedingungs-Wirkungs-Verhältnisses ist nicht abschließend geklärt.

Vergleiche von Evaluationsergebnissen, die aus *verschiedenen Veranstaltungen eines Dozenten* gewonnen wurden mit Resultaten aus *Veranstaltungen gleichen Themas, die verschiedene Dozenten hielten*, ermöglichen einen Rückschluß auf die Generalisierbarkeit von Evaluationen. Hierbei ließ sich nachweisen, daß Dozentenvariablen weitgehend dozentenbezogen, studentische oder Themenvariablen jedoch mehr veranstaltungsspezifisch beurteilt werden.

## V. Angestrebte Weiterentwicklung

Weiterentwicklungen und Forschungsarbeiten sollen um drei Schwerpunkte kreisen:

1. Erprobungen in verschiedenen Fächern sollen die Berechnungen von weiteren fachspezifischen Normen erlauben.

2. Weitere frei wählbare Items sollen „behavioral“ (verhaltensnah) formuliert und deren Auswirkung auf den didaktischen Prozeß und die Veranstaltungsbeurteilung untersucht werden.
3. Die Gestaltung eines effektiven Rückmelde- und Beratungsprozesses soll untersucht werden.

## VI. Besondere Hinweise zur Verwendung des Verfahrens

Generell ist zu empfehlen, Evaluationen in mehreren verschiedenen Veranstaltungen (Mindeststichprobengröße jeweils  $N \geq 10$  Studierende) vorzunehmen, falls es Ziel ist, die Lehrqualität eines Dozenten zu bestimmen. Der Einfluß eventuell auftretender Störvariablen kann so minimiert werden. Bei systematischen Abweichungen durch verschiedene studentische Stichprobensamensetzungen können Korrekturen vorgenommen werden.

Der HILVE-Fragebogen ist auch in einer maschinell lesbaren Form auslieferbar. Bei maschineller Auswertung können in kurzer Zeit fehlerarm und kostengünstig große Mengen von HILVE-Bögen ausgewertet werden. Dies erlaubt durch eine kurzfristige Rückmeldung des Feedbacks und ergänzende Beratung, Lehrevaluationsergebnisse für die Optimierung der laufenden Veranstaltung einzusetzen.

Rindermann, H. & Amelang, M. (1994). Das Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE). Heidelberg: Asanger. ISBN: 3-89334-277-X.

Rindermann, H. & Amelang, M. (1994). Entwicklung und Erprobung eines Fragebogens zur studentischen Veranstaltungsevaluation. *Empirische Pädagogik*, 8(2), 131-151.

Das Manual wurde veröffentlicht im Asanger-Verlag:

Roland Asanger Verlag GmbH, Rohrbacherstr. 18, 69115 Heidelberg; Tl. 06221/183104, Fax 06221/160415.

## **8.6 Bielefelder Fragebogen zur Bewertung von Lehrveranstaltungen von W. D. Webler, IZHD/Arbeitsgruppe Studentische Veranstaltungskritik**

### **I. Kurze Beschreibung des Instruments**

Maschinenlesbarer Fragebogen mit 6 Varianten. Die Items wurden nach Sichtung internationaler und deutscher Fragebogen nach einem eigenen Modell guter Lehre generiert. Die Varianten erfassen alle Arten von Lehrveranstaltungen.

Die Langfassung A (geistes- und sozialwissenschaftliche Seminare) hat 74 Items und 2 offene Fragen. Die Versionen für Vorlesungen sowie die für (naturwissenschaftliche) Übungen und Praktika sind kürzer.

Eine ab 1994 eingesetzte, um ein Drittel gekürzte Fassung hat (trotz Maschinenlesbarkeit) 6 ad hoc formulierbare Fragemöglichkeiten und 2 offene Fragen. Das Instrument wurde nach einem Pilotversuch (1991 in der Fakultät für Soziologie) ab 1992 in 11 von 13 Fakultäten flächendeckend in allen Lehrveranstaltungen (mehr als 700) eingesetzt. Ab 1994 ging der Einsatz zurück (Abnutzungseffekte wegen zu häufigen Einsatzes) und wurde z.T. durch andere schriftliche/ mündliche Methoden der Rückmeldung ersetzt.

Eine Variante des Fragebogen-Instruments wurde vom IZHD für Veranstaltungsbewertungen im Rahmen des Lehrbericht-Projekts an Fachbereichen der hessischen Hochschulen ebenfalls in einer großen Zahl unterschiedlicher Lehrveranstaltungen eingesetzt.

### **II. Hinweise zur Durchführung und Auswertung**

Die Dauer zum Ausfüllen des Fragebogens beträgt je nach Version 10-20 Minuten. Ausgefüllt wird in der Mitte des Semesters in allen Veranstaltungen einer Semesterwoche, während einer laufenden Veranstaltung. Die Auswertung erfolgt durch Belegleser und ein Auswertungsprogramm. Der Belegleser liest 3.000 doppelseitige Fragebogen pro Stunde.

### **III. Ergebnisdarstellung**

Die Ergebnisse werden pro Veranstaltung in absoluten Zahlen, Balkendiagrammen und Mittelwerten ausgedruckt (z.B. auf Overheadfolien zur Diskussion in der Lehrveranstaltung; „Bielefelder Diskursmodell“). Eine Datenbank wird nicht angelegt (vereinbarte Randbedingung).

#### IV. Gütekriterien

Keine Analysen nach Gütekriterien durchgeführt.

#### V. Angestrebte Weiterentwicklung

Zur Zeit nicht beabsichtigt.

#### VI. Besondere Hinweise zur Verwendung des Verfahrens

Hinweise zur Verwendung des Verfahrens sind in folgenden Berichten enthalten:

W.-D. Webler u.a.: Lehrberichte. Empirische Grundlagen, Indikatorenauswahl und Empfehlungen. (Bundesministerium für Bildung und Wissenschaft (Hg.): Schriftenreihe Studien zu Bildung und Wissenschaft, Bd. 107). Bad Honnef 1993. (Anhang mit verschiedenen Fragebogen-Varianten).

W.-D. Webler: Evaluation der Lehre: Praxiserfahrungen und Methodenhinweise. In: Hellmut Winkler (Hg.): Qualität der Hochschulausbildung. (Wiss. Zentrum für Berufs- und Hochschulforschung der Universität GH Kassel; Werkstattberichte 40), Kassel 1993.

### **8.7 Fragebogen zur „Evaluation der Lehre“ an der Universität Mannheim**

Das Evaluationsprojekt „Evaluation der Lehre“ wurde aufgrund eines Antrags des Mannheimer AStAs Ende 1992 begonnen und vom Landesministerium Baden-Württemberg gefördert. Die Senatskommission für Lehre begleitete das Projekt, dessen Leitung in den Händen von Dr. Hans-Dieter Daniel lag.

Das große und umfassende Projekt an der Universität Mannheim hatte zum Ziel, eine Vollerhebung im betriebswirtschaftlichen Grundstudium (BWL I und BWL II) durchzuführen. Später erfolgten weitere Erhebungen zur Lehrbewertung in den Geisteswissenschaften. Alle beteiligten Lehrenden nahmen auf freiwilliger Basis an der Lehrbewertung teil.

Entsprechend der Art der Veranstaltung (Vorlesungen, Übungen, Tutorien) wurden von Daniel veranstaltungsspezifische Fragebogen entwickelt. Außerdem wurde ein allgemeiner Fragebogen ausgearbeitet, der veranstaltungsübergreifende Informationen und einige Angaben zur Person abfragte.

Die Bearbeitungsdauer betrug aufgrund der relativ umfassenden Fragebogen 25-30 Minuten, maximal nach 45 Minuten konnte die Erhebung abgeschlossen werden.

Folgende Fragebogen wurden entwickelt:

Fragebogen zu den Lehrveranstaltungen in „Grundzüge der Betriebswirtschaftslehre“ 1992/93 (Allgemeiner Teil)

Fragebogen zu den Vorlesungen in „Grundzüge der Betriebswirtschaftslehre“ 1992/93

Fragebogen zu den Übungen in „Grundzüge der Betriebswirtschaftslehre“ 1992/93

Fragebogen zu den Tutorien in „Grundzüge der Betriebswirtschaftslehre“ 1992/93

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 (Allgemeine Fragen)

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 -Fragebogen für Einführungsveranstaltungen -

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 -Fragebogen für Vorlesungen -

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 -Fragebogen für Seminare -

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 -Fragebogen für sprachpraktische Übungen -

Hörerbefragung in der Fakultät für Sprach- und Literaturwissenschaft 1993/94 -Fragebogen für Tutorien -

Evaluation der Lehre, Fragebogen für Einführungsveranstaltungen (Anglistisches Seminar, Kurzfragebogen)

Fragebogen für Studierende des Diplomstudienganges „Philologie mit wirtschaftswissenschaftlicher Qualifikation“ mit dem Kernfach Slavistik

Eine Überprüfung der Gütekriterien dieser Instrumente wurde nicht durchgeführt.

Das Projekt wurde von einer Senatskommission für Lehre betreut, die die Endfassung der jeweiligen Fragebogen verabschiedete. Folge dieser Konstellation

sind ausgesprochen umfangreiche Fragebogen, die vor allem zu wissenschaftlichen Zwecken oder zur Itemrequirierung geeignet erscheinen. Wird eine Vollerhebung geplant, so ist mit der Notwendigkeit von viel Personal zu rechnen.

Hinweise zur Verwendung des Verfahrens sind in folgenden Berichten enthalten:

Daniel, H.D., Thoma Michaela & Bandilla, W. (1995). Das Modellprojekt „Evaluation der Lehre“ an der Universität Mannheim. Teil 1: Planung und Durchführung von Befragungen in Lehrveranstaltungen. In: Mohler, P. (Hrsg.) *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Waxmann Verlag: Münster.

Daniel, H.D. (1995). Das Modellprojekt „Evaluation der Lehre“ an der Universität Mannheim. Teil 2: Statistische Auswertung von Befragungen in Lehrveranstaltungen. In: Mohler, P. (Hrsg.) *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Waxmann Verlag: Münster.

Daniel, H.D. (1995). Bewertung der Lehre aus Sicht der Studierenden und Absolventen. In D.Müller-Böling (Hrsg.), *Qualitätssicherung in Hochschulen* (S.160-185). Gütersloh: Verlag Bertelsmann Stiftung.

## **9 Möglichkeiten der Datenverwendung: Reichweite und Wirksamkeit**

### **9.1 Die Wirkungslosigkeit reiner Datenerhebung**

Nachdem in Deutschland seit einigen Jahren Projekte zur studentischen Veranstaltungskritik laufen, erscheint es möglich, bereits einige Ergebnisse und Probleme festzustellen. So beschreiben z.B. Gralki und Heidemarie Hecht: "Die in Deutschland laufenden lehrveranstaltungsbezogenen Evaluationsvorhaben brechen i.d.R. nach der Ergebnismeldung ab. Es gibt kaum entwickelte Vorstellungen oder Modelle, was mit den Ergebnissen weiter geschehen soll, wie sie zielgerichtet zu einer Verbesserung der Vermittlungsqualifikation genutzt werden können. Die Auffassung, daß allein die Kenntnisnahme der Ergebnisse zu einer Verbesserung des eigenen Lehrverhaltens beiträgt, ist weit verbreitet" (Gralki und Heidemarie Hecht, 1992, S. 110). Daß diese Ansicht jedoch trügerisch ist, belegen eine ganze Reihe von Stellungnahmen.

Rüdiger Reißert hat an der Technischen Universität Berlin in über 1400 Lehrveranstaltungen Studenten interviewt. Seine Erfahrung: "Kein einziger der Vorschläge, die als Konsequenzen aus den Bewertungen der Lehrveranstaltungen sowie aus zahlreichen Gesprächen mit Lehrenden und Studierenden empfohlen wurden, sind von der Hochschulleitung umgesetzt worden" (Zitiert nach: DER SPIEGEL, 1994, S. 57). "Die Mehrheit der Professoren hat kein Interesse am Urteil der Studenten", sagt Privatdozent Hans-Dieter Daniel, Projektleiter in Mannheim. "Veränderungen sind nicht erwünscht. Didaktische Weiterbildung oder Einsatz neuer Medien - Fehlanzeige" (Zitiert nach: DER SPIEGEL, 1994, S. 57).

Die gleichen Auswirkungen beschreibt Schmyk (1992) für das Evaluationsprojekt der TU Berlin: "Das großangelegte Evaluationsprojekt an der TU droht nach zwei Jahren fleißigen Auftürmens von (ebenfalls anonymisierten) Datenbergen im Sande zu verlaufen, weil sich unter den Verantwortlichen niemand mehr findet, der über notwendige Schlußfolgerungen aus den Befragungsergebnissen sprechen möchte" (S.88).

Auch Megerle stellte bereits 1991 fest, daß reine Erhebungen keine weiteren Konsequenzen nach sich ziehen, und er begründet auch warum: "... Dozentschelte allein führt gewiß nicht weiter. Sie kann im Einzelfall anspornend wirken oder den Druck verstärken (allerdings u.U. auch trotzige Verweigerung provozieren). Eine generelle Kursänderung kann sie aufgrund der von der wissenschaftlichen und politischen Öffentlichkeit festgelegten ungleichen Wertigkeit von Forschung und Lehre nicht bewirken. Dazu ist sie nicht der richtige Ansatz" (Megerle, 1991, S. 35).

Umfassender hat Centra (1980) untersucht, unter welchen Bedingungen ein Feedback der Ergebnisse allein ausreicht, um Lehrende zur Veränderung der Unterrichtsgestaltung zu motivieren. Dabei kommt er zu dem Schluß, daß dies nur der Fall ist, wenn Lehrende a) in Lehraspekten, die sie als wichtig einstufen und b) in denen sie eine weitaus bessere Selbsteinschätzung hatten, schlechtere Werte erzielten als von sich selbst erwartet.

Lehrende waren allein durch die Ergebnisse der Studierendenbefragungen zu Änderungen nur dann bereit, wenn sie eine unrealistische Selbsteinschätzung über die ihnen relevanten Lehraspekte hatten. Da nicht davon ausgegangen werden kann, daß sehr viele Lehrende eine in größerem Ausmaß unkorrekte Einschätzung ihrer Lehrleistung haben, müssen Strategien gefunden werden, die den Lehrenden Hilfestellungen geben, eine ergebnisorientierte Verwendung der Bewertungsdaten ermöglichen und zu Veränderungen motivieren. Um Verbesserungen zu erreichen, muß Lehrevaluation Folgen haben.

### **Wichtige Elemente der Evaluation: Empfehlung und Rechenschaft**

Nach einer Phase der überwiegenden Datensammlung ist ein Übergang zu beobachten, die Lehrevaluation in ein breiteres Verfahren einzubetten (vgl. Reisert, 1994). Die aktuellen Lehrbewertungen, ob von HIS, dem Wissenschaftsrat, dem IZHD oder dem norddeutschen Evaluationsverbund, orientieren sich inzwischen alle am „niederländischen Modell“ (zu diesem Modell vgl. Richter 1995). Dabei werden nicht nur die Ergebnisse der Lehrbewertung ermittelt, sondern auch Empfehlungen für die Lehrverbesserungen ausgesprochen.

Ein weiteres wichtiges Element darf bei diesem Modell nicht übergangen werden: die Rechenschaftslegung. Nach einem gewissen Zeitraum gibt es eine Art „Kontrollpunkt“, zu dem eine weitere Bewertung erfolgt, mit dem Ziel festzustellen, welche der Empfehlungen umgesetzt wurden. In Deutschland sind die ersten größeren Projekte mittlerweile soweit, daß dieser „Kontrollpunkt“ erreicht ist. Noch ist es allerdings zu früh, um zu beurteilen, ob sich aufgrund der in den meisten Ländern eingeführten Lehrberichtspflicht eine maßgebliche Änderung in der Hochschullehre ergeben wird.

### **Notwendigkeit eines Gesamtkonzeptes zur Verbesserung der Lehre**

Eine reine Erhebung studentischer Veranstaltungskritik dürfte insgesamt wenig Folgen haben, wenn sie nicht eingebettet wird in ein Gesamtkonzept zur Verbesserung der Lehre, das Lehrende für ihre Aufgaben in diesem Bereich sowohl qualifiziert als auch motiviert. Allein auf den guten Willen der Hochschullehrenden zu setzen, aktiv an der Lehrverbesserung mitzuwirken und Belohnungen oder

Preise für gute Lehre zu verteilen, reicht für eine systematische Lehrverbesserung nicht aus.

Man kann davon ausgehen, daß bei den meisten Lehrenden der gute Wille zur guten Lehre und zu ihrer Verbesserung nicht fehlt. Der Blick auf den Hochschulalltag der Lehrenden zeigt aber, daß er bei den meisten Wissenschaftlern und Wissenschaftlerinnen reichlich ausgefüllt ist. Engagieren sie sich stärker in dem einen oder anderen Bereich (Forschung, Lehre, Selbstverwaltung), so müssen sie sich in einem anderen einschränken. Gerade in Bereichen wie der wissenschaftlichen Tätigkeit ist eine mehr oder weniger eindeutige Prioritätensetzung unerlässlich.

Weiter kann unterstellt werden, daß der allergrößte Teil des wissenschaftlichen Personals in irgendeiner Form ambitioniert ist und im allgemeinen recht rational handelt. Akzeptiert man diese Annahmen und betrachtet sich dabei die Mechanismen der 'scientific community', die Mechanismen der Anerkennung, des beruflichen Fortkommens, der größeren wissenschaftlichen und finanziellen Freiräume, so stellt sich ernsthaft die Frage, wieso sich ein kleiner Teil von Wissenschaftlern und Wissenschaftlerinnen überhaupt überdurchschnittlich in der Lehre einsetzt.

Natürlich gibt es dafür Gründe. Und natürlich sind nicht alle Lehrende ausschließlich am beruflichen Fortkommen interessiert. Auch Lust an der Lehre, Beliebtheit bei den Studierenden und vielleicht die Verpflichtung einer jüngeren Generation gegenüber können motivierend wirken. Dennoch ist klar festzustellen, daß es ein System von Anforderung und Belohnung, von zu erfüllenden Kriterien und Aufstiegsmöglichkeiten, von Bedingung und Erfüllung, wie es sich in der Forschung bewährt hat, in der Lehre nicht gibt.

Dies ist keine neue Erkenntnis. Es gibt bereits mehr oder weniger systematische Ansätze, dies zu beheben. Preise für gute Lehre zu verteilen ist "in". Zwar ist erfreulich, daß ein ausdauerndes Bemühen in der Lehre jetzt manchmal honoriert wird. Aber es ist nicht davon auszugehen, daß sich dadurch prinzipiell etwas ändert. Vielleicht sind diejenigen, die sich den Lehranforderungen schon immer stärker gewidmet haben, jetzt noch etwas motivierter, aber ein Lehrpreis wird einen schlechten Didaktiker, eine planlose Lehrende kaum zu Fortbildungsmaßnahmen motivieren. Preise treffen die positive Spitze, die glanzvollen Vorzeigelehrenden. Dies ist erfreulich und gerecht. Aber ein systematischer Anreiz für alle Lehrende sind sie nicht.

Lehrleistungen müssen eine festgelegte Größe in den Regelkreisen der Wissenschaft werden. Bei einem solchen Konzept hat die Lehrbewertung die Funktion einer Befundaufnahme, an die sich Vorschläge der Verbesserung anschließen müssen, und deren Ausführung gelegentlich überprüft wird. Sowohl besonderes

Engagement als auch die passive Haltung in der Lehrverbesserung müssen daraufhin Konsequenzen haben. In welchem Ausmaß und auf welche Art Lehrleistungen in einer wissenschaftlichen Laufbahn berücksichtigt werden sollten, wird bereits diskutiert. Eine systematische Berücksichtigung, welchen Ausmaßes auch immer, wäre für eine bessere Nutzung von Evaluationsdaten wünschenswert.

## 9.2 Verwendungsmöglichkeiten von Beurteilungsdaten

Lehrevaluationen können ganz unterschiedlichen Zwecken dienen. Diese unterschiedlichen Verwendungsmöglichkeiten sollten einen maßgeblichen Einfluß sowohl auf die Erhebungsinstrumente als auch auf die Konzeption entsprechender Verwendungssysteme haben. Es erscheint daher sinnvoll, prinzipiell zu unterscheiden, welche Möglichkeiten der Datenverwendung überhaupt bestehen.

Insgesamt können fünf Verwendungsmöglichkeiten unterschieden werden (vgl. Gralki und Heidemarie Hecht, 1992):

- Datenerhebungen für Forschungszwecke,
- für eine bessere Transparenz über die Lehre,
- als Grundlage für den Dialog über Lehrqualität,
- oder für Qualifikationsmaßnahmen ,
- und schließlich Evaluationsdaten als Basis entsprechender Steuerungssysteme.

Dieses Einteilungsschema wird im folgenden genutzt, sowohl um die Diskussion in diesem Feld zu strukturieren als auch um die sich daraus ergebenden Konsequenzen systematisch darstellen zu können. Die Reihenfolge dieser Ansätze wird an dieser Stelle unter dem Gesichtspunkt der Reichweite ihres Veränderungsanspruches bestimmt.

Den geringsten Veränderungsanspruch besitzt die Datenerhebung zu *Forschungszwecken*, da nur der bestehende Ist-Zustand ermittelt wird und Grundlagen der Evaluationsmöglichkeiten geprüft werden sollen. Es wird vor allem zur Hypothesengewinnung im Bereich der Lehre genutzt.

Der *Transparenzansatz* stellt die Informationsauskunft, sowohl als Feedback für die Lehrenden als auch als Orientierungs- und Entscheidungsmöglichkeit für eine Kursbelegung der Studierenden, in den Mittelpunkt.

Steht die *Kommunikation* im Mittelpunkt, so sollen die erhobenen Daten als Grundlage für einen Dialog über die Lehre genutzt werden. Lehre soll im Fachbe-

reich bzw. der Hochschule stärker thematisiert, der Erfahrungsaustausch dadurch unterstützt werden.

Die Verwendung der Daten als *Qualifikationsgrundlagen* soll den Lehrenden ermöglichen, ihre Stärken und Schwächen besser zu erkennen. Die Weiterbildung des wissenschaftlichen Personals wird betont. Die vorhandenen Daten sollen zur Entscheidungshilfe für Fortbildungsmaßnahmen herangezogen werden.

Ein umfassender Veränderungsanspruch ist erst beim *Steuerungsansatz* vorhanden. Lehrbewertung dient dann z.B. als Basis für einen bestimmten Teil der Mittelvergabe, um so eine höhere Motivation für Verbesserungen in der Lehre zu erreichen.

Nachfolgend werden diese fünf Ansätze mit ihren unterschiedlichen Schwerpunkten und Reichweiten diskutiert, um sie in ihrer praktischen Wirksamkeit für die Verbesserung der Lehre als Teil der Studienreform einzuschätzen.

### **Forschung: Ermittlung und Diskussion**

Der Forschungsansatz ist für eine wissenschaftliche Aufarbeitung der Lehrqualität von Bedeutung. Es kann u.a. dazu dienen, Instrumente und Methoden der Lehrbewertung zu entwickeln und zu verbessern, Hypothesen zur Lehre zu testen und Grundlagen der empirischen Sozialforschung auf Bereiche der Lehrevaluation zu übertragen.

Wie die Veranstaltungskritik selbst, kann jedoch auch die Forschung in diesem Bereich nur Wissen bereit stellen. Auswirkungen per se sind von bloßen Befunden nicht zu erwarten. Dennoch darf man die Wichtigkeit gerade dieses Bereiches nicht unterschätzen. Auf genau diesem Terrain setzen sich die Befürworter/-innen und Gegner/-innen der Lehrevaluation mit Vorliebe auseinander (vgl. Kapitel 4 und 5).

Gegner und Gegnerinnen der Lehrevaluation vermuten bei den Studierenden, daß sie die Lehre gar nicht beurteilen könnten, da sie zu unreif und zu unerfahren seien, studentische Veranstaltungskritik sei daher nichts weiter als ein 'Beliebtheitswettbewerb'. Bewertungsbogen seien prinzipiell unreliabel und nicht valide. Und eine Vielzahl externer Variablen (Kursgröße, Notengebung) beeinflusse die Bewertung. In diesem Feld der Auseinandersetzung kann die empirische Forschung viel zur Aufklärung und damit Versachlichung der Debatten beitragen.

Die Fachdiskussion in diesem Bereich bezieht sich in Deutschland dabei vor allem auf zwei Themen, die noch einiger empirischer Untersuchungen und Aufklärungen bedürfen (vgl. zum Beispiel die Beiträge in Mohler, 1995):

1. Können Bewertungen unterschiedlicher Kurse (d.h. mit einer unterschiedlich zusammengesetzten Studierendenschaft) überhaupt zu Vergleichen herangezogen werden?
2. Welche Einflüsse spielen bei der Lehrbewertung von Seiten der Studierenden eine Rolle?

Freilich wird schon an diesem ersten Ansatz deutlich, daß die Aufteilung und Abgrenzung nur theoretisch möglich ist, sich in der Realität aber anders darstellt. So hat allein die Veröffentlichung von Untersuchungsergebnissen in einigen großen Magazinen (vor allem DER SPIEGEL, 1989, aber auch Kriz im STERN, 1993) für öffentliche Aufmerksamkeit, Vergleiche und so viel sozialen Druck gesorgt, so daß in der Folge an den Hochschulen im Bereich der Lehrevaluation einiges in Angriff genommen wurde.

Streng genommen ist bei diesem Ansatz davon auszugehen, daß Veröffentlichungen höchstens in entsprechenden Fachpublikationen vorliegen dürfen, um dessen theoretischen Rahmen nicht zu überschreiten. Um Wirksamkeit zu entfalten, müssen forschungsbezogene Untersuchungen zur Lehrevaluation mit zuverlässigen Instrumenten arbeiten und ihre Ergebnisse in verständlicher, nachvollziehbarer Form veröffentlicht werden.

### **Transparenz: Veröffentlichung und Orientierung**

Der Transparenzansatz geht eine Stufe weiter als das Forschungsmodell. Die Daten sollen erhoben und in einem gewissen Rahmen öffentlich gemacht werden. Mit der Veröffentlichung hätten Lehrende die Möglichkeit, ihre Leistungen einzuschätzen, Studierende die Möglichkeit, bei der Wahl ihrer Kurse diese Informationen zu berücksichtigen.

Implizit wird vermutet, daß Lehrende sich aufgrund dieser Öffentlichkeit und der damit entstehenden Konkurrenz stärker um Lehrverbesserungen bemühen würden. Dieser Druck würde noch wachsen, wenn Studierende anhand der Daten eine entsprechende Kursbelegung durchführen könnten. Sind die Lehrbeurteilungen somit öffentlich und werden die Studierenden explizit aufgefordert, sie als Orientierungshilfe zu benutzen, so findet bereits eine indirekte Steuerung mit Hilfe sozialen Drucks statt.

Betrachtet man diesen Ansatz, so wird ersichtlich, daß zwei Vorbedingung notwendig sind, um die impliziten Auswirkungen herbeizuführen. Erstens müßten die Studierenden überhaupt dieses Ausmaß an Wahlmöglichkeiten besitzen, um einen Marktmechanismus im Sinne von Angebot und Nachfrage etablieren zu können. Dies erscheint bei der jetzigen Situation an den deutschen Hochschulen für die meisten Fächer illusorisch.

Zweitens müßten die gewonnenen Ergebnisse veröffentlicht werden. Eine öffentliche Präsentation dieser Daten findet in den Projekten der Lehrevaluation häufig nicht so detailliert statt. Dies hängt auch mit den Unklarheiten zusammen, ob Ergebnisse der studentischen Veranstaltungskritik überhaupt veröffentlicht werden dürfen (vgl. auch Kapitel 1.2).

Dabei werden vor allem Bestimmungen zum Datenschutz thematisiert. Es wird u.a. diskutiert, ob Regelungen der "Dienstaufsicht" von Hochschulleitungen oder Wissenschaftsministerien so geändert werden könnten, daß Lehrbewertung organisiert durchgeführt und die Ergebnisse in bestimmte Entscheidungen miteinbezogen werden könnten. Gestritten wird, ob Lehre eine universitätsöffentliche Angelegenheit sei, und ob Veröffentlichungen von Evaluationsdaten überhaupt mit dem "informationellem Selbstbestimmungsrecht" vereinbar seien.

Der Transparenzansatz ist somit in vielerlei Hinsicht weniger ein eigenständiges Modell, vielmehr beschreibt es einen Verwendungszweck studentischer Lehrevaluation: ihre Veröffentlichung. Im Modellfall würde der Veranstaltungsbesuch über die veröffentlichten Bewertungen gesteuert. Auswirkungen auf die Lehrqualität sind aus den genannten Gründen kaum zu erwarten, eher sind kontraproduktive Reaktionen zu befürchten, wenn keine begleitenden Maßnahmen vorhanden sind. Die bloße Veröffentlichung kann zwar Aufregungen und Diskussionen auslösen, sicherlich aber nicht eine Verbesserung der Lehre in kontinuierlicher und geregelter Form.

### **Kommunikation: Dialog und Koordination**

Kommunikation ist die grundsätzlichsste Vorbedingung bei der Einführung studentischer Veranstaltungskritik. Sie ist zum einen notwendig, um Studierende und Lehrende überhaupt zu motivieren, an Erhebungen zur Lehrbewertung teilzunehmen. Zum anderen ist sie unerläßlich, um nach der Datenerhebung weitere Maßnahmen vorzubereiten und durchzuführen. Das Kommunikationsmodell wird zwar als "niedrigste" Ebene eines Rahmenkonzeptes betrachtet, ist damit aber auch grundlegende Basis.

Kommunikation zwischen Lehrenden und Studierenden scheint in vielen Bereichen nicht optimal. Peisert u.a. (1988) ermittelten, daß es nur ca. 25% der Studierenden gelingt, zumindest manchmal Kontakt zu den Lehrenden herzustellen. Nur 10% der Studierenden sind mit dieser Situation zufrieden: 60% der Hochschullehrer und -lehrerinnen beklagen sich dagegen, daß die von ihnen angebotenen Sprechstunden zu wenig genutzt werden. Preißer (1991) spricht von einer "fehlenden Gesprächskultur".

In den entsprechenden Hochschulgremien und unter den Lehrenden selbst findet wenig Diskussion zum Thema Qualität der Lehre statt. Studentische Veranstal-

tungskritik kann ein Instrument sein, die Kommunikation in einem Kurs anzuregen. Wird nur ein Dialog über die Lehrqualität angestrebt, so spielt auch die Qualität des benutzten Fragebogens eine nicht so große Rolle: Auch aus einer "Krücke von Fragebogen" (Comelli, 1991) kann man die Grundlage für Kommunikationsprozesse schaffen.

Preißer (1991) betrachtet studentische Lehrbewertung auch deswegen als notwendig, weil Veranstaltungsfeedback häufig nicht gegeben wird. Um ein direktes Feedback zu ermöglichen, ist es unerlässlich, die Ergebnisse der erhobenen Daten zumindest im jeweiligen Kurs zu besprechen. Werden jedoch nur Daten erhoben, erhält der Dozent/ die Dozentin ein mehr oder weniger unkommentiertes Ergebnis.

Einen anderen Weg gehen die Projekte, die sich an dem sogenannten 'Niederländischen Modell' orientieren. Dort werden nach der Erhebung sehr unterschiedlicher Daten Empfehlungen formuliert und diese in einem mehr oder weniger großen Rahmen des jeweiligen Faches besprochen. Allerdings steht dabei die Lehrqualität des einzelnen Lehrenden weniger im Mittelpunkt, was die Kommunikation erheblich erleichtern dürfte.

Der allererste Schritt in jedem Projekt sollte das Gespräch über die Lehre zwischen den Kursteilnehmern und -teilnehmerinnen auf der einen und den Lehrenden auf der anderen Seite sein. Dies darf sich jedoch nicht auf die Ebene der Beziehungen zwischen Lehrenden und Kurs beschränken, sondern müsste sich auf der Ebene der Fachbereiche und der Hochschule fortsetzen, und zwar möglichst in kontinuierlicher Weise.

Beschränkt man das Kommunikationssystem nicht nur auf das interne System Hochschule, so sollte auch das Umfeld (Schulen, künftige Arbeitgeber und Arbeitgeberinnen etc.) in die Interaktion einbezogen werden. "Im Vergleich zum Dialog in den USA mutet die hermetisch abgeschlossene Dreigliederung von Schule, Hochschule und Arbeitsmarkt noch immer wie ein indisches Kastenwesen an" (Schnitzer, 1991, S. 17). Der von Schnitzer beschriebene Koordinationsmangel trägt mit zu den bekannten Schwierigkeiten im Studium bei (z.B. häufiges Wechseln bestimmter Fachkombinationen). Die Institutionalisierung eines Dialogs über die verschiedenen Ebenen erscheint notwendig.

Vorteilhaft an diesem Kommunikationsansatz ist, daß im Vergleich zu den weitergehenden Ansätzen mit geringeren Widerständen zu rechnen wäre, da betroffene Lehrende keinerlei Sanktionen zu fürchten hätten und die "Öffentlichkeit" als erstes auf den entsprechenden Kurs beschränkt bleiben könnte. Letzteres würde vor allem für ein mündliches Veranstaltungsfeedback gelten, in geringerem Maße jedoch auch für eine Datenauswertung ohne Kursvergleiche und öffentlichen Aushang zutreffen.

Problematisch dagegen ist die völlige Abhängigkeit der Lehrverbesserung vom "goodwill" der Lehrenden. Es ist fraglich, ob die Motivation oder die Fähigkeit etwas am eigenen Kurs zu ändern, sich allein durch die Kommunikation über die Ergebnisse der Lehrbewertung verbessern würde. Dabei geht es nicht nur um eine geringe Motivation der Lehrenden. Comelli (1991) z.B. geht davon aus, daß für das Gelingen solcher Feedback-Mechanismen, wie sie die studentische Veranstaltungskritik darstellen soll, soziale und vor allem kommunikative Fähigkeiten notwendig sind, die in diesem Bereich z.T. völlig fehlen.

Kommunikation muß zwar unabdingbarer Bestandteil eines Rahmenkonzeptes sein, in dem studentische Veranstaltungskritik die wichtigste Datengrundlage darstellt, sie bleibt jedoch ohne weitere Maßnahmen unzureichend. Die Ergebnisse bisheriger Projekte der Lehrevaluation lassen die Schlußfolgerung zu, daß sich an der Qualität der Lehre nur punktuell etwas ändern würde.

### **Qualifikation: Kompetenzsteigerung und Fortbildung**

Im Qualifikationsmodell als Ausgangspunkt zur Verbesserung der Lehre wird der/ die Lehrende als zentrales Organ der Hochschule in den Mittelpunkt der Überlegungen gestellt. Eine höhere Qualifizierung der Lehrenden führt in diesem Sinne zu einer Verbesserung der Lehre. Im Vergleich zum Kommunikationsmodell beinhaltet das Qualifikationsmodell einen weitreichenderen Schritt. Als Voraussetzungen für eine Verbesserung werden Qualifikationsmaßnahmen gesehen.

Beurteilt man den Ist-Zustand der bisherigen Qualifikation der Dozenten und Dozentinnen in Bezug auf die Lehre, so kann man zu folgendem Ergebnis kommen: "Die Hochschule verlangt von den Hochschullehrern Dinge, für die sie nicht ausgebildet sind und für die sie nicht systematisch die Ressourcen bereitstellt. Sie versetzt sie nicht in die Lage, die gestiegenen Anforderungen an ihre Tätigkeit qualifiziert zu verrichten, entlastet sie aber auch nicht. An diesem kurzen Rückblick sieht man, daß die Lehre nicht verbessert werden kann, ohne die Institution der Hochschullehrer zu reformieren. Denn Lehren ist nur ein Bestandteil dieser unqualifizierten Tätigkeiten, die die Hochschullehrer zu verrichten haben. Die Lehrenden an der Hochschule haben Lehren systematisch nie gelernt. Die Hochschule stellt dafür keine Ressourcen bereit, um Lernen und Lehren zu lernen sowie Lernen zu lehren" (Preißer, 1992, S. 209).

Im Vergleich zu dem vorgestellten Kommunikationsansatz wird beim Qualifikationsmodell nicht davon ausgegangen, daß die alleinige Rückmeldung der Lehrbewertungsergebnisse bereits zu Verbesserungen führt. Die Qualifikation der Schlüsselakteure an den Hochschulen wird somit als zentrales Mittel der Lehrverbesserung gesehen. Bereits heute haben die Hochschulen den Auftrag, die Weiterbildung ihres Personals zu fördern (Hochschulrahmengesetz vom 23.4.87 § 2,4)

und dafür zu sorgen, daß die Lehre den didaktischen und methodischen Erkenntnissen entspricht (ebd. § 8,2).

Werden die tatsächlichen Weiterbildungsmöglichkeiten des wissenschaftlichen Personals in Bezug auf die Lehre betrachtet, so stellt sich heraus, daß Weiterbildungseinrichtungen nicht flächendeckend vorhanden sind. Obwohl Konzepte zur hochschuldidaktischen Aus- und Fortbildung existieren (z.B. Interdisziplinäres Zentrum für Hochschuldidaktik, 1995), sind aktuelle Anwendungen für eine systemintegrierte und damit verpflichtende didaktische Aus- und Weiterbildung des wissenschaftlichen Personals an Hochschulen noch Mangelware.

Zwar ist das Interesse an didaktischer Fortbildung in den letzten Jahren gestiegen, es ist jedoch immer noch eine kleine Minderheit der Lehrenden, die an solchen Kursen teilnimmt.

Obwohl im Hochschulrahmengesetz (§44 Einstellungsvoraussetzungen für Professoren) festgelegt wird, daß Professoren eine 'pädagogische Eignung' besitzen müssen, wird diese aber nicht näher spezifiziert. Diese Eignung sollte in der Regel durch Erfahrungen in der Lehre oder Ausbildung nachgewiesen werden. Auch in den Landesgesetzen ist meist vorgeschrieben, daß zum Abschluß einer Promotion ein wissenschaftlicher Vortrag gehört. In einigen Hochschulgesetzen ist bei einer Berufung auch eine Probelehrveranstaltung vorgesehen. Diese wird jedoch selten als tatsächliche Überprüfung einer Lehrbefähigung gestaltet.

Es gibt bislang wenig Anreize, sich um eine höhere Qualifikation in der Lehre zu bemühen. Berufungen erfordern fast ausschließlich Forschungsleistungen, ihre Grundlage sind Publikationslisten. Mit Ausnahme einer Probelehrveranstaltung findet keine Überprüfung der Lehrbefähigung statt. Nicht einmal in dem genannten Probevortrag wird die Lehrqualität systematisch ermittelt (vgl. Krieger, 1992).

Der erste Schritt in einem Qualifikationsmodell müßte somit die Beantwortung der Frage nach den notwendigen Kompetenzen und Fähigkeiten der Hochschullehrer und -lehrerinnen sein, vor allem nach dem Teil der Kompetenzen bei Hochschullehrenden, der über die fachwissenschaftliche Qualifikation hinausgeht. So fordert Preißer von den Lehrenden: "Sie müssen das lernen, was in vergleichbaren Positionen in der freien Wirtschaft gang und gebe ist: Arbeitseinsatz und Mitarbeiterführung (wie kommuniziere ich mit meiner Sekretärin und mit meinen Assistenten), Entscheidungsvorbereitung (für Gremiensitzungen eine unabdingbare Voraussetzung), Problemlösungstechniken, Haushaltsplanung (ohne sie wird die Hochschule keine Ressourcen intelligent einsparen können), Arbeitsplanung (weil sie permanent keine Zeit haben, aber auch permanent Zeit verträdeln); Präsentations- und Kommunikationstechniken (damit auch die Lehre, aber auch Vorträge auf Kongressen besser werden) usw." (Preißer, 1992, S. 204).

Studentische Veranstaltungskritik wäre im Qualifikationsmodell das Instrument zur Ermittlung von Qualifikationsbedarf, sowohl bei dem einzelnen Hochschullehrer oder der einzelnen Hochschullehrerin als auch über den gesamten Lehrkörper gesehen. Diese Forderung wirkt sich auf den Anspruch an den zu benutzenden Fragebogen aus. Ein solches Instrument muß zum einen eine relativ präzise Stärken-Schwächen-Analyse ermöglichen, zum anderen aber auch über hohe soziale Validität verfügen.

Ähnlich wie beim Kommunikationsmodell ist am Qualifikationsansatz bedenklich, daß es allein von der Motivation der Lehrenden abhängt, ob Verbesserungen angestrebt und durchgeführt werden. Ist diese Motivation vorhanden, so reicht manchmal schon das studentische Feedback, um eine ständige Optimierung der eigenen Veranstaltung durchzuführen. Betrachtet man in Deutschland allein das Interesse an hochschuldidaktischen Qualifikationsmöglichkeiten, so ist zu bezweifeln, ob die für diesen Ansatz notwendige Motivation vorhanden ist: "Hochschuldidaktische Fragen sind - mit Ausnahmen für die Fachwissenschaftler der Disziplin Hochschuldidaktik - kein aktuelles Thema unter deutschen Professoren und auch nicht innerhalb des akademischen Mittelbaus" (Krieger, 1992, S.56). Trotz des Interesses an solchen Fortbildungen hat sich die beschriebene Situation noch nicht grundlegend verändert.

Insgesamt kann geschlossen werden, daß Absichten zur Verbesserung der Lehre, die sich nur auf das 'reine' Qualifikationsmodell stützen, die notwendige Basis dafür nicht erstellen, zumal wenn Lehrkompetenz nicht verbindliche Grundlage von Einstellung und Berufungen wird und Weiterbildung nur freiwillig bleibt. Eine wichtige Voraussetzung zur Anwendung des Qualifikationsmodells wäre zunächst, eine stärkere Verpflichtung für die didaktische Aus- und Weiterbildung zu verankern.

### **Steuerung: Anreize und Sanktionen**

Steuerung ist kein Ziel an sich, sondern ein Prinzip, um Erwünschtes zu fördern und Negatives zu unterbinden. Mittel einer Steuerung sind somit Anreizsysteme und Sanktionsmöglichkeiten. Vor allem zwei Richtungen werden mit dem Gedanken eines Steuerungsmodells an den Hochschulen verfolgt: einen effizienteren Mitteleinsatz zu ermöglichen und eine Verbesserung der Lehre zu bewirken. Steuerungsmechanismen können sowohl von staatlicher Seite den Hochschule gegenüber angewandt oder aber innerhalb der Hochschulen selbst praktiziert werden. Staatliche Steuerung kann direkt über rechtliche Vorschriften und indirekt über "Markt"-Mechanismen ausgeführt werden.

Unter den Steuerungsmöglichkeiten ist die Verteilung finanzieller Mittel das am häufigsten eingesetzte Instrument. Bevor der Aspekt der Mittelsteuerung mit Hilfe von Ergebnissen studentischer Veranstaltungskritik betrachtet wird, ist es not-

wendig, sich insgesamt anzuschauen, ob und inwieweit eine finanzielle Mittelsteuerung an den deutschen Hochschulen stattfindet. Dabei sind zwei Ebenen zu unterscheiden: zum einen der Mittelzufluß von staatlichen Stellen an die Hochschulen (z.B. durch Mittelzuweisung über Studierendenzahlen), zum anderen seine Verteilung innerhalb der Hochschulen.

Festgestellt werden kann, daß bisher die Haushalte der Hochschulen von Jahr zu Jahr zum allergrößten Teil identisch blieben. Die grobe Verteilung der einzelnen Haushaltsposten änderte sich über die Jahre kaum. Eine Mittelsteuerung von staatlicher Seite findet immer noch vor allem über sog. Sonderprogramme statt, mit denen Mittel zweckgebunden an die Hochschulen ausgeschüttet werden. Die interne Mittelsteuerung liegt in den meisten Hochschulen erst in den Anfängen.

Verteilungskämpfe sind kaum bekannt, der Status quo bleibt über Jahre erhalten. "Evaluationskonzepte, die eine Steuerungsfunktion erhalten sollen, sind andersartig zu gestalten. Voraussetzung für die Einführung derartiger Konzepte ist die Bereitschaft einer Universität, Ressourcen für die Lehre bedarfs- und leistungsgerecht zu verteilen. Im Bereich der Forschung ist dieses Prinzip anerkannt" (Kurth, 1992, S. 164).

Ist diese Bereitschaft vorhanden, kann innerhalb der Hochschulen studentische Veranstaltungskritik eine Grundlage der Mittelverteilung sein. "Ein denkbarer Weg, die Ergebnisse der Evaluation wirkungsvoll umzusetzen, ist die Kopplung der Zuweisung eines Teils der Ressourcen, die einem Lehrstuhl zur Verfügung stehen, an die Ergebnisse der Evaluation. Besonders belastete und nachgefragte Lehrstühle erhalten einen Ausgleich für Engagement in der Lehre. In erster Linie ist hier an Hilfskraftmittel zu denken. Ferner könnten zeitlich befristete Stellen an die Belastung mit Lehr- und Prüfungsaufgaben geknüpft werden. Denkbar ist auch, Räume variabel zuzuweisen" (Kurth, 1992).

Ein Beispiel für eine bereits weit entwickelte Mittelsteuerung ist an der Universität Dortmund zu finden. Dort wird ein beträchtlicher Teil der finanziellen Mittel mit Hilfe festgelegter Kriterien verteilt. Diese Kriterien beziehen sich allerdings noch auf rein quantitative Evaluationszahlen, wie z.B. die Anzahl der Studierenden in einem Studiengang.

Bereits solche Steuerungsmodelle stoßen anderenorts allerdings auf großen Widerstand: "Notwendig ist also ein Verfahren, das in Kriterien und Erträgen auch von den Handelnden akzeptiert wird, dessen Ergebnisse nicht für andere Zwecke (z.B. finanzielle Sanktionen) mißbraucht werden" (Hochschulrektorenkonferenz, 1994, S. 30).

Dieser Widerstand ist auch dann gegeben, wenn die Forderung von Preißer (1992) nach einer Stärkung der Selbstverwaltung, z.B. mit Hilfe einer höheren Deckungsfähigkeit der Haushaltstitel untereinander, erfüllt wird. Mitglieder des

Lehrkörpers sind auf Leistungsvergleiche, Haushaltsverhandlungen und Verteilungskämpfe nicht eingestellt. "Ein wichtiger Grund für die bisherige Beschränkung besteht sicher darin, daß bei uns eine kritische Selbst- und Fremdevaluation noch wenig geübt ist und noch eher als persönliche Zumutung, wenn nicht als unzulässiger Eingriff in die Lehrfreiheit gedeutet wird" (Gerlach, 1992, S. 2).

Bei diesem Vergleich muß jedoch bedacht werden, daß die Kultur der Kritik, die in der Forschung gepflegt wird, auch nicht sehr öffentlicher Natur ist. Natürlich werden Gutachten erstellt und die in Fachzeitschriften ausgetauschte Kritik tangiert gerade in den Geisteswissenschaften manchmal die Grenze zur persönlichen Feindschaft, zu dem sind die "großen Namen eines Faches" bekannt, doch systematische Vergleiche der Produktivität, z.B. über die Anzahl veröffentlichter Artikel oder über die Häufigkeit von Zitationen, finden nicht statt. Werden solche Vergleiche systematisch durchgeführt und veröffentlicht (z.B. Keul, Gigerenzer und Stroebe, 1993), führt das zu harscher Kritik.

Schlußfolgernd kann gesagt werden, daß direkte Leistungsvergleiche über die Lehrenden eines Faches nicht nur in der Lehre sondern auch in der Forschung unüblich sind. Somit beinhaltet die Bewertung von Lehrveranstaltungen für die Professoren und Professorinnen eine völlig neue Perspektive.

Eine Steuerung von staatlicher Seite kann auch über die indirekte Aufwertung der Lehre erfolgen. Im Vergleich zur Forschung nimmt Lehre einen weitaus geringeren Stellenwert ein, da Lehrqualifikationen nicht honoriert werden und sie keine Bedeutung für eine Hochschulkarriere haben. Eine Änderung der gesetzlichen Vorschriften für die Habilitations- und Berufsordnung, die in den Hochschulgesetzen der Länder verankert sind, wäre dafür notwendig.

Weiss (1991) beschreibt für Österreich solche Möglichkeiten: "Wie für den Schüler gibt es aber auch beim Universitätslehrer Beurteilungen, die Berechtigungen verleihen oder der Auslese dienen: Beurteilung des "didaktischen Geschicks" im Rahmen des Habilitationsverfahrens, Beurteilung der "didaktischen Fähigkeiten" im Rahmen des Habilitationsverfahrens oder der Verlängerung eines Dienstvertrages als Universitätsassistent" (Weiss, 1991, S. 36).

Wissenschaftsminister von Trotha (Baden-Württemberg) hat in einer Presseerklärung im Mai 1991 vorgeschlagen, daß bei Habilitations- und Berufungsverfahren stärker auf die Lehrbefähigung geachtet werden soll (Ministerium für Wissenschaft und Forschung, Baden-Württemberg, 1991). Dieser Vorschlag hatte jedoch nur Apellfunktion und kann nicht als Steuerungsmaßnahme betrachtet werden.

Ähnliches wird in anderen Bundesländern diskutiert. So wird in NRW im Aktionsprogramm zur Qualität der Lehre (Ministerium für Wissenschaft und Forschung, 1992) vorgeschlagen, an den Hochschulen 'Hospitationen' einzuführen, um wissenschaftliche Mitarbeiter und Mitarbeiterinnen in der Lehre anzuleiten,

das Habilitationsverfahren um eine didaktische Komponente zu erweitern, insgesamt die Lehqualifikationen bei Berufungsverhandlungen und das Lehrengagement bei Bleibeverhandlungen stärker zu berücksichtigen. Weiter wird die Einführung eines Lehrprobejahres bei Erstberufenen genannt.

Vergleichbare Vorschläge sind im Berliner Hochschulstrukturplan zu finden (Senatsverwaltung für Wissenschaft und Forschung, 1993). Ebenso sind mögliche Maßnahmen in den Materialien zur Studienstrukturreform in Hessen genannt (Hessisches Ministerium für Wissenschaft und Kunst, 1994), wobei dort ausgearbeitet wurde, nach welchen Kriterien man die im Rahmengesetz geforderte 'pädagogische Eignung' überprüfen könnte.

Die genannten Steuerungsmöglichkeiten greifen jedoch vor allem beim wissenschaftlichen Nachwuchs. Marianne Kriszio (1992) z.B. erwartet Auswirkungen studentischer Veranstaltungskritik am ehesten bei Nachwuchswissenschaftlern und -wissenschaftlerinnen, da dort die Einflußmöglichkeiten größer sind als bei bereits langjährigen Professoren und Professorinnen.

Eine weitere indirekte Steuerung ließe sich über sozialen Druck erzielen. Dies ist z.B. bei der Veröffentlichung der Ergebnisse studentischer Veranstaltungskritik möglich. Ein anderes Beispiel dafür wäre das in den Niederlanden praktizierte Verfahren der "peer-evaluation", bei dem Fachkollegen und -kolleginnen den entsprechenden Bereich besuchen und beurteilen (vgl. Acherman, 1992).

Ein direktes Steuerungsprinzip dagegen wäre z.B. die Schaffung entsprechender Anreizmechanismen. Diese Mechanismen könnten sich auf den jeweiligen Lehrstuhl beziehen, zum Beispiel Sondermittel für die Lehre beinhalten. Ein Beispiel für ein solches Anreizsystem wäre LARS: Aus dem Erfahrungsbericht des Programms zu "Leistungsanreizsysteme in der Lehre an den Fachhochschulen" (LARS) geht hervor, daß etwas mehr als 10% der an badenwürttembergischen Fachhochschulen lehrenden Professoren und Professorinnen 1993 teilnahmen. Gefördert wurden vor allem: der Einsatz neuer Medien und die Erprobung neuer Studien- und Arbeitsformen (Ministerium für Wissenschaft und Forschung, Baden-Württemberg, 1994). Ein direkt auf den Lehrenden/ die Lehrende bezogener Anreizmechanismus wären z.B. Gehaltszulagen für besonders gut beurteilte Lehrende.

Entsprechende Förderungen und Anreize könnten mit geringen Maßnahmen hohe Wirkungen gerade im Kreis des wissenschaftlichen Nachwuchses erzielen. Solche Programme erfordern häufig nicht einmal hohe Extrakosten. Ein Beispiel von der TH Darmstadt zeigt, daß Tutoren und Tutorinnen schon allein durch eine Lohnfortzahlung (bis zehn Stunden) für eine freiwillige Teilnahme an hochschuldidaktischen Aus- und Fortbildungsangeboten motiviert werden konnten (Heger, 1991, S. 66).

Außer Anreizen wären als Steuerungsmöglichkeiten noch Sanktionen zu nennen. Während Anreize vor allem die engagiertesten Lehrenden erreichen, würde mit Sanktionsmechanismen vor allem eine "Minimalsteuerung" (d.h. Steuerung, wenn eine bestimmte Norm unterschritten wird) erfolgen. Sanktionsmöglichkeiten dieser Art wären z.B. die Gewährung eines Forschungsfreisemesters bei kritisch beurteilten Lehrenden erst nach einer didaktischen Weiterbildung. Gerade die Forschungsfreisemester wären ein relativ leichter Einstieg für die Etablierung gewisser Steuerungsmechanismen. Sie sind begehrt, stellen somit einen realen Anreiz bei der Professorenschaft dar und sie sind eine 'freiwillige Leistung', die Kriterien für ihre Vergabe könnten daher relativ einfach verändert werden.

Wird studentische Veranstaltungskritik als Grundlage für Steuerungsmechanismen benutzt, müssen die ihr zugrundeliegenden Instrumente entsprechenden Gütekriterien genügen. Solche Fragebogen müssen sorgfältig konstruierte und überprüfte Instrumente sein. Ihre soziale Validität hat zwar nicht den Stellenwert wie im Qualifikationsmodell, ist aber dennoch ausgesprochen relevant. Bedenkt man jedoch Penfields (1978) Feststellung, daß systematische Lehrverbesserungen selbst beim Einsatz massiver Steuerungsmechanismen, wie z.B. Beförderung, in den Vereinigten Staaten nicht auftreten, so ist festzustellen, daß auch der alleinige Einsatz einzelner Steuerungsmechanismen nicht ausreicht, um entscheidende und weitreichende Steigerungen der Qualität der Lehre zu erreichen.

### **9.3 Gesamtkonzept und Strategien der Verwendung**

#### **Erstellung der Datengrundlage**

Instrumente der Datenerhebung müssen entwickelt werden, die entsprechenden Qualitätskriterien genügen. Lita Linzer Schwarz (1980) listet auf, welche verschiedenen Evaluationsmöglichkeiten es im Ganzen gibt: Unterrichtsbesuche, Lehrpläne, studentische Kommentare, Kollegenbeurteilungen, Selbstevaluation und studentische Veranstaltungskritik. Obwohl letzteres die wohl umfassendste und kostengünstigste Variante darstellt, ist es sinnvoll auch andere Datensammlungen in die Phase der Befunderhebung miteinzubeziehen.

Sowohl bei der Entwicklung als auch bei der Verwendung entsprechender Instrumente muß bereits vorher der Verwendungszweck deutlich sein. Es bestehen große Unterschiede bei den Anforderungen an Instrumente, die unterschiedlichen Zwecken dienen sollen. Es bedarf nur eines relativ einfachen Verfahrens, um einer Professorin ein kurzes Feedback zu geben, das sie als Grundlage für ein Gespräch mit Studierenden verwenden kann. Ganz andere Punkte müssen beachtet werden, wenn ein Verfahren z.B. dazu dienen soll, einem Privatdozenten bei der Wahl von didaktischen Weiterbildungsangeboten zu unterstützen. Sollen gar Res-

sources entsprechend der Evaluationsergebnisse verteilt werden, so müssen bei der Erhebung entsprechende Gütekriterien (Zuverlässigkeit, Meßgenauigkeit, Objektivität) noch stärker beachtet werden.

Sollen Fachbereiche beurteilt oder Studiengänge reformiert werden, müssen wiederum andere Daten gesammelt und aufgearbeitet werden. Dabei können Verfahren, die stärker für die Bewertung einzelner Kurse gedacht sind, wie die studentische Veranstaltungskritik im engeren Sinne, als ein Teil dieser Befundsammlung gesehen werden.

Zusammenfassend kann somit gesagt werden, daß bei Überlegungen, welches Instrument verwandt werden soll und welchen Kriterien es genügen muß, vor allem der Zweck der Erhebung ausschlaggebend sein sollte.

### **Nutzung der gewonnen Daten**

Um die Daten studentischer Veranstaltungskritik als Lehrevaluationen und Teil der Lehrberichtsgestaltung zur Verbesserung der Lehre wirksam nutzen zu können, reicht es nicht aus, sie nur zu erheben.

Sieht man von dem wissenschaftlichen Zwecken dienlichen Nutzungsmodell ab, so wird deutlich, daß der Qualitätsanspruch an die anzuwendenden Verfahren vor allem mit dem Verwendungszweck zusammenhängen sollte. Daraus ergibt sich auch, daß auf unterschiedlichen Wegen gewonnene Daten unterschiedlich verwendet werden sollten.

Eine Veröffentlichung gewonnener Daten ist zu folgenden Zwecken notwendig: Lehre gewinnt an Transparenz, Studierende können sich dementsprechend orientieren und die so hergestellte Öffentlichkeit soll zu einer stärkeren Sensibilisierung für Lehrbelange beitragen. Zumindestens kurz- bis mittelfristig ist davon auszugehen, daß mit solchen Maßnahmen das Thema „Lehrqualität“ stärker in der hochschulpolitischen Öffentlichkeit thematisiert würde.

Wird einen Schritt weitergegangen, so ist eine aktive Auseinandersetzung um die Lehrqualität im Dialog vorstellbar. Dies erscheint vor allem für die Verbesserung der Lehre im einzelnen Kurs sinnvoll. Weiterreichende Vorstellungen können damit wohl kaum verwirklicht werden.

Nutzt man entsprechende Daten zur weiteren Qualifikation der Lehrenden, so erscheint eine vertraulichere Verwendung der erhobenen Daten sinnvoll. Allerdings ist nach den bisherigen Erkenntnissen die didaktische Qualifikation der Lehrenden nicht das größte Problem der Lehrqualität (eher ihr Kommunikationsverhalten oder die Strukturierung und Abstimmung der Lehrangebote).

Steuerungsmodelle sehen direktere und umfassendere Mechanismen zur Lehrverbesserung vor. Durch Belohnungen und Sanktionen sollen Anreize geschaffen werden, die die Lehrqualität steigern sollen.

Die Verwendung gesammelter Daten im Sinne einzelner Nutzungsansätze erscheint als nicht ausreichend. Diese behandeln jeweils nur Teilaspekte der möglichen Verwendung. Es erscheint daher sinnvoll, sie zu einem Gesamtkonzept zu verbinden. Qualität, Verbesserung und Evaluation der Lehre müssen erst einmal stärker thematisiert werden, bevor weitere Maßnahmen folgen können. Sie müssen zudem in ein verbindliches Konzept des „Qualitätsmanagements“ eingebunden werden (vgl. auch Barz, 1995).

Auf direkter Ebene einzelner Kurse spielen in erster Linie die Kommunikation zwischen Lehrenden und Studierenden und erst in zweiter Linie die Qualifikation der Lehrenden eine Rolle. Dabei ist die Frage nach der „Qualität“ der Instrumente vernachlässigbar. Daten können mit relativ einfachen Verfahren gewonnen werden, wenn sie nur dazu dienen sollen, eine Grundlage für ein Feedbackgespräch zu bieten und u.U. nur kursintern öffentlich sind.

Instrumente ganz anderer Güte erscheinen notwendig, wenn Daten dazu verwendet werden sollen auf Fachbereichs- oder Hochschulebene Ressourcen zu verteilen oder z.B. Sanktionen zu formulieren. Aber auch hier spielen Kommunikation und Transparenz eine herausragende Rolle. Neue Steuerungsmechanismen vermögen Ängste und auch Abwehrmechanismen auszulösen. Daher erscheint es nicht sinnvoll, Ansätze zu solchen Maßnahmen per staatlicher Verordnung einzuführen.

Stellvertretend für andere steht Comellis (1991) Reaktion auf einen solchen Versuch: "Bei mir als Arbeits- und Organisationspsychologen, der sich seit mehr als zwei Jahrzehnten unter anderem auch mit Fragen und Problemen betrieblicher Beurteilungssysteme beschäftigt, provoziert hier Art und Vorgehensweise (der Einführung studentischer Veranstaltungskritik in NRW/ Anm. d. Verf.) kritische Überlegungen. Ich weiß nur noch nicht: Soll mich der Zorn packen oder die Resignation?...kein Wirtschaftsunternehmen, wenn es sich auch nur einigermaßen sachkundig gemacht hätte, würde es wagen, in der hier zu beobachtenden Form ein Beurteilungssystem zu implementieren....Kein Beurteilungssystem kann ohne begleitende Schulungsmaßnahmen erfolgreich implementiert werden. Das bedeutet zunächst, daß allen Beteiligten das Beurteilungssystem erläutert und begründet wird. Dies geht nicht per Rundschreiben oder Erlaß, sondern nur durch Schulung und Diskussion" (S. 12).

Die Einführung von Steuerungsmaßnahmen erscheint danach nur vertretbar, wenn sie von einer ausführlichen Kommunikations- und Unterstützungsstrategie begleitet werden.

## **Chancen und Grenzen der Lehrevaluation**

Es gibt prinzipiell zwei Gründe für Schwierigkeiten im Bereich der Lehre und der Lehrorganisation. Kulik und McKeachie (1975, zitiert nach Friedlander, 1978) unterscheiden zwei Faktoren: fehlende Motivation und fehlende Qualifikation. Eine unzureichende Qualität der Lehre kann entweder auf Qualifikationsmängel der Lehrenden zurückzuführen sein oder darauf, daß Lehrenden andere Bereiche wichtiger sind, sie also weniger motiviert sind (bzw. motiviert werden), sich intensiver mit diesem Teil ihrer Aufgabe auseinanderzusetzen.

Mit Hilfe von Steuerungsmechanismen (von denen ein indirektes Mittel öffentlicher Druck ist) wäre es möglich, dafür zu sorgen, daß sich Hochschulen und Lehrende mit bestimmten Themen auseinandersetzen. Diese Steuerungsmöglichkeiten beziehen sich vor allem auf mögliche Änderungen von Rahmenbedingungen (Gesetzgebung, Mittelzuweisung).

Hauptansatzpunkt ist dabei die Stellung der Lehre an den Hochschulen. Deren Stellung läßt sich zu einem gewissen Teil durch Steuerungsmechanismen beeinflussen. Wird der Wert der Lehre erhöht, so würde sich dies auf alle Ebenen der Hochschule (Gesamtheit, Fachbereich, Lehrveranstaltung) auswirken. Hochschulleitungen könnten eher eine interne Steuerung entsprechend solcher Faktoren vornehmen, Fachbereiche den organisatorischen Aspekt der Lehre als wichtig begreifen, einzelne Dozenten und Dozentinnen ihre Lehrverpflichtungen ernster nehmen. Dabei spielt die Frage nach der Qualifikation der Lehrenden eine wichtige Rolle.

Die Anforderungen bezüglich Lehrqualifikationen und Managementfähigkeiten haben sich nicht nur aufgrund der gestiegenen Studierendenzahlen drastisch erhöht. Dies hat jedoch weder in einer verstärkten Fortbildung noch im Ausbildungskanon der Nachwuchswissenschaftler und -wissenschaftlerinnen Spuren hinterlassen. Es müssen somit Möglichkeiten der Qualifikation (auch im Bereich von Managementqualifikationen) eröffnet und Lehrende motiviert werden, diese wahrzunehmen.

Oder man entscheidet sich dafür, Lehrende stärker als bisher von Aufgaben, für die sie nicht ausgebildet wurden, zu entlasten und in diesem Bereich qualifizierte und ausgebildete Verwaltungsfachleute zu beschäftigen. Dabei geht es vor allem um Bereiche der Lehrorganisation und -verwaltung; zugleich müssen die Konsequenzen eines solchen Schrittes bedacht werden: Eine Entlastung der Lehrenden in diesem Bereich ist ohne eine gewisse Umverteilung der Entscheidungsspielräume und eine finanzielle Mehrausstattung der Fachbereiche nicht möglich.

Daher erscheint es in Deutschland wahrscheinlicher, daß diese Aufgaben weiterhin von den Lehrenden wahrgenommen werden; diese dafür allerdings besser

ausgebildet werden sollten. Dabei kommen der Motivierung durch Kommunikation und verschiedenen Steuerungsmechanismen eine große Bedeutung zu.

Größer dagegen erscheinen Möglichkeiten der Motivationssteigerung beim wissenschaftlichen Nachwuchs, der sich "noch in der Ausbildung" befindet. Dort müssen die Anforderungen deutlich werden, die inzwischen an den Hochschulen, im Wissenschaftsbetrieb und in der Lehre notwendig sind.

Die Debatte um Strategien der Verwendung von Lehrberichten und (studentischer) Lehrevaluation wird von zwei Polen markiert: Soll Veränderung durch Kommunikation oder durch Steuerung erreicht werden? Als 'Veränderung durch Kommunikation' können Appelle, Diskussionen, aber auch Preisverleihungen u.ä. gelten. Dies alles sind Mechanismen, die vor allem dazu führen können, Lehrende und die Öffentlichkeit für das Thema zu sensibilisieren. Dahinter steht die Hoffnung, daß damit die Lehre ganz selbstverständlich aufgewertet wird und diese Aufwertung zu allgemeinen Verbesserungen der Lehrqualität führt.

Eine Veränderung an den Hochschulen ohne den Einsatz von Steuerungsmechanismen ist jedoch nicht wahrscheinlich. Dies hat verschiedene Gründe: "Evaluation bedeutet immer potentiellen Veränderungsdruck; das muß in restriktiv-stabilen Systemen Ängste und Abwehr erzeugen" (Heger, 1991, S. 53). "Die Evaluation der Lehre gefährdet potentiell die bestehenden Strukturen und eingefahrenen Verfahren in der Hochschule, indem sie transparent gemacht und der öffentlichen Diskussion zugänglich gemacht werden. ... Die Eigengesetzlichkeit der hochschulpolitischen Rituale blockiert grundlegende Reformmaßnahmen.... Der Konservatismus des Bestandsinteresses, des Interesses der Institution an sich selbst ... ist wichtiger als die Notwendigkeit zur Reform" (Preißer, 1992, S. 205).

Ein Gesamtkonzept zur Einführung von Lehrbewertungen muß somit Folgemaßnahmen berücksichtigen und alle genannten Verwendungsmöglichkeiten beachten. Ohne eine neue Steuerungsstrategie, abgestimmt zwischen staatlicher Seite, den Gremien der Hochschulleitungen und Fachbereichen, besteht ansonsten die Gefahr einer reinen Symptombekämpfung.

Zusammenfassend kann gesagt werden, daß ein Ansatz notwendig erscheint, der das gesamte Lehrsystem umfaßt und dort Maßnahmen eines Qualitätsmanagements etabliert. Dafür müßten folgende Voraussetzungen geschaffen werden:

- Es muß eine gesicherte und regelmäßig zu erneuernde Datengrundlage geben. Diese muß sowohl die quantitativen Daten (Studierendenzahlen, Studiendauer, etc.) als auch Beurteilungsdaten (Ergebnisse studentischer Veranstaltungskritik, organisatorische Probleme im Fachbereich, etc.) erfassen.
- Weiter sind notwendig: Kriterien, nach denen diese Daten beurteilt werden sollen und entsprechende Gremien, die dies auch konsequent handhaben.

- Sowohl Verbesserungsvorschläge als auch weitere Empfehlungen müssen nicht nur formuliert, sondern in ihrer Umsetzung kontrolliert werden.
- Um eine höhere Verbindlichkeit der jeweiligen Empfehlungen zu ermöglichen, müssen entsprechende Rahmenbedingungen geschaffen werden (klare Verantwortlichkeiten, Anreiz- und Sanktionssysteme, etc.).
- Die genannten Punkte müssen nicht nur für Lehrende, sondern auch für die Fachbereiche und für die einzelnen Hochschulen gelten.

Qualitätsmanagement bedeutet nicht, daß an einem einzelnen Zeitpunkt Mißstände aufgedeckt und behoben werden, um die Sache wieder für einige Jahre auf sich beruhen zu lassen. Sondern es ist ein Verfahren der kontinuierlichen Qualitätssicherung und -entwicklung mit Hilfe von regelmäßig durchgeführten Bestandsaufnahmen, daraus abgeleiteten Folgerungen, verantwortlicher Umsetzung und erneuter Rechenschaft bzw. Vergewisserung.

Der von der Hochschulrektorenkonferenz formulierte Anspruch, daß „Evaluation nicht (einseitig) als Instrument zur Behebung von Defiziten gesehen wird, sondern als Anspruch des ständigen Bemühens um einen hohen Standard und um kontinuierliche Reflexion des individuellen wie institutionellen Handelns“; ... (Hochschulrektorenkonferenz, 1994, S.30) umreißt zutreffend den Grundgedanken von Qualitätsmanagement und den Stellenwert der Lehrevaluation. Entsprechend zielen auch die Empfehlungen des Wissenschaftsrates darauf ab, Evaluation und Veranstaltungskritik nicht als Selbstzweck zu sehen, sondern als Beitrag zur Stärkung der Lehre in den Hochschulen und als wichtiges Element zur Qualitätsentwicklung in der Lehre (vgl. Wissenschaftsrat 1996). Sowohl die Thesen und Ansprüche der Hochschulrektorenkonferenz als auch die Erfahrungen und Empfehlungen des Wissenschaftsrates bieten Grundlagen, Lehrevaluation und studentische Veranstaltungskritik als gewichtige Teile der Hochschulentwicklung anzuerkennen und zu verwenden.



## Literatur

- Abrami, P. C., Perry, R. P. und Leventhal, L. (1982). The Relationship Between Student Personality Characteristics, Teacher Ratings, and Students Achievement. *Journal of Educational Psychology*, 74, S. 111-125
- Acherman, H.A. (1992). Externe Qualitätsbewertung EQA (External Quality Assessment) - Selbstregulierung der Ausbildungsqualität im holländischen Hochschul-system. In Holtkamp, R. und Schnitzer K. (Hrsg.), *Evaluation des Lehrens und Lernens. Ansätze, Methoden, Instrumente*, S. 13-19
- Aleamoni, L. M. (1987). Typical Faculty Concerns About Student Evaluation of Teaching. *New Direction for Teaching and Learning*, 31, S. 25-31
- Aleamoni, L. M. und Thomas, G. (1980). Differential Relationships of Student, Instructor, and Course Characteristics to General and Specific Items On a Course Evaluation Questionnaire. *Teaching of Psychology*, 7 (4), S. 233-235
- Aleamoni, L.M. und Hexner, Pamela (1980). A Review of the Research on Student Evaluation and a Report on the Effect of Different Sets of Instructions on Student Course and Instructor Evaluation. *Instructional Science*, 9, S. 67-84
- Ames, R. und Lau, S. (1979). An Attributional Approach to the Validity of Student Ratings of Instruction. *Contemporary Educational Psychology*, 4, S. 26-39
- Arkin, R. M. und Maruyama, G. M. (1979). Attribution, Affect, and College Exam Performance. *Journal of Educational Psychology*, 71 (1), S. 85-93
- Ashton, R. H. (1975). Correlates of Ratings of Teaching Effectiveness Gordon's Survey of Interpersonal Values. *Psychological Reports*, 36, S. 890
- Bargel, T., Multrus, F. und Ramm, M. (1996). Studium und Studierende in den 90er Jahren. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (Hrsg.), Bonn
- Barz, A. (1995). Fünf Thesen über die Sorge um Qualität der Lehre. Vom 'muddling through' zum 'management'. In Technische Universität Dresden (Hrsg.), *Evaluation der Lehre. Ein Kolloquium an der Technischen Universität Dresden*, Dresden: University Press, S. 127-134
- Beatty, M. J. und Zahn, C. J. (1990). Are Student Ratings of Communication Instructors Due to 'Easy' Grading Practices?: An Analysis of Teacher Credibility and Student-Reported Performance Levels. *Communication Education*, 39, S. 275-282
- Blass, T. (1980). What Do Positive Correlations Between Student Grades and Teacher Evaluation Mean? *Teaching of Psychology*, 7 (3), S. 186-187

- Bledsoe, J. C. (1978). Insight into one's Own Teaching: Stability of Students' Evaluations Across Classes. *Psychological Reports*, 42, S. 1071-1074
- Blount, H. P., Stallings, W. M. und Gupta, V. G. (1978). The Effects of Different Instructions on Student Ratings of University Courses and Teachers. *Journal of Educational Research*, 71, S.149-152
- Brandenburg, D. C., Slinde, J. A. und Batista, E. E. (1977). Student Ratings of Instruction: Validity and Normative Interpretation. *Research in Higher Education*, 7, S. 67-78
- Bray, J. H. und Howard, G. S. (1980). Interaction of Teacher and Student Sex and Sex Role Orientations and Student Evaluations of College Instruction. *Contemporary Educational Psychology*, 5, S. 241-248
- Brown, D. L. (1976). Faculty Ratings and Student Grades: A University-wide Multiple Regression Analysis. *Journal of Educational Psychology*, 68 (5), S. 573-576
- Bruton, B. T. und Crull, Sue R. (1982). Causes and Consequences of Student Evaluation of Instruction. *Research in Higher Education*, 17 (3), S. 195-206
- Bundesministerium für Bildung und Wissenschaft (Hrsg.) (1993). *Qualität und Wettbewerb in der akademischen Lehre - Zwischenbilanz zum Modellprogramm des Bundesministeriums für Bildung und Wissenschaft*. Reihe: Bildung - Wissenschaft - Aktuell, Bd. 5, Bonn
- Centra, J. A. (1977). Student Ratings of Instruction and Their Relationship to Student Learning. *American Educational Research Journal*, 14, S. 17-24
- Centra, J. A. (1980). *Determining Faculty Effectiveness*. San Francisco: Jossey-Bass Publishers
- Chacko, T. I. (1983). Student Ratings of Instruction: A Function of Grading Standards. *Educational Research Quarterly*, 8 (2), S. 19-25
- Cohen, P. A. (1981). Student Ratings of Instruction and Student Achievement: A Metaanalysis of Multisection Validity Studies. *Review of Educational Research*, 51 (3), S. 281-309
- Coleman, J. und McKeachie, W.J. (1981). Effects of Instructors/ Course Evaluations on Student Course Selection. *Journal of Educational*, 73 (2), S. 224-226
- Comelli, G. (1991). Studentische Veranstaltungskritik. Mit wachen Augen trotzdem mitmachen! *Die Neue Hochschule*, 4 ,S. 12-14
- Cooper, Pamela J., Steward, Lea P. und Gudykunst, W. B. (1982). Relationship with Instructor and other Variables Influencing Student Evaluations of Instruction. *Communication Quarterly*, 30 (4), S. 308-314

- Cranton, Patricia A. und Smith, R. A. (1986). A New Look at the Effect of Course Characteristics on Student Ratings of Instruction. *American Educational Research Journal*, 23 (1), S. 117-128
- Daniel, H. D. (1994). Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung. *Empirische Pädagogik*, 1994, 8 (2), S. 109-130
- Daniel, H. D. (1995a). Bewertung der Lehre aus Sicht der Studierenden und Absolventen. In D. Müller-Böling (Hrsg.), *Qualitätssicherung in Hochschulen* (S.160-185). Gütersloh: Verlag Bertelsmann Stiftung
- Daniel, H. D. (1995b). Das Modellprojekt „Evaluation der Lehre“ an der Universität Mannheim. Teil 2: Statistische Auswertung von Befragungen in Lehrveranstaltungen. In Mohler, P. (Hrsg.), *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Münster: Waxmann Verlag
- Daniel, H. D. (1995c). Datenbedarf zu und Anfertigung von Lehrberichten. Erfahrungsbericht aus Baden-Württemberg. In Statistisches Bundesamt Wiesbaden (Hrsg.), *Effizienzbeurteilung der Hochschulausbildung auf statistischer Grundlage*, Stuttgart: Metzler-Poeschel, S. 62-65
- Daniel, H. D. (1995d). Die Wahl von Studienort und Hochschule. In E. Dichtl & M. Lingenfelder (Hrsg.), *Effizient studieren: Wirtschaftswissenschaften*, Wiesbaden: Gabler, S. 3-22
- Daniel, H. D. und Hornbostel, M. (1993). Provozierende Fragen. *SPIEGEL-SPEZIAL*, S. 157
- Daniel, H. D., Thoma, M. & Bandilla, W. (1995). Das Modellprojekt „Evaluation der Lehre“ an der Universität Mannheim. Teil 1: Planung und Durchführung von Befragungen in Lehrveranstaltungen. In Mohler, P. (Hrsg.), *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Münster: Waxmann Verlag, S. 83-92
- Danielsen, A. L. und White, R. A. (1976). Some Evidence on the Variables Associated with Student Evaluation of Teachers. *The Journal of Economic Education*, 7, S. 117-119
- DER SPIEGEL (1989). Die neuen Unis sind die besten. 50, S. 70-87
- DER SPIEGEL (1994). *Famose Ergebnisse*. 15, S. 55-57
- Detmer, H. (1991). Studentische Lehrkritik und Datenschutz. *Mitteilungen des Hochschulverbandes*, 6, S. 288
- Diehl, J. M. & Kohr, H. U. (1977). Entwicklung eines Fragebogens zur Beurteilung von Hochschulveranstaltungen im Fach Psychologie. *Psychologie in Erziehung und Unterricht*, 24, S. 61-75

- Diehl, J. M. (1983). *Untersuchungen zur Einstellung von Psychologiestudenten gegenüber Statistik und Statistik-Veranstaltungen*. Berichte aus dem Fachbereich 06 Psychologie, Universität Gießen
- Diehl, J. M. (1994). *Fragebogen zur studentischen Evaluation von Hochschulveranstaltungen. Manual zum VBVOR und VBREF*. Fachbereich 06 Psychologie (Abteilung Methodik), Universität Gießen
- Drews, D. R., Burroughs, J. W. und Nokvich, DeeAnn (1987). Teacher Self-Ratings as a Validity Criterion for Student Evaluations. *Teaching of Psychology*, 14 (1), S. 23-25
- Driscoll, Laura A. und Goodwin, W. L. (1979). The Effects of Varying Information About Use and Disposition of Results on University Students' Evaluation of Faculty and Courses. *American Educational Research Journal*, 16, S. 25-37
- Drummond, R. J. und McIntire, W. G. (1977). The Role of Cognitive Style in Student Evaluation of Instruction. *College Student Journal*, 11, S. 220-223
- DuCette, J. und Kenney, Jane (1982). Do Grading Standards Affect Student Evaluations of Teaching? Some New Evidence on an Old Question. *Journal of Educational Psychology*, 3, S. 308-314
- Elmore, Patricia B. und LaPointe, K. A. (1975). Effect of Teacher Sex, Student Sex, and Teacher Warmth on the Evaluation of College Instructors. *Journal of Educational Psychology*, 67 (3), S. 368-374
- Elmore, P. B. und Pohlmann, J. T. (1978). Effect of Teacher, Student, and Class Characteristics on the Evaluation of College Instructors. *Journal of Educational Psychology*, 70, S. 187-192
- Englert, Steffani (1993). *Fachkultur, Geschlecht und soziale Reproduktion*. Weinheim: Deutscher Studien Verlag
- Erichsen, H. U. (1993). *Hochschulen auf dem Wege - Jahresbericht 1993 des Präsidenten der Hochschulrektorenkonferenz (HRK) - Professor Dr. Hans-Uwe Erichsen*. Bonn: Hochschulrektorenkonferenz, S.1-15
- Faßnacht, G. (1979). *Systematische Verhaltensbeobachtung*. München: Ernst Reinhardt Verlag
- Firth, M. (1979). Impact of Work Experience on the Validity of Student Evaluation of Teaching Effectiveness. *Journal of Educational Psychology*, 71, S. 726-730
- Frank, Andrea (1991). Wodurch ist der Hochschullehrer/ die Hochschullehrerin wichtig für das studentische Lernen? In Webler, W. und Otto, H. (Hrsg.) *Der Ort der Lehre in der Hochschule. Lehrleistungen, Prestige und Hochschulwettbewerb*. Weinheim: Deutscher Studienverlag

- Friedlander, J. (1978). Student Perception on the Effectiveness of Midterm Feedback to Modify College Instruction. *The Journal of Educational Research*, 71, S. 140-143
- Gaffuri, Adriane, Wrench, D., Karr, C. und Kopp, R. (1982). Exploring Some Pitfalls in Student Evaluation of Teaching. *Teaching of Psychology*, 9 (4), S. 229-230
- Gerlach, J. W. (1992). Vorwort des Präsidenten der Freien Universität Berlin. In Grünh, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 1-2
- Giesen, H. und Jansen, R. (1983). Universität aus der Sicht ihrer Studenten. *Zeitschrift für Entwicklungspsychologische und Pädagogische Psychologie*, 15, S. 222-223
- Giffen, Katherine van (1990). Influence of professor gender and perceived use of humor on course evaluations. *Humor*, 3 (1), S.65-73
- Gillmore, G. M. (1977). How Large is the Course Effect? *Research in Higher Education*, 7, S. 187-189
- Gleich, J. M., Meran, G. und Bargel, T. (1982). *Studenten und Hochschullehrer*. Ministeriums für Wissenschaft und Kunst (Ba-Wü) (Hrsg.), Serie: Bildung in neuer Sicht, Bd. 48
- Gralki, H. O. und Hecht, Heidemarie (1992). Hochschuldidaktische Aspekte der Beurteilung von Lehrveranstaltungen durch Studenten. In Grünh, D. und Gattwinkel, H. (Hrsg.), *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 99-113
- Gralski, H. O. (1992). Provokative Evaluation - der falsche Weg zur Verbesserung der Lehre. In Grünh, D. und Gattwinkel, H. (Hrsg.), *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 35-41
- Grottian, P. (1992). Der geschlossene Vortrag - ein Lehrstück für die Pervertierung der Lehrbewertung. In Grünh, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 27-33
- Heckhausen, H. (1986). Die Pädagogische Psychologie vor neuen Herausforderungen. In Weidemann, B. und Krapp, A. (Hrsg.), *Pädagogische Psychologie*. S. 786-788
- Heger, M. (1991). Lehrveranstaltungskritik als angewandte Hochschuldidaktik. *Zeitschrift für Hochschuldidaktik*, 15, S. 43-69
- Hessisches Ministerium für Wissenschaft und Kunst (1994). *Der hessische Weg. Aktualisierte Materialien zur Studienstrukturreform*. Wiesbaden: Oscar Brandstetter GmbH

- Hessisches Ministerium für Wissenschaft und Kunst (Hrsg.) (1995). *Autonomie und Verantwortung - Hochschulreform unter schwierigen Bedingungen*. Bericht der Hochschulstrukturkommission des Landes Hessen. Frankfurt/ Main: Campus Verlags GmbH
- Hochschul-Informationen-System (1994). *Exmatrikuliertenbefragung an der Universität Bayreuth im Studienjahr 1993/94*. HIS Kurzinformationen, A11/94. Hannover: HIS
- Hochschul-Informationen-System (1995). *-Studienabbruch: Gründe und anschließende Tätigkeiten. Ergebnisse einer bundesweiten Befragung im Studienjahr 1993/94*. HIS Kurzinformationen, A1/95. Hannover: HIS
- Hochschulrektorenkonferenz (Hrsg.) (1994). *Qualität von Studium und Lehre*. Dokumente zur Hochschulreform, Bonn: Hochschulrektorenkonferenz
- Hochschulrektorenkonferenz (1995a). *Rundschreiben Nr.3/95 vom 06.03.95*
- Hochschulrektorenkonferenz (1995b). *Evaluation der Lehre*. Pressemitteilung HRK/16/95
- Hochschulrektorenkonferenz (1995c). *Zur Evaluation im Hochschulbereich unter besonderer Berücksichtigung der Lehre*. Entschließung des 176. HRK-Plenums vom 3.7.1995
- Hofman, J. E. und Kremen, Liya (1980). Attitudes Toward Higher Education and Course Evaluation. *Journal of Educational Psychology*, 72 (5), S. 610-617
- Hofman, J. M. (1988). Studienmotivation und Veranstaltungsbeurteilung. *Psychologie, Erziehung, Unterricht*, 35, S. 119-126
- Holtkamp, R. & Schnitzer, K. (1992). *Evaluationen des Lehrens und Lernens. Ansätze, Methoden, Instrumente*. Hannover: HIS
- Hufen, F. (1995). Rechtsfragen der Lehrevaluation an wissenschaftlichen Hochschulen, Rechtsgutachten, Bonn: Deutscher Hochschulverband
- Ipsen, D. & Portele, G. (1976). *Organisation von Forschung und Lehre an westdeutschen Hochschulen*. Serie: Hochschulplanung. München: Verlag Dokumentation
- Interdisziplinäres Zentrum für Hochschuldidaktik (1995). *Grundlinien und Alltag der Hochschuldidaktik in Bielefeld*. Bericht über die Jahre 1991-1994. Universität Bielefeld
- Keil, W. & Piontkowski, U. (1973). *Strukturen und Prozesse im Hochschulunterricht*. Weinheim, Basel: Beltz Verlag
- Keul, A. G., Gigerenzer, G. und Stroebe, W. (1993). Wie international ist die Psychologie in Deutschland, Österreich und der Schweiz? Eine SSCI-Analyse. *Psychologische Rundschau*, 4, S. 259-269

- Kovac, R. (1976). Personality Correlates of Faculty and Course Evaluations. *Research in Higher Education*, 5, S. 335-344
- Krieger, W. (1992). Evaluation von Hochschulunterricht. In Grün, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 45-59
- Kriszio, Marianne (1992). USA als Vorbild? - Bewertung der Lehre im Rahmen der gegenwärtigen Personalstruktur. *Forum Wissenschaft*, 1, S. 17-19
- Kriz, J. (1993). Wie gut sind unsere Universitäten? *Stern*, 1993, 16, S. 171-184
- Kromrey, H. (1993). Studentische Vorlesungskritik. Empirische Daten und Konsequenzen für die Lehre. *Soziologie*, Heft 1/1993, S. 39-56
- Kromrey, H. (1994a). Wie erkennt man gute Lehre? Was studentische Vorlesungsbefragungen (nicht) aussagen. *Empirische Pädagogik*, 2, S. 153-168
- Kromrey, H. (1995a). Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In Mohler, P. (Hrsg.), *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*, Münster: Waxmann Verlag, S. 105-128
- Kromrey, H. (1995b). Studentische Befragung zu Lehre und Studium. Von der Lehrevaluation zur Qualitätsentwicklung. In Technische Universität Dresden (Hrsg.), *Evaluation der Lehre. Ein Kolloquium an der Technischen Universität Dresden*, Dresden: University Press, S. 67-87
- Kromrey, H. (1995c). Buchbesprechung über „Das Heidelberger Inventor zur Lehrveranstaltungs-Evaluation“ von Rindermann, H. und Amelang, M.
- Kultusminister- und Hochschulrektorenkonferenz (1994). *Umsetzung der Studienstrukturreform*. Bonn: Hochschulrektorenkonferenz
- Kurth, M. (1992). Evaluationskonzepte zur Mittelsteuerung im Haushalt der Universität. In Grün, D. und Gattwinkel, H. (Hrsg.), *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 163-167
- Landeshochschulstrukturkommission (1992). *Stellungnahmen und Empfehlungen zu Struktur und Entwicklung der Berliner Hochschulen*. Berlin: Görlitz
- Lester, D. (1982). Students' Evaluation of Teaching and Course Performance. *Psychological Reports*, 50, S. 1126
- Levinthal, C. F., Lansky, L. M. und Andrews, O. E. (1971). Student Evaluations of Teacher Behaviors as Estimations of Real-Ideal Discrepancies: A Critique of Teacher Rating Methods. *Journal of Educational Psychology*, 62 (2), S. 104-109

- Liberales Hochschulgruppen (1992). *Lehre unter der Lupe*. Augsburg: Pinus Druck
- Linzer Schwartz, Lita (1980). Criteria for Effective University Teaching. *Improving College and University Teaching*, 28 (3), S. 120-123
- Lynch, D. J.; Tamburrino, M. und Nagel, R. (1989). Students' Reactions as Guide to Course Revision. *Psychological Reports*, 65, S. 989-990
- Marquez, T. E., Lane, D. M. und Dorfman, P. W. (1979). Toward the Development of a System for Instructional Evaluation: Is there Consensus Regarding What Constitutes Effective Teaching? *Journal of Educational Psychology*, 71 (6), S. 840-849
- Marsh, H. W. (1982). The Use of Path Analysis to Estimate Teacher and Course Effects in Student Ratings of Instructional Effectiveness. *Applied Psychological Measurement*, 6 (1), S. 47-59
- Marsh, H. W. (1982b). Validity of Students' Evaluations of College Teaching. A Multitrait-Multimethod Analysis. *Journal of Educational Psychology*, 74 (2), S. 264-279
- Marsh, H. W. und Hocevar, D. (1984). The Factorial Invariance of Student Evaluations of College Teaching. *American Educational Research Journal*, 21 (2), S. 341-366
- Marsh, W. und Ware Jr., J. E. (1982). Effects of Expressiveness, Content Coverage, and Incentive on Multidimensional Student Rating Scales: New Interpretation of the Dr. Fox Effect. *Journal of Educational Psychology*, 74 (1), S. 126-134
- Mazer, G. E. (1977). Evaluating the Evaluations: A Factor Analysis of Student Ratings. *Counselor Education and Supervision*, 17, S. 6-11
- McBean, E. A. und Lennox, W. C. (1985). Effect of Survey Size on Student Ratings of Teaching. *Higher Education*, 14, S. 117-125
- McGaghie, W. C. (1975). Student and Faculty Ratings of Instruction: Another Look. *Journal of Medical Education*, 50, S. 387-389
- McKeachie, W.J. und Lin, Y.G. (1978). A Note on Validity of Student Ratings of Teaching. *Educational Research Quarterly*, 4 (3), S. 45-47
- Megerle, K. (1991). Dozentenschelte allein führt nicht weiter. *Frankfurter Rundschau*, Nr. 276 vom 28.11.1991, S. 35
- Meier, R. S. und Feldhusen, J. F. (1979). Another Look at Dr. Fox: Effect of Stated Purpose for Evaluation, Lecturer Expressiveness, and Density of Lecture Content on Student Ratings. *Journal of Educational Psychology*, 71 (3), S. 339-345

- Meredith, G. M. (1981). Focus-Scan Learning Strategy Correlates of Students' Appraisal of Instruction. *Perceptual and Motor Skills*, 53, S. 620
- Meredith, G. M. (1981b). Preferred Length of Scales for Students' Evaluation of Instruction. *Perceptual and Motor Skills*, 53, S. 490
- Meredith, G. M. (1982). Grade-related Attitude Correlation of Instructor/ Course Satisfaction among College Students. *Psychological Reports*, 50, S. 1142
- Meredith, G. M. und Schmitz, E. D. (1986). Student-taught and Faculty-taught Seminars in Undergraduate Education: Another Look. *Perceptual and Motor Skills*, 62, S. 593-594
- Ministerium für Wissenschaft und Forschung (NRW) (1992). *Aktionsprogramm Qualität der Lehre. Abschlußbericht*. Düsseldorf: Neusser Druckerei und Verlag GmbH
- Ministerium für Wissenschaft und Kunst (BaWü) (1991). *Pressemitteilung 68/1991 vom Mai 1991*
- Ministerium für Wissenschaft und Kunst (BaWü) (1994). *Pressemitteilung 11/1994 vom 4.2.1994*
- Mishra, S. P. (1979). The Use of Instructors' Self-Selected Items in Evaluating Teaching Effectiveness. *Journal of Psychology*, 102, S. 173-177
- Mohler, P. (1995). *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Münster: Waxmann Verlag
- Morstain, B. R. (1977). Relationship of Student and Instructor Educational Orientations with Course Ratings. *Journal of Educational Psychology*, 69 (4), S. 388-398
- Müller-Wolf, H.-M. und Fittkau, B. (1971). Lehrverhalten von Hochschullehrern und seine Bedeutung für Einstellungen und Verhalten von Studenten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 3 (3), S. 165-180
- Multrus, F. (1995). Zur Lehr- und Studienqualität. Dimensionen, Skalen und Befunde des Studierendensurveys. Hefte zur Bildungs- und Hochschulforschung, Nr. 12. Arbeitsgruppe Hochschulforschung der Universität Konstanz. Konstanz
- Murray, H. G. (1983). Low-Interference Classroom Teaching Behaviors and Student Ratings of College Teaching Effectiveness. *Journal of Educational Psychology*, 75 (1), S. 138-149
- Naftulin, D. H., Ware, J. E. und Donnelly, F. A. (1973). The Dr. Fox Lecture: A Paradigm of Educational Seduction. *Journal of Medical Psychology*, 48, S. 630-635

- Neumann, Lily und Neumann, Y. (1981). Comparison of Six Lengths of Rating Scales: Students' Attitudes Toward Instruction. *Psychological Reports*, 48, S. 399-404
- Orpen, C. (1980). The Susceptibility of Student Evaluation of Lecturers to Situation Variables. *Higher Education*, 9, S. 293-306
- Overall, J. U. und Marsh, H. W. (1980). Students' Evaluations of Instruction: A Longitudinal Study of Their Stability. *Journal of Educational Psychology*, 72 (3), S. 321-325
- Pasen, R., Frey, P., Menges R. und Rath, G. (1978). Different Administrative Directions and Student Ratings of Instruction: Cognitive Versus Affective Effects. *Research In Higher Education*, 9, S. 161-167
- Peisert, H., Bargel, T. und Framhein, Gerhild (1988). *Studiensituation und studentische Orientierungen an Universitäten und Fachhochschulen*. In Schriftenreihe: Studien zu Bildung und Wissenschaft, Bundesministerium für Bildung und Wissenschaft (Hrsg.), 59
- Penfield, D. A. (1978). Student Ratings of college Teaching: Rating the Utility of Rating Forms. *The Journal of Educational Research*, 72, S. 19-23
- Petchers, M. K. und Chow, J. C. (1988). Sources of Variation in Students' Evaluation of Instruction in a Graduate Social Work Program. *Journal of Social Work Education*, 1, S. 35-42
- Preißer, R. (1991). Bewertung der Qualität der Lehre an der Technischen Universität Berlin. In Webler, W. D. und Otto, H. U. (Hrsg.) *Der Ort der Lehre in der Hochschule. Lehrleistungen, Prestige und Hochschulwettbewerb*. Weinheim: Deutscher Studienverlag
- Preißer, R. (1992). Verwirklichungsbedingungen der Evaluation der Lehre und der Verbesserung der Lehre - Konsequenzen aus den bisherigen Erfahrungen mit Lehrveranstaltungskritik. In Grün, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. S. 197-217, Berlin: Zentrale Universitäts-Druckerei, S. 197-217
- Prosser, M. und Trigwell, K. (1991). Student Evaluations of Teaching and Courses: Student Learning Approaches and Outcomes as Criteria of Validity. *Contemporary Educational Psychology*, 16, S. 293-301
- Reissert, R. (1992). *Dokumentation: Evaluation der Lehre - Aktuelle Aktivitäten an deutschen Hochschulen Teil 1+2*, Hochschul-Informationssystem (Hrsg.), Hannover: HIS
- Reissert, R. (1994). *Evaluation der Lehre - interne Selbsevaluation und externe Begutachtung durch Peers*. HIS-Kurzinformation A8/94, Hannover: HIS

- Richardson, M. D. (1978). A Study of Class Level Course Evaluation Trends. *College Student Journal*, 12, S. 132-134
- Richter, R. (Hrsg.). *Qualitätssorgen in der Lehre. Leitfaden für die studentische Lehrevaluation*. Neuwied: Luchterhand
- Rindermann, H. & Amelang, M. (1994a). *Das Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE)*. Heidelberg: Asanger
- Rindermann, H. & Amelang, M. (1994b). Entwicklung und Erprobung eines Fragebogens zur studentischen Veranstaltungsevaluation. *Empirische Pädagogik*, 8 (2), S. 131-151.
- Rindermann, H. (1996a/im Druck). *Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen: Analysen der Validität und zu Auswirkungen ihres Einsatzes anhand des HILVE*. Landau: Verlag Empirische Pädagogik.
- Rindermann, H. (1996b/im Druck). Verbesserung der Lehre durch Veranstaltungsevaluation? In E. Witruk (Hrsg.) *Tagungsband der 5. Fachgruppentagung Pädagogische Psychologie*. Leipzig
- Ring Christlich-Demokratischer Studierender (1992). *Mehr Ehre für die Lehre*. Erlangen: Union Betriebsgesellschaft (UBG)
- Ritter, U. P. (1992). Studentische Partizipation bei der Evaluation von Lehrveranstaltungen. *Das Hochschulwesen*, 2, S. 87-90
- Romeo, Felicia F. und Weber, W. A. (1985). An Examination of Variables which Influence Student Ratings of University Faculty. *College Student Journal*, 19, S. 133-140
- Romney, D. (1976). Course Effect vs. Teacher Effect on Student' Ratings of Teaching Competence Research. *Higher Education*, 5, S. 345-350
- Schmyk, C. (1992). Viele kleine Süppchen im großen Evaluationstopf. Evaluation der Lehre und hochschulpolitische Interessenlagen - ein Beitrag aus studentischer Sicht. In Grünh, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. Berlin: Zentrale Universitäts-Druckerei, S. 83-89
- Schnitzer, K. (1991). *Transparenz der Lehre auf amerikanisch*. In Hochschul-Informationen-System (Hrsg.). HIS-Kurzinformation A10/91. Hannover: HIS
- Schott, E. (1973). *Zur empirischen und theoretischen Grundlegung eines Bewertungsinstrumentes für Vorlesungen*. Serie: Blickpunkt Hochschuldidaktik, Bd. 28
- Scott, C. S. (1977). Student Ratings and Instructor-Defined Extenuationg Circumstances. *Journal of Educational Psychology*, 69, S. 744-747

- Seiler, L. H., Weybrigt, L. und Stang, D. J. (1977). How Useful Are Published Evaluations Ratings to Student Selecting Courses and Instructors? *Teaching of Psychology*, 4, S. 174-177
- Senatsverwaltung für Wissenschaft und Forschung Berlin (1993). *Berliner Hochschulstrukturplan*. Berlin
- Shatz, M. und Best, J. B. (1986). Selection of Items for Course Evaluation by Faculty and Students. *Psychological Reports*, 58, S. 239-242
- Sherman, Barbara R. und Blackburn, R. T. (1975). Personal Characteristics and Teaching Effectiveness of College Faculty. *Journal of Educational Psychology*, 67 (1), S. 124-131
- Snyder, C. R. und Clair, M. (1976). Effects of Expected and Obtained Grades on Teacher Evaluation and Attribution of Performance. *Journal of Educational Psychology*, 68 (1), S. 75-82
- Sommer, J. und Petermann, F. (1978). Deskriptive und präskriptive Aussagen in einem Fragebogen zur Beurteilung akademischer Lehrveranstaltungen. *Zeitschrift für Entwicklungspsychologie und Pädagogischer Psychologie*, 10 (4), S. 336-346
- SPIEGEL-SPEZIAL (1993). Welche Uni ist die beste? Bd. 3, Hamburg
- Stillman, Paula L., Gillers, M., Heins, M., Nicholson, G. und Sabers, D. L. (1983). Effect of Immediate Student Evaluations On a Multi-Instructor Course. *Journal of Medical Education*, 58, S. 176-178
- Strom, B., Hocevar, D., Zimmer, J. und Michael, W. B. (1982). The Course Structure Inventory: Discriminant and Construct Validity. *Educational and Psychological Measurement*, 42, S. 1125-1133
- Sturm, M. (1991). Lehrveranstaltungsbewertung als Grundlage für die kontinuierliche Verbesserung der universitären Lehre. *Zeitschrift für Hochschuldidaktik*, (1-2), S. 16-34
- Süllwold, F. (1992). Welche Realität wird bei der Beurteilung von Hochschullehrern durch Studierende erfahren? *Mitteilungen des Hochschulverbandes*, 1, S. 34-36
- Terry, R. L. und McIntosh, D. E. (1988). Do Students' Expectancies Affect Their Course Evaluations? *Educational and Psychological Measurement*, 48, S. 787-798
- Thomas, D., Ribich, F. und Freie, J. (1982). The Relationship Between Psychological Identification with Instructors and Student Ratings of College Courses. *Instructional Science*, 11, S. 139-154
- Thorndike, R. L. und Hagen, Elisabeth (1969). *Measurement and Evaluation in Psychology and Education*. New York: John Wiley & Sons, Inc.

- Thurston Barnett, L. und Littlepage, G. (1979). Course preferences and evaluations of male and female professors by male and female students. *Bulletin of the Psychonomic Society*, 13 (1), S. 44-46
- Ware Jr., J. E. und Williams, R. G. (1975). The Dr. Fox Effect: A Study of Lecturer Effectiveness and Rating of Instruction. *Journal of Medical Education*, 50, S. 149-156
- Webler, W.-D. (1992). Evaluation der Lehre - Praxiserfahrungen und Methodenhinweise. In Grün, D. und Gattwinkel, H. (Hrsg.) *Evaluation von Lehrveranstaltungen*. S. 143-161, Berlin: Zentrale Universitäts-Druckerei
- Webler, W.-D. (1993). Evaluation der Lehre: Praxiserfahrungen und Methodenhinweise. In Winkler, H. (Hrsg.) *Qualität der Hochschulausbildung*. (Wiss. Zentrum für Berufs- und Hochschulforschung der Universität GHS Kassel (Werkstattberichte 40). Kassel
- Webler, W.-D., Domeyer, V. und Schiebel B. (1993). *Lehrberichte: empirische Grundlagen, Indikatore Auswahl und Empfehlungen zur Darstellung der Situation der Lehre in Lehrberichten*. Bundesministerium für Bildung und Wissenschaft (Hrsg.). Bad Honnef: Bock Verlag
- Weinert, F. E. (1986). Lernforschung als eine zentrale Aufgabe der Pädagogische Psychologie, In Weidemann, B. und Krapp, A. (Hrsg.), *Pädagogische Psychologie*. S. 783-786, Urban Schwarzberg: Psychologische Verlags Union
- Weiss, R. (1991). Ziele und Probleme einer Lehrveranstaltungskritik. *Zeitschrift für Hochschuldidaktik*, 15, S. 35-42
- Westdeutsche Rektorenkonferenz (1986). *Zur Beurteilung und Entwicklung der Ansätze zur Leistungsbewertung und -messung von Hochschulen. Stellungnahme des 149. Plenum der Westdeutschen Rektorenkonferenz*, Bonn
- Williams, R. G. und Ware, J. E. Jr. (1977). An Extended visit with Dr. Fox: Validity of Student Satisfaction with Instruction Ratings after Repeated Exposures to a Lecturer. *American Educational Research Journal*, 14 (4), S. 449-457
- Wissenschaftsrat (1996). *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Drs. 2365/96. Köln
- Yates, J. W. und Karmos, J. S. (1971). Final Grade Predictions and Instructional Evaluation at Different Levels of Course Achievement. *Journal of Instructional Psychology*, 4 (3), S. 38- 40



## Anhang 1

### Ansprechpartner und -partnerinnen für die Lehrevaluation

Instituion	Name	Telefonnr.
<b>Bund</b> Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie, Heinemannstr. 2, 53175 Bonn	Herr Lömker	0228-572219
<b>Landesministerien</b> Baden-Württemberg: Ministerium für Wissenschaft und Forschung, Königstr. 46, 70173 Stuttgart	Herr Dr. Kneser	0711-279-3174
Bayern: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst, Salvatorplatz 2, 80333 München	Herr Fiebler	089-2186-2366
Berlin: Senatsverwaltung für Wissenschaft und Forschung, Bredtschneiderstr. 5-8, 14057 Berlin	Herr Reich	030-3065-5780
Brandenburg: Ministerium für Wissenschaft, Forschung und Kultur, Friedrich-Ebert-Str. 4, 14473 Potsdam	Herr Dr. Boschan	0331-866-4750
Bremen: Senator für Bildung und Wissenschaft, Rembertiring 8-12, 28195 Bremen	Frau Dr. Sywottek	0421-361-6864
Hamburg: Behörde für Wissenschaft und Forschung, Hamburger Str. 37, 22083 Hamburg	Frau Tursich 3814	040-2984
Hessen: Ministerium für Wissenschaft und Kunst, Rheinstr. 23-25, 65185 Wiesbaden	Herr Weber	0611-165-510
Mecklenburg-Vorpommern: Kultusministerium, Werderstr. 124, 19055 Schwerin	Herr Vosz	0385-588-7301
Niedersachsen: Niedersächsisches Ministerium für Wissenschaft und Kultur, Leibnizufer 9, 30169 Hannover	Frau Valentien	0511-120-2452
Nordrhein-Westfalen: Ministerium für Wissenschaft und Forschung, Völklinger Str. 49, 40221 Düsseldorf	Frau Jorzek	0211-896-4403
Rheinland-Pfalz: Ministerium für Bildung und Kultur, Mittlere Bleiche 61, 55116 Mainz	Herr Mentges	06131-16-2954
Saarland: Ministerium für Wissenschaft und Kultur, Hohenzollernstr. 60, 66117 Saarbrücken	Frau Klos	0681-503-305
Sachsen: Sächsisches Staatsministerium für Wissenschaft und Kunst, Wigardstr. 17, 01097 Dresden	Herr Dr. Binjoschek	0351-564-6320

Sachsen-Anhalt: Kultusministerium, Breiter Weg 31, Breiter Weg 31, 39104 Magdeburg	Frau Dr. Berg	0391-567-7672
Schleswig-Holstein: Ministerium für Wissenschaft, For- schung und Kultur, Düsternbrooker Weg 64, 24105 Kiel	Herr Schlütz	0431-988-5822
Thüringen: Thüringer Ministerium für Wissenschaft, For- schung und Kultur, Juri-Gagarin-Ring 158, 99084 Erfurt	Frau Muck	0361-5966-142
<b>Weitere Institutionen</b> Hochschulrektorenkonferenz (HRK), Ahrstr. 39, 53175 Bonn	Frau Dr. Steiger	0228-887-140
Ständige Konferenz der Kultusminister der Länder (KMK), Nassestr. 8, 53113 Bonn	Frau Ackermann- Schneiders	0228-501-299
Wissenschaftsrat (WR), Brohler Str. 11, 50968 Köln	Herr Dr. Husung	0221-3776-228
Verbund der norddeutschen Hochschulen, Universität Hamburg, Edmund-Siemers-Allee 1, 20146 Hamburg	Frau Dr. Fischer-Bluhm	040-41236317
Niedersächsische Evaluationsagentur, Universität Hannover, Welfengarten 1, 30167 Hannover	Herr Reuke	0511-7623531
Interdisziplinäres Zentrum für Hochschuldidaktik (IZHD), Universität Bielefeld, Universitätsstr. 25, 33615 Bielefeld	Herr Dr. Webler	0521-106-4680
Hochschul-Informations-System (HIS), Goseriende 9, 30159 Hannover	Herr Dr. Reissert	0511-1220-244
Gemeinnütziges Centrum für Hochschulentwicklung (CHE), Carl-Bertelsmann-Str. 256, 33311 Gütersloh	Herr Dr. Barz	05241-9761-25
Arbeitsgruppe Hochschulforschung, Universität Konstanz, Sozialwissenschaftliche Fakultät, Postfach 5560, 78434 Konstanz	Frau el Hage	07531-882896
<b>Ansprechpartner für Instrumente</b> Professor Dr. Diehl, Universität Gießen, Fachbereich 06 Psychologie, Otto-Behagel-Str. 10, 35394 Gießen		0641-702-1
Professor Dr. Kromrey, FU Berlin, Institut für Sozio- logie, Babelsbergerstr. 14-16, 10715 Berlin		030-85002-229
Dr. Rindermann, Ludwig-Maximilian-Universität, Institut für Pädagogische Psychologie, Leopoldstr. 13, 80802 München		089-21806289



Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie unentgeltlich abgegeben. Sie ist nicht zum gewerblichen Vertrieb bestimmt. Sie darf weder von Parteien noch von Wahlbewerbern oder Wahlhelfern während eines Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für Bundestags-, Landtags- und Kommunalwahlen sowie für Wahlen zum Europäischen Parlament. Mißbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen und an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung.

Unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Schrift dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Bundesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte.



**bmb+f**

Bundesministerium für  
Bildung, Wissenschaft,  
Forschung und Technologie