

Beurteilung von Lehrveranstaltungen durch Studierende

- 1. Evaluation an Hochschulen**
- 2. Merkmale guter Lehre**
- 3. Können Studierende die Hochschullehre angemessen beurteilen?**
- 4. Auswirkungen und Nutzen studentischer Lehrevaluationen**
- 5. Empfehlungen zur Durchführung von Lehrevaluation**
- 6. Anforderungen an Praxis und Rahmenbedingungen von Lehrevaluationen**
- 7. Literatur**
- 8. Lebenslauf**

Seit mehr als 10 Jahren werden an deutschen Universitäten Lehrveranstaltungen bewertet. Lehrevaluationen bleiben jedoch umstritten, zudem tragen sie bei konventioneller Anwendung – keine Anknüpfung an Beratung und Weiterbildung sowie gleich bleibende Rahmenbedingungen – häufig nicht zu einer Verbesserung der Lehre bei. Möglichkeiten, wie Lehrveranstaltungsevaluationen durchgeführt werden können, damit sie zweckdienlich sind, werden auf dem Hintergrund nationaler und internationaler Erfahrungen und auf Basis wissenschaftlicher Studien des Autors diskutiert.

1. Evaluation an Hochschulen

Evaluationen befassen sich mit Institutionen, Personen, Programmen und Produkten. Über adäquate *Beschreibung* hinausgehend sollen Evaluationen *bewerten* und bestehende Prozesse *optimieren*. Vor allem in angelsächsischen Ländern spielt Evaluation der Lehre eine wichtige Rolle: So verwenden über 90% aller US-amerikanischen Hochschulen studentische Beurteilungen. Ergebnisse dieser Befragungen

- helfen einzelnen Dozenten, Schwächen eigener Lehre zu erkennen und abzustellen (Qualifikationsmodell),
- fördern die Kommunikation zwischen Lehrenden und Studierenden über die Lehre (Kommunikationsmodell),
- dienen den Studenten bei der Kurswahl (Transparenzmodell),

- werden von der Hochschule als ein Qualifikationsmaß bei Bleibe-, Gehalts- und Berufungsverhandlungen eingesetzt (Steuerungsmodell)
- und dienen gemeinsam mit Forschungsleistungen dem Vergleich zwischen Universitäten (Transparenzmodell).

Gute Lehre ist nicht nur durch didaktisch erfolgreiche Lehrveranstaltungen erreichbar, sondern muß in ein entsprechendes Umfeld in Form eines inhaltlich adäquaten und gut strukturierten Curriculums und in Form von günstigen institutionellen (Studienbedingungen an einem Fachbereich) und gesellschaftlichen Rahmenbedingungen (Zulassungssystem, Finanzierung u.ä.) eingebettet sein. Hierzu – und im besonderen bei inhaltlichen Fragen – sollten nicht nur studentische Einschätzungen, sondern auch Selbst- und Fachkollegen-Beurteilungen sowie quantitative Indikatoren herangezogen werden.

2. Merkmale guter Lehre

Was zeichnet „gute Lehre“ aus? Welche Merkmale charakterisieren gute Dozenten, welche zusätzlich gute Veranstaltungen? Zur Beantwortung dieser Fragen können Lehr- und Ausbildungszielkriterien definiert werden. Ebenso lassen sich instruktionspsychologische Theorien heranziehen oder Gesetzestexte, Bildungsvorstellungen, Curricula und Lehrevaluationsbögen interpretieren. Der direkteste Weg, Kennzeichen guter Lehre zu finden, besteht darin, Dozenten und Studenten, Personen also, die tagtäglich gute oder schlechte Lehre, Vorträge und Veranstaltungen erleben, zu befragen.

Zu den für erfolgreiche Lehre verantwortlichen Lehrmethoden zählen zuallererst Strukturierung und Klarheit. Zu den Determinanten auf Seiten des Dozentenverhaltens gehören darüber hinaus auch verständliche Erklärungen und Darstellungen, rhetorische Kompetenz, Motivierung, Verarbeitungstiefe, Zeit- und Schwierigkeitsmanagement, Engagement, Kooperativität/Kommunikation und in Seminaren das Leiten von Diskussionen. Weitere Bedingungsvariablen des Erfolges einer Veranstaltung stellen aber auch das studentische Verhalten wie Fleiß, die veranstaltungsinterne Beteiligung und Referate dar. Dazu können Kontextvariablen (das gestellte Thema, die Anforderungen, Besuchszahl, Existenz einer Prüfung, Pflichtmäßigkeit des Besuches, Veranstaltungstyp u.ä. Rahmenbedingungen) den Lehrerfolg wie auch die Veranstaltungskritik beeinflussen (s. Abbildung 1).

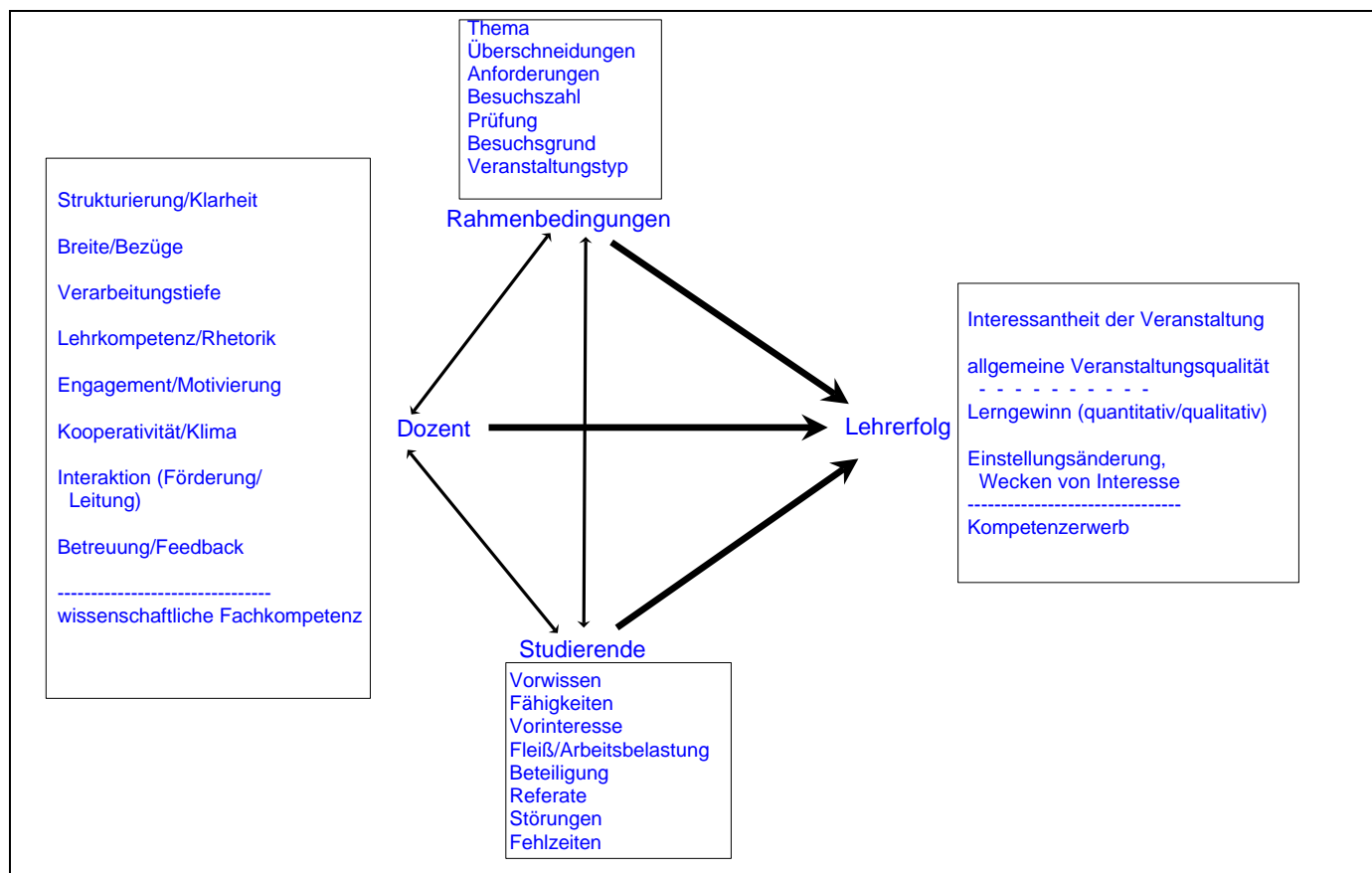


Abbildung 1: Bedingungsmodell des Lehrerfolgs (aus Rindermann, 2001)

Die Erhebung der Zufriedenheit mit der Lehrveranstaltung reicht zur Evaluation von Lehrveranstaltungen nicht aus. Allein gibt sie nur die *Bewertung* der Lehre wieder, eine differenzierte *Beschreibung* fehlt, welche Studenten und Dozenten Aufschluss über spezifische Stärken und Schwächen der Veranstaltung gibt und zwischen Aspekten des Dozenten, der Veranstaltung und der Studierenden unterscheiden kann.

Kritisch ist es, die Fachkompetenz des Dozenten oder die Inhalte der Veranstaltung durch Studierende beurteilen zu lassen, da beide Aspekte durch Studenten – zumindest in den Anfangssemestern – kaum adäquat beurteilt werden können. Zudem sind die Inhalte der Veranstaltung weitgehend durch Curricula den Dozenten vorgegeben. Sinnvoll ist jedoch, Themenrelevanz und –interessantheit (und Besuchsgrund) als „Biasvariablen“ zu erheben, weil diese Umstände – falls deren Einschätzungen unabhängig von der Lehre des Dozenten sind – die Beurteilung der Lehre verzerren können.

3. Können Studierende die Hochschullehre angemessen beurteilen?

Obwohl in der Regel gute durchschnittliche Beurteilungen von Lehrenden und Lehrveranstaltungen durch Studierende erzielt werden, ist die Skepsis vieler Professoren gegenüber studentischen Evaluationen nicht übersehbar. Selbst Gerichte kamen zur Ansicht, dass Studenten „zu einer objektiven Beurteilung generell nicht in der Lage seien“. Diese Äußerungen stehen aber in auffälligem Kontrast zum internationalen Forschungsstand. So vertreten z.B. Marsch und Roche (1997) die Überzeugung, dass studentische, über Veranstaltungen gemittelte Be-

urteilungen multidimensional, messgenau und stabil, primär eine Funktion des Lehrverhaltens des Dozenten und nicht der Kursthemen, relativ valide hinsichtlich vieler Indikatoren effektiver Lehre (z.B. Lerntests, Fremdbeurteilungen) und relativ unbeeinflusst von einer Vielzahl von potentiellen Verzerrungsvariablen (z.B. Geschlecht, Intelligenz) sind und als nützlich betrachtet werden: für Dozenten als Feedback, für Studenten in ihrer Kurswahl und für die Hochschulleitung bei Personalentscheidungen.

Von den Kritikern an studentischen Lehrevaluationen wird oft die Heterogenität der Urteile oder mangelnde Vergleichbarkeit von Veranstaltungskritik aus Veranstaltungen mit verschiedenen Themen betont. Zwar trifft es durchaus zu, dass zwischen *einzelnen* Studierenden nur Urteilerübereinstimmungen in geringer Höhe beobachtbar sind und damit auch nicht die Urteilerkonvergenz überschritten wird, die bei Begutachtungen von Manuskripten im Revierprozess erreicht wird. Allerdings liegen bei studentischen Veranstaltungen i.d.R. Angaben von 10 bis 20 und mehr Studierenden vor. In diesen Stichprobengrößen sind messgenaue Werte erzielbar. Studentische Rohwerte sind weniger messgenau als Veranstaltungsmittel.

Ergebnisse studentischer Lehrevaluationen können das Lehrverhalten des einzelnen Dozenten oder aber auch die spezifische Teilnehmerzusammensetzung oder die Themeninteressantheit reflektieren. Diese Faktoren lassen sich analysieren, wenn Beurteilungen aus mehreren Veranstaltungen miteinander verglichen werden: Vergleiche von Evaluationsergebnissen, die aus verschiedenen Veranstaltungen eines Dozenten stammen, mit Resultaten aus Veranstaltungen gleichen Themas, die verschiedene Dozenten hielten, ermöglichen einen Rückschluss auf die personale oder situative Stabilität von Evaluationen.

Hierbei ließ sich zeigen, dass Dozentenvariablen (z.B. Lehrkompetenz) eher dozentenbezogen, studentische Skalen (z.B. Beteiligung) jedoch mehr veranstaltungsspezifisch beurteilt werden. Jede Dimension misst aber in unterschiedlichen Anteilen sowohl Lehrverhalten des Dozenten als auch Veranstaltungsaspekte. Einschätzungen der Dozentenvariablen sind somit nur weitgehend und nicht völlig auf das Lehrverhalten zurückzuführen. Aus *einer* Veranstaltungsevaluation kann deshalb kein zuverlässiges Urteil über die Lehrqualifikationen von Dozenten gefällt werden. Dies hat zur Konsequenz, dass immer mehrere verschiedene Veranstaltungen (verschiedene Themen, Veranstaltungsformen und Semester) eines Dozenten evaluiert werden müssen, bevor zuverlässige dozentenbezogene Aussagen getroffen werden können.

Die Gültigkeit studentischer Veranstaltungsbeurteilungen wird von vielen angezweifelt. „One need not to talk with faculty very long to be aware of their concern about possible biases in student ratings“ (Cashin, 1995, S. 6). Zu diesen potentiellen Verzerrungsfaktoren zählen nach Befragungen von Dozenten verschiedener Länder Themenschwierigkeit, Benotungsstrenge, Popularität, Vorinteresse, Arbeitslast, Besuchszahl und Besuchsgrund. Können diese Befürchtungen als gerechtfertigt angesehen werden?

Evaluationsresultate gelten als valide (gültig), wenn sie das Geschehen in einer Lehrveranstaltung adäquat beschreiben. Zur Bestimmung der Validität müssen verschiedene Verfahren herangezogen werden, die unterschiedliche Aspekte der Validität abdecken. Ist man primär daran interessiert, ob Studierende das Geschehen einer Lehrveranstaltung adäquat beschrei-

ben, ist ein Vergleich der Resultate mit den Urteilen geschulter Fremdurteiler die Methode der Wahl. Ist man im engeren Sinne an der Effektivität der Lehre in Form von Lerngewinn interessiert, können auf diesen Lehrerfolg bezogene Prozess- und Produktvariablen (z.B. Beurteilungsskalen Struktur oder Lernerfolg) durch Zusammenhang mit Leistungstestergebnissen validiert werden.

Soll der Einfluss der Rahmenbedingungen oder des Lehrverhaltens der Dozenten bestimmt werden, empfiehlt sich ein Vergleich der Beurteilungsstabilität von (a) Kursen bei verschiedenen Dozenten und gleichen Themen mit (b) Kursen bei gleichen Dozenten und verschiedenen Themen (Generalisierbarkeit). Verfälschungsmöglichkeiten durch Verzerrungsvariablen lassen sich durch Interkorrelationen zwischen Veranstaltungsbeurteilungen und potentiellen Verzerrungsvariablen untersuchen.

In Metaanalysen (s. Rindermann, 2001) korrelieren die Urteile von Fremdeinschätzern und Studierenden in mittlerer Höhe ($r=.54$). Sie liegen damit in dem Bereich, der als mittlere Größe für Validität beschrieben wurde. Studentische Beurteilungen können somit als zumindest „mittelvalide“ bezeichnet werden.

Zusammenhänge zwischen Klausurresultaten und Evaluationsergebnissen (beides Veranstaltungsmittel) können bei Parallelveranstaltungen (verschiedene Dozenten, gleicher Lehrinhalt, gleicher Test, dozentenunabhängige Benotung, Beurteilung der Lehre vor der Prüfung, Gruppenmittel) als Nachweis der Gültigkeit studentischer Beurteilungen im Sinne von Förderung prüfbareren Wissens gewertet werden. Der Zusammenhang zwischen Evaluationsergebnissen und Leistungsmaßen (im Schnitt $r=.52$) ist überraschend hoch.

Merkmale, die nur mit der Fragebogenbeantwortung und nicht mit den tatsächlichen Geschehnissen einer Veranstaltung zusammenhängen, werden als Biasvariablen bezeichnet. Sie verzerren die Evaluationen; der Nachweis fehlender Biaseffekte spricht dagegen für die Validität von Evaluationen. Das Auftreten von Zusammenhängen kann jedoch keine ursächlichen Beeinflussungen nachweisen: Z.B. gibt es einen Zusammenhang in der Größe von $r=.57$ zwischen Popularität und Dozentenvariablen. Ob aber der Bias „Sympathie für den Dozenten“ die Fragebogenbeantwortung beeinflusst oder ob gutes Lehrverhalten die Popularität steigert oder gar warme Ausstrahlung und Freundlichkeit (die u.a. zu Sympathie führen) auch die Lehreffektivität erhöhen, bleibt in dieser Untersuchungsform ungeklärt.

Generell können über alle Studien hinweg studentische wie dozentische biographische Variablen (Alter, Geschlecht, Semesterzahl) als relevante Biasvariablen ausgeschlossen werden. Ebenso übt das Leistungsniveau der Studierenden, erhoben über Abiturnoten, veranstaltungsbezogene Klausurresultate oder Universitätsprüfungen nur einen zu vernachlässigenden Einfluss auf Veranstaltungsbeurteilungen aus. Die Notenerwartung kann nicht eindeutig als Biasvariable definiert werden, da günstige Noten auch aufgrund einer guten Lehrveranstaltung erwartet werden können.

Entgegen der Erwartung korrelieren in amerikanischen Stichproben Arbeitsaufwand und Schwierigkeit *positiv* mit der Veranstaltungsbeurteilung, bei eigenen Studien (Rindermann,

2001) sind negative und positive Zusammenhänge mit der Wahrnehmung der Anforderungen zu beobachten (s. Abbildung 2).

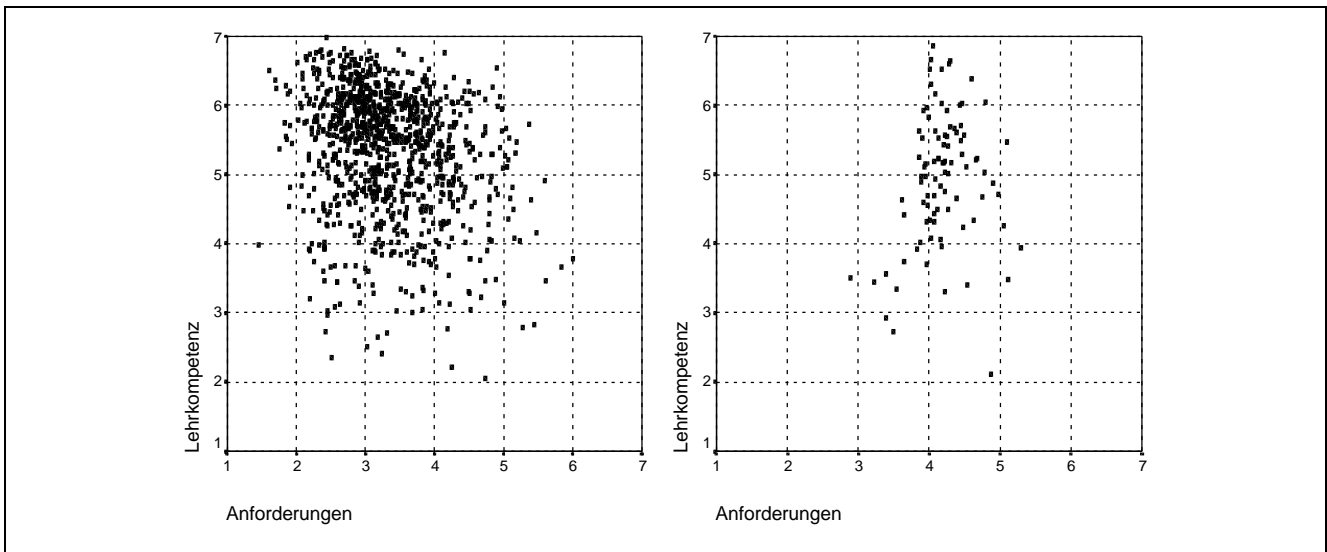


Abbildung 2: Streudiagramm in Veranstaltungsmitteln zwischen Lehrkompetenz und Anforderungen für HILVE-I (links, 1.000 Kurse, $r = -.24$) und HILVE-II (rechts, 92 Kurse, $r = .18$); HILVE ist das Heidelberger Inventar zur Lehrveranstaltungsevaluation in verschiedenen Entwicklungsstufen; eine numerisch höhere Ausprägung (z.B. 7 statt 1) entspricht höheren Werten

Bedeutendere und stabile Zusammenhänge können mit der Beurteilung des Veranstaltungsthemas, mit dem (z.T. retrospektiv eingeschätzten) Vorinteresse und dem Besuchsgrund beobachtet werden. Studierende, die Interesse für ein Veranstaltungsthema empfinden, Vorinteresse angeben und aus Interesse (vs. Pflicht) eine Veranstaltung besuchen, beurteilen diese, den Dozenten und den Lehrerfolg günstiger als Studierende, die kein Interesse entwickeln können und aus extrinsischen Gründen die Veranstaltung belegen. Müssen studentische Beurteilungen deshalb als verzerrt und invalide eingeschätzt werden? Eine solche Interpretation muss aus verschiedenen Gründen angezweifelt werden:

Die beobachtbaren Zusammenhänge bleiben bis auf die Korrelation zur Themainteresantheit und Relevanz im Schnitt auf niedrigem Niveau.

Veranstaltungsthema, aber auch Anforderungen und Popularität des Dozenten und *retrospektiv* eingeschätztes Vorinteresse und (retrospektiv) eingeschätzter Besuchsgrund können nicht eindeutig als Kausalfaktoren interpretiert werden, da das Lehrverhalten des Dozenten deren Beurteilung beeinflussen kann.

Gute Lehre kann für Themen interessieren – interessant empfundene Themen erleichtern Lehre und Lernerfolg und/oder führen zu positiv verzerrten Beurteilungen – die Stärke des jeweiligen Bedingungs-Wirkungs-Verhältnisses ist nicht abschließend geklärt. Allerdings kann eine gewisse Verzerrung empirisch auch nicht ausgeschlossen werden. Deshalb können – sollten Lehrevaluationen für Berufungsentscheidungen herangezogen werden – nicht aus einzelnen Evaluationsresultaten Schlussfolgerungen gezogen werden. Hierzu müssen zuvor mehrere verschiedene Veranstaltungen aus unterschiedlichen Bereichen des zu vertretenden Faches evaluiert werden.

Besuchsgrund, Vorinteresse, Themainteresse und Anforderungen sollten als mögliche Kontrollvariablen in Lehrinventaren miterhoben werden. Der Einfluss von potentiellen Verzerrungsvariablen auf die Beurteilung von Lehrveranstaltungen ist aber weit geringer, als allgemein befürchtet wird.

Generell ist zu empfehlen, Evaluationen in mehreren verschiedenen Veranstaltungen (Mindeststichprobengröße jeweils N=10-15 Studierende aus fünf Veranstaltungen) vorzunehmen, falls es Ziel ist, die Lehrqualität eines Dozenten zu bestimmen. Der Einfluss eventuell auftretender Störvariablen kann so minimiert werden. Nicht sinnvoll sind Vergleiche zwischen Dozenten verschiedener Fächer (beispielsweise Chemie versus Geschichte).

4. Auswirkungen und Nutzen studentischer Lehrevaluationen

Studentische Veranstaltungskritik sollte einen Beitrag zur Optimierung der Lehre leisten, Evaluation ist nie Selbstzweck. Kann Lehre durch Evaluationsmaßnahmen nachweisbar verändert werden?

Alle bisherigen nationalen und internationalen Studien zeigen, dass bloße Durchführung von Lehrevaluation, oder nur Lehrevaluation und Feedback, oder nur Lehrevaluation, Feedback und veranstaltungsinterne Besprechung zwischen Studierenden und Lehrenden nicht ausreichend sind zur erkennbaren Verbesserung der Lehre. Vielmehr sind zusätzlich externe Beratung und Training notwendig.

Diese Ergebnisse verweisen auf die Notwendigkeit, Feedback studentischer Lehrevaluationen mit mündlicher Erläuterung und persönlicher Beratung durch Kollegen oder Fachdidaktiker zu ergänzen: Nur-Feedback-Ansätze greifen zu kurz, um stabile handlungssteuernde Strukturen der Dozenten modifizieren zu können. Weiterbildung und Beratung sind unabdingbar, um substantielle und zeitstabile Veränderungen zu bewirken. Evaluation dient hierbei der Bestimmung von Defiziten (wo besteht Handlungsbedarf?) und sie dient als Informationsgrundlage für Beratung und Training. Zusätzlich prüft sie den Erfolg von Rückmeldung, Beratung und Training.

5. Empfehlungen zur Durchführung von Lehrevaluation

5.1 Rückmeldung für die Lehrenden

Die Beurteilung der Veranstaltungsqualität sollte durch eine ausgewogene Mischung vorgegebener standardisierter Items, veranstaltungsspezifischer (und/oder selbst formulierter) Fragen und offener Fragen, auf die schriftlich in eigener Formulierung geantwortet werden kann, vorgenommen werden. Ein Fragebogen mit standardisierten Items (Fragen) liefert vergleichbare Daten aus vielen inhaltlichen Bereichen und ist ein statistisch auf Gütekriterien überprüfbares Instrument. Zudem können diese Instrumente maschinenlesbar gestaltet werden. Ein durch einen standardisierten Fragebogen gestütztes Urteilsverfahren ist der freien Urteilsbildung hinsichtlich inhaltlicher Repräsentativität, Objektivität, Messgenauigkeit und Validität deutlich ü-

berlegen. Die Angabe von Normen erleichtert durch ein Bezugssystem die Interpretation der eigenen Resultate. Zusätzlich müssen aber veranstaltungsspezifische Items herangezogen werden.

Eine solche Befragung sollte Mitte des Semesters vorgenommen werden, damit die Resultate noch Auswirkungen für die laufende Veranstaltung zeigen können. Studierende und Lehrende dürfen zeitlich mit Evaluationsaufgaben aber nicht überfordert werden. So sollten Evaluationen aller Veranstaltungen nicht häufiger als alle drei bis fünf Semester durchgeführt werden. Aufwendige Verfahren (Diskussion, moderierte Rückmeldung, Beratung, Trainings etc.) sollten nur bei kritisch bewerteten Lehrkräften oder jungen, auf eigene Lehrerfahrung noch nicht zurückblickende Dozenten zur Verbesserung der Lehrqualität eingesetzt werden. Evaluation dient der Verbesserung der Lehre, keinesfalls ihrer Ersetzung!

5.2 Lehrevaluation im Beratungs- und Weiterbildungsprozess

Bei einem formativen Ansatz – d.h. wenn das Ziel der Lehrevaluation die Verbesserung der Qualität der Lehre darstellt – sollte immer eine Begleitung durch hochschuldidaktische Maßnahmen erfolgen.

Beispiel: Beratung in Berkeley

- Durchführung von studentischen Lehrevaluationen zur Verwendung als Informationsgrundlage
- Übermittlung des Feedbacks durch hochschuldidaktische Berater
- Diskussion des Feedbacks zwischen Dozent und Berater; Erörterung verhaltensnaher, spezifischer, konkreter und umsetzbarer Vorschläge für die Lehrveranstaltung
- Heranziehung von Lehrtechniken sehr gut evaluierter Dozenten als eine Informationsquelle für Beratungsgespräche
- Aufrechterhaltung des Kontakts und „Verpflichtung“ durch brieflichen und telephonischen Nachkontakt
- erneute Lehrevaluation zur Bestimmung von Veränderungen

5.3 Feedback für Fachbereich oder Fakultät

Lehrevaluationsresultate informieren bei entsprechender Fragensammensetzung nicht nur einzelne Dozenten über didaktische Desiderata, sondern lassen auch Rückschlüsse auf curriculare Defizite und problematische Rahmenbedingungen zu. Sind beispielsweise Veranstaltungsinhalte aufeinander abgestimmt, sind Dozenten außerhalb von Veranstaltungen zugänglich? Lehrevaluationsergebnisse in allgemeiner und veranstaltungsbezogener Form können in Podiumsdiskussionen zwischen Studierenden und Lehrenden besprochen werden. Vorteil einer solchen moderierten Besprechung ist, dass institutsspezifische Probleme der Lehr- und Studiensituation und deren Lösung erörtert werden können: z.B. Organisation von aufeinander auf-

bauenden Veranstaltungen, Prüfungsordnung, Bibliothekszeiten, kritische Bereiche der Lehre usw. Der Schwerpunkt sollte auf die Entwicklung von Lösungsvorschlägen gelegt werden.

5.4 Informationen für die Veranstaltungswahl der Studierenden

Studierende können Evaluationsresultate und ergänzende Informationen zur Veranstaltung für ihre eigene Veranstaltungswahl – soweit ein Spielraum z.B. in Form von Alternativveranstaltungen besteht – nutzen. Voraussetzung dafür ist die Veröffentlichung veranstaltungsbezogener Resultate.

5.5 Lehrevaluationen bei der Vergabe von Lehrpreisen sowie bei Berufungs- und Bleibeverhandlungen

Lehrpreise stellen eine Möglichkeit dar, die Bedeutung der Lehre an den Hochschulen zu stärken. Sie können die Aufmerksamkeit auf gute Lehre richten und diese in Vorbildfunktion öffentlich bekannt machen sowie eine Diskussion über Defizite in der Lehre in Gang setzen. Bei der Setzung von Anreizen ist es wichtig, Breitenwirkung zu erzielen. Lehrpreise erzielen zwar Öffentlichkeitswirksamkeit; außerdem tragen sie dazu bei, ein positives Umfeld für Anstrengungen in der Lehre zu schaffen. Jedoch erreicht man damit nicht den Problembereich, nämlich die ca. 10% der Veranstaltungen und Dozenten, bei denen besondere Defizite durch Evaluationen aufgedeckt werden. Hier sollten Beratungsmöglichkeiten zur Eröffnung von Handlungsalternativen angeboten werden.

Für Hochschullehrer können bewerbungsverwertbare Unterlagen aus Lehrevaluationen einen wichtigen Anreiz darstellen, falls bei Berufungen Leistungen in der Lehre gebührend berücksichtigt werden.

5.6 Lehrevaluationsabhängige Mitteldistribution

Die leistungsabhängige Mittelverteilung (Sach- und Personalmittel, Forschungsfreisemester, an der Leistung orientierte Besoldungsstruktur) könnte sich bei entsprechender institutioneller Flexibilität und Veränderungsmöglichkeit als breitenwirksamer Motor zur Verbesserung der Lehre herausstellen. Die Auswahl der Kriterien muss jedoch sorgfältig geschehen und sollte erst nach Konsultation betroffener Fachbereiche und fachübergreifender Wissenschaftsverbände erfolgen. Die unterschiedliche Beeinflussung dieser Kriterien durch (un)günstige Rahmenbedingungen muss geklärt werden. Denkbar sind als Kriterien auf personaler und institutioneller Ebene Lehrevaluationsergebnisse (aus mehreren Veranstaltungen je Dozent), Absolventenquoten, Anzahl betreuter Diplomarbeiten und Dissertationen, besondere Lehrprojekte, Fachstudienzeiten.

6. Anforderungen an Praxis und Rahmenbedingungen von Lehrevaluationen

6.1 Weiterbildungsangebote

Lehrveranstaltungsbeurteilungen mit Feedback aber ohne Weiterbildung und ohne Beratung konnten in allen bisher veröffentlichten Studien nur geringfügige Veränderungseffekte zeigen, wohingegen bei Ergänzung durch Beratung und Weiterbildung positive Veränderungen in studentischen Beurteilungen, in studentischen Einstellungen und sogar in studentischen Leistungen erkennbar waren. In den USA bieten 77% aller Universitäten solche Kurse an, in Harvard nehmen ca. 50% aller Lehrenden jährlich an solchen Weiterbildungen teil. Dozenten werden diese Angebote aber nur dann aufgreifen, wenn sich Engagement in der Lehre lohnt, welche in Zeitkonkurrenz zu Forschung, Verwaltung, Nebentätigkeiten und Freizeit steht.

6.2 Hochschulpolitik

Die Qualität der Lehre und der Ausbildungssituation (Studiendauer, Studienabbruchquote etc.) an Hochschulen kann durch bloße Messung und Qualitätsfeststellung nicht verändert werden. Falls die Qualität der Lehre und des Studiums optimiert werden sollen – eine hochschulpolitische Entscheidung – müssen solche Systembedingungen geschaffen werden, die Qualitätsverbesserungen ermöglichen und fördern.

Lehr- und Lehrveranstaltungsevaluationen müssen in ein *allgemeines Verfahren zur Bestimmung und Förderung* von Lehr-, Ausbildungs- und Forschungsqualität integriert werden. Die alleinige Einschätzung von Lehrqualität ist nicht ausreichend für eine Verbesserung. Insbesondere ist ein Ansatz abzulehnen, Evaluation – die bloße Messung der Lehrqualität – von oben zu verordnen, ohne Personen und Institutionen zu ermöglichen, Qualität zu verbessern. Die umfangreiche nationale und internationale Hochschulforschung zeigt, dass Evaluation in einem solchen Fall keine positiven Veränderungseffekte zeigt, im schlimmeren Fall führt sie zu Frustration und Entmutigung bei kritisch bewerteten Dozenten und Fachbereichen, manche Hochschulforscher sprechen sogar von „*professorial melancholia*“.

Der Hochschulforscher Doyle hält einen solchen Ansatz (nur Messung ohne Qualitätsmanagement) auch für inhuman und unökonomisch: „I do not think it is humane to open someone to the possibility of a negative evaluation without at the same time providing some meaningful help toward improvement. Business firms know that they waste money if they discourage or dismiss potentially productive employees and so they spend large sums of money on intensive training programs.“ (Doyle, 1991, S. 126). Ähnlich Marsh & Roche (1999, S. 517): „There is an ethically dubious but widespread custom of giving potentially negative feedback to teachers without providing access to cost-effective interventions to assist them to improve their teaching effectiveness. This, perhaps, is the most serious indictment of the current practice.“

Wenn man an der Verbesserung der Qualität der Lehre in Form von Lehrveranstaltungen interessiert ist, sollte man die Lehre im Rahmen eines *Beratungs-Rückmelde-Verfahrens* eva-

luieren. Dagegen ist eine Evaluation ohne Möglichkeit für die Beteiligten, die Qualität zu verbessern, ohne Beratung und fördernde Rahmenbedingungen weder effektiv noch vertretbar.

Ist man an der Verbesserung der lehrveranstaltungsübergreifenden Ausbildungsqualität interessiert, sind interne Selbst- und externe Fremdevaluationsprozesse durch Kollegen (Peers) in *effektive Kommunikations- und Entscheidungsprozesse* zu überführen, an deren Ende Anreize für gute Lehre stehen müssen. Moderne Management- und Organisationskonzepte sowie Ergebnisse der Hochschul- und Lehrevaluationsforschung liegen hierfür bereit. Es macht Sinn etwas, z.B. eine Leistung oder Lehre und Forschung an einer Universität zu evaluieren, wenn man, falls notwendig, diese nach der Bestimmung verbessern oder zu diesen eine Entscheidung treffen könnte. Ist dies nicht möglich, wird vornehmlich Mehrarbeit gefördert, nicht Qualität.

Universitäre Lehre und Forschung entstehen nicht in einem Vakuum. Ausbildungserfolg – Gegenstand und Ziel von Lehrevaluationen und Lehrberichten – ist *eingebettet in ein Umfeld*, beispielsweise ist er angewiesen auf die Vorbildung der Studienanfänger seitens der Schulen, auf Beschlüsse der Kultusministerkonferenzen, die Schulen und Universitäten Rahmenbedingungen vorgeben, auf Zulassungsverfahren und Studienfinanzierungsmöglichkeiten, Verwaltungseffizienz, das Besoldungssystem, rechtlich vorgegebene Organisationsstrukturen innerhalb der Universitäten usw. Die Frage nach der Qualität bzw. nach ihrem Ausbleiben ist deshalb weniger eine Frage, die man Individuen stellen darf, sondern eher eine Frage, die durch Analyse von Struktur- und Rahmenbedingungen der Universität zu beantworten ist. Auch vermeintlich personenbezogene Merkmale wie didaktische Kompetenz oder Motivation von Lehrenden, bei denen zwar wie überall interindividuelle Unterschiede bestehen, weisen einen hohen strukturellen Anteil auf, da diese durch didaktische Beratung und Training (wie andernorts üblich und institutionell verankert) und Anreize sowie personale Entscheidungsmöglichkeiten beeinflusst werden.

Eine Institution zu evaluieren macht nur dann Sinn, wenn Möglichkeiten bestehen, auf Ergebnisse zu reagieren und Veränderungen einzuleiten. Hochschulautonomie – *Eigenverantwortlichkeit und Steuerungsfähigkeit* – ist Voraussetzung für Evaluation und Qualität. Das Mittel *qualitätsorientierter Evaluation* bleibt sonst systemfremd und bewirkt nur wenig mehr als Verwaltungsarbeit für alle Beteiligten. Qualität könnte sogar vermindert werden, weil Zeit für Forschung und Lehre verloren geht. Im jetzigen Zustand dienen Evaluation und Berichterstellung primär der Befolgung extern festgelegter Vorgaben. Innerhalb des gegebenen Rahmens ist der Spielraum für qualitätsorientierte Evaluation nicht groß. Qualität bleibt so von *individuellem Engagement* abhängig. Beispielsweise kann *Lehrveranstaltungsevaluation* eingesetzt werden, um Rückmeldungen zur Qualität eigener Lehre zu erhalten. Lehrevaluation und Lehrberichte können in einen *Kommunikationsprozess* zwischen Lehrenden und zwischen Lehrenden und Studierenden münden, in dem Probleme erörtert und Lösungen entwickelt werden. Die Leistungsfähigkeit der Lehrveranstaltungsevaluation wäre bei Beratung, Trainingsangeboten und Anreizen aber größer.

7. Literatur

Cashin, W. E. (1995). *Student ratings of teaching: the research revisited*. Manhattan: Center for Faculty Evaluation and Development (IDEA No. 32).

Doyle, K. O. (1991). Report on a trip downtown. In M. Theall & J. Franklin (Hg.), *Effective practices for improving teaching (New directions for teaching and learning 48)* (S. 123-126). San Francisco: Jossey-Bass.

Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52(11), 1187-1197.

Marsh, H. W. & Roche, L. A. (1999). Reply upon SET research. *American Psychologist*, 54(7), 517-518.

Rindermann, H. (2001). *Lehrevaluation – Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Mit einem Beitrag zur Evaluation computer-basierter Unterrichts*. Landau: Empirische Pädagogik.

8. Lebenslauf

Heiner Rindermann (Jahrgang 1966) studierte Psychologie in Heidelberg und schloss dieses Studium 1995 dort mit Promotion ab. 1994-1999 arbeitete er an der Universität München im Institut für Pädagogische Psychologie in einem Forschungsprojekt zum Thema Hochbegabung; seit 1999 lehrt und forscht er als Wissenschaftlicher Assistent im Bereich Methodenlehre, Diagnostik und Evaluation am Psychologischen Institut der Universität Magdeburg. Er beschäftigt sich mit den Themen Lehrevaluation und Unterrichtsforschung, Intelligenz und Hochbegabung, Selbstkonzept und Persönlichkeit, evolutionäre Psychologie und Wissenschaftsforschung.

Anschrift:

Dr. Heiner Rindermann

Institut für Psychologie, Otto-von-Guericke-Universität Magdeburg

Postfach 4120, D-39016 Magdeburg

TI. 0391/6711919, Fax. 0391/6711965, Sekr. 0391/6718464

E-Mail: heiner.rindermann@gse-w.uni-magdeburg.de

<http://www.uni-magdeburg.de/ipsy/methoden/hr/webseite-rindermann.htm>