

Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation

Heiner Rindermann

Otto-von-Guericke-Universität Magdeburg

1. Evaluation und Lehrevaluationsforschung an deutschen Hochschulen

Anfang der 90er Jahre begannen Universitäten im deutschen Sprachraum als Reaktion auf verschiedene Krisensymptome in der Hochschulausbildung und im Gefolge studentischer Proteste und Verwaltungsreformen der Länder die Qualität von Lehre und Ausbildung zu evaluieren. Als Verfahren etablierten sich neben *interner Evaluation* („Selbstevaluation“, „Lehrberichte“) und *externer Evaluation* durch Experten („niederländisches Modell“) vor allem *studentische Lehrveranstaltungsevaluationen*. Evaluationsmaßnahmen sind oft in Akkreditierungsverfahren eingebunden, die der Sicherung von Mindeststandards in der Ausbildungsqualität dienen und über Zielvereinbarungen diese Ausbildungsqualität zu verbessern suchen. Heute ist in den meisten Ländern Lehrevaluation gesetzlich verankert und an Hochschulen institutionalisiert. Obwohl die Ergebnisse im Schnitt positiv ausfallen (Daniel 1996), bleibt Lehrevaluation umstritten, zudem trägt ihre Durchführung häufig nicht zu einer Verbesserung der Lehre bei.

Unter *Evaluation* wird die systematische Analyse und empirische Untersuchung von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zum Zwecke ihrer Bewertung und Modifikation verstanden. Über adäquate *Beschreibung* und *Messung* hinausgehend *bewerten* Evaluationen und *optimieren*. Formative Ansätze legen hierbei den Schwerpunkt auf die Verbesserung von laufenden Verfahren, summative Ansätze nehmen über die Bereitstellung fundierter Informationen Wirkungsbeurteilungen vor und unterstützen so Entscheidungen. In Evaluation und Evaluationsforschung lassen sich verschiedene Modelltypen voneinander abheben (s. Rindermann 2001; Beywl 1988). Allen Varianten ist aber gemeinsam, dass über einen Gegenstandsbereich der Praxis durch methodisches Vorgehen relevante Informationen für die Gestaltung dieser Praxis ermittelt werden sollen. Wissenschaftliche Methodik *und* Praxisorientierung kennzeichnen die Evaluationsforschung. Evaluation versteht sich als eine Form *rationaler Problemlösestrategie*, die für Entscheidungen zuverlässige empirische Datengrundlagen liefert

sowie Entscheidungen mittels Klärung von Zielen und Dekomposition von komplexen Problemen rationaler gestalten möchte.

Die Evaluationsmaßnahmen *wissenschaftlicher* Institutionen zeichnen sich erstaunlicherweise oft gerade durch fehlende Rezeption wissenschaftlicher Forschung aus; die Evaluation an Hochschulen geht praktizistisch vor, gesetzliche Auflagen werden verwaltungskorrekt erfüllt oder – vor allem zu Beginn der 90er Jahre – studentische Fachschaften engagieren sich. Hierbei werden aber etwa diejenigen, in vielen nationalen und internationalen Studien gewonnenen Ergebnisse kaum rezipiert und im Handeln berücksichtigt, die nachweisen, dass Lehr(-veranstaltungs)evaluationen ohne beratende Rückmeldung oder Trainingsangebote und ohne Einbettung in ein gute Lehre förderndes und honorierendes Umfeld nicht oder nur wenig die Qualität von Lehre verbessern können. Oder Erhebungsverfahren für Veranstaltungen werden ohne Anknüpfung an die Unterrichtsforschung erstellt, zentrale Determinanten, Merkmale und Erfolgskriterien guter Lehre werden so nicht erhoben. Zudem scheinen klassische Gütekriterien wie Objektivität, Reliabilität und Validität und übliche Verfahren zur Prüfung von Dimensionalität, Kausalität etc. nicht zur Anwendung zu kommen.

Der Eindruck drängt sich auf, dass Lehrevaluation für die Praxis sich in nicht wenigen Fällen schon in der Lösung von Fragebogeneinsammel-, Belegleserauswertungs- und automatisierten Rückmeldeproblemen erschöpft. Funktionieren diese technischen Routinen zur organisatorischen Bewältigung rechtlicher Auflagen, stellen sich weitere Fragen nicht mehr. Die Befolgung von Verordnungen steht vor der Qualitätsorientierung.

Defizite treten aber nicht nur in der Praxis, sondern auch in der Forschung auf, die Forschung bleibt meist an eigenen Fächertraditionen statt an spezifischen Fragestellungen orientiert. Forschungsansätze, Theorien und Methoden anderer Disziplinen werden übersehen: Die in der Psychologie verbreiteten Konstruktionsprinzipien, Versuchsplanungskonzepte und statistischen Auswertungsverfahren; die in der pädagogischen Forschung entwickelten Konzepte zur „guten Lehre“ in Schule und Hochschule; hochschuldidaktische Ansätze der Beratung und des Trainings; soziologische Studien zu Habitus und Feldbedingungen an Universitäten; systemische, hochschulwissenschaftliche und betriebswissenschaftliche Analysen der Institution Universität und ihrer Verwaltung; philosophisch begründete sowie wissenschaftshistorisch analysierte Bestimmungen von Ideen und Erscheinungsformen der Universität. Lehrevaluationsforschung greift zu selten theoretisch wie praktisch relevante Fragestellungen auf.

Forschung und Anwendung, Theorie und Praxis der Lehrevaluation sind wenig und in einer für beide Seiten zu unproduktiven Weise miteinander verknüpft. Dieser Beitrag versucht nun ausgehend von einer kurzen Schilderung des Stands der Forschung die kurz skizzierte Problematik zu explizieren und zu klären, warum Evaluation an Hochschulen praktizistisch vorgeht. These ist, dass die Ziele von Evaluation auf diese Weise nicht erreicht werden können. Die wechselseitige Rezeption von Praxis und Theorie, von Lehrevaluationsforschung und Grundlagen- wie Anwendungsforschung verschiedener Disziplinen wird als sinnvolle Alternative zum bisherigen Vorgehen in ihrer theoretischen Bedeutung und in ihrer praktischen Relevanz beschrieben.

2. Stand der Forschung

Der Stand der bisherigen Lehrevaluationsforschung wird an dieser Stelle nur kurz als Basis für die folgenden Erörterungen geschildert. Es werden Erfahrungen internationaler und nationaler Studien zusammenfassend referiert, insbesondere werden hierzu Ergebnisse aus Studien, die mittels des HILVE (Heidelberger Inventar zur Lehrveranstaltungsevaluation in verschiedenen Entwicklungsstufen) an mehreren Hochschulen und Ausbildungseinrichtungen (Heidelberg, München, Bonn, Leipzig, Berlin, Halle, Magdeburg, Nürnberg, Reichenau, Wiesbaden) durchgeführt wurden, herangezogen.¹

2.1 Merkmale guter Lehre

Was zeichnet „gute Lehre“ aus? Lehrevaluation lässt sich nicht durchführen ohne diese Frage zuvor konzeptuell und empirisch zu klären. Zunächst sei hier Lehre von Lehrveranstaltung unterschieden: *Lehre*, *Lehrqualität* oder *Lehrevaluation* sind veranstaltungs- und dozentenübergreifend und beziehen sich auf die Hochschullehre insgesamt inklusive Veranstaltungen, Ausbildungsinhalten, Bibliotheksausstattung, Studienzeiten, Absolventenchancen u.ä., während unter *Lehrveranstaltung*, *Veranstaltungskritik* oder *Lehrveranstaltungsevaluation* ausschließlich die Evaluation von Vorlesungen und Seminaren verstanden wird. Die Lehrveranstaltung stellt den Kern der Hochschullehre dar. In ihr findet der Wissens- und Kompetenzerwerb in Interaktion zwischen Dozenten und Studenten statt. In der Regel werden aber für beide Bereiche die umfassenderen und kürzeren Begriffe „Lehre“ und „Lehrevaluation“ verwendet.²

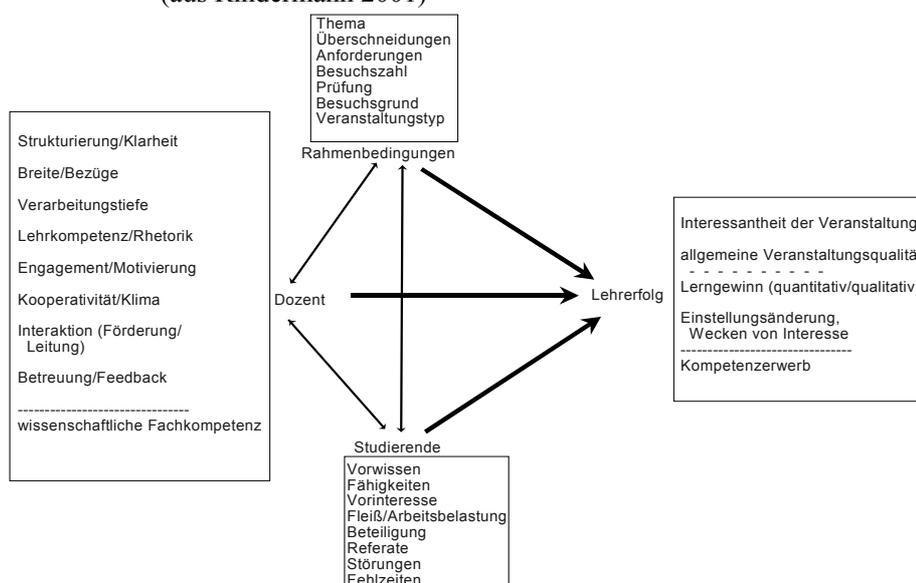
Welche Merkmale zunächst charakterisieren gute Dozenten, welche gute Veranstaltungen? Zur Beantwortung können Lehr- und Ausbildungszielkriterien axiomatisch definiert werden. Ebenso lassen sich instruktionspsychologische Theorien heranziehen oder gegebene Gesetzestexte, Bildungsvorstellungen, Curricula und Lehrevaluationsbögen interpretieren. Der direkteste Weg, Kennzeichen guter Lehre zu finden, besteht darin, Dozenten und Studenten zu befragen.

Zu den für erfolgreiche Lehrveranstaltungen verantwortlichen Lehrmethoden zählen als Ergebnis verschiedener Studien unterschiedlicher Autoren in mehreren Hochschulsystemen (direkte Befragungen, Regressionsanalysen mit Lehrevaluationsdaten, Faktorenanalysen, Strukturgleichungsmodelle, Auswertung von Konzepten guter Lehre, Unterrichtsforschung; Rindermann 2001, 1999) zuallererst Strukturierung und Klarheit. Als Determinanten auf Seiten des *Dozentenverhaltens* gelten darüber hinaus auch verständliche Erklärungen und Darstellungen, rhetorische

- 1 Ausführlichere Informationen zu Forschungsmethoden und Forschungsstand finden sich u.a. in Marsh (1987), Marsh/Roche (1997), Cashin (1995), Perry/Smart (1997), Rindermann (2001), Spiel (2001), Diehl (2001) und Greimel (2002) sowie in den Heften der *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 1996, 16(2), *Zeitschrift für Pädagogische Psychologie*, 1996, 10(3/4), *Zeitschrift für Pädagogik*, 1999, 45(1) und 2000, Beiheft 41, und in *Psychologie in Erziehung und Unterricht*, 2003, 50(1).
- 2 Im englischen Sprachraum analog *teaching effectiveness, evaluation of teaching* usw.

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 *studentische Determinanten* wie Fleiß, Vorwissen, die veranstaltungsinterne Beteiligung und Referate dar (Gold/Souvignier 2001). 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). Hinzu kommt auf Dozentenseite die wissenschaftliche Fachkompetenz, die durch Studierende aber nicht eingeschätzt werden sollte. Die Determinanten sind nicht unabhängig voneinander, der Dozent kann etwa Einfluss auf Thema, Anforderungen und Besuchgrund ausüben, der Veranstaltungstyp auf Beteiligung und Betreuung, das Vorwissen auf Anforderungen, der Fleiß auf die Betreuung usw. Lehrerfolg wird multikriterial aufgefasst, innerhalb des Lehrerfolges lässt sich zwischen Erfolgsmerkmalen für die Veranstaltung und intendierten Veränderungen auf Seiten der Studierenden unterscheiden. Das vier Komponenten unterscheidende Modell – drei Bedingungsfaktoren und der Lehrerfolg – kann neben der Fragebogenkonstruktion auch als präskriptives Lehrmodell für die Beratung von Lehrenden herangezogen werden. Aus diesem Grunde ist das Modell auch differenzierter als die Skalen der meisten Lehrinventare (vgl. Diehl 2001; Rindermann 2001) und differenzierter als die empirisch in der Regel zu ermittelnde Faktorenstruktur: So lassen sich etwa die Skalen Lehrkompetenz und Engagement selten im HILVE statistisch trennen, für die Beratung ist es jedoch wichtig, zwischen Kompetenz und Motivierung zu trennen, denn über letzteres ist es einfacher, einen Einstieg in die Veränderung von Unterrichtshandeln zu finden.

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



Eine differenzierte Beschreibung anhand theoretisch begründeter und empirisch überprüfter Skalen, die das komplexe Geschehen in Lehrveranstaltungen wiedergeben, ist Voraussetzung für Bewertung und sich hieran anknüpfende Maßnahmen.

Die Qualität der Lehre an Hochschulen geht über die der Veranstaltungen hinaus.³ Sie umfasst inhaltliche sowie formale Ziel-, Prozess- und Bedingungsgrößen wie ein inhaltlich adäquates und im zeitlich vorgegebenen Rahmen studierbar gestaltetes Curriculum, Breite und Inhalt des Veranstaltungsangebots und der vertretenen Fachrichtungen, Abstimmung der Inhalte und Zeiten verschiedener Veranstaltungen und Studienabschnitte, geringe Veranstaltungsausfälle, kurze Studiendauer, Bibliotheksausstattung und -öffnungszeiten, technische Ausstattung, Zugänglichkeit von Dozenten, Qualität des Studienanfängerzulassungsverfahrens, Absolventenquote und Studienabbruch, Bearbeitungszeiten und Betreuung von Abschlussarbeiten, Absolventenqualifikation und deren Berufschancen. Die Studienbedingungen an einem Fachbereich sind eingebettet in generelle Rahmenbedingungen von Ausbildung und Hochschule (Vorbildung, Zulassungssystem, Finanzierung, Arbeitsmarkt u. ä.).

Gute Lehre ist nicht nur durch einzelne didaktisch erfolgreiche Lehrveranstaltungen erreichbar, sondern muss in ein entsprechendes Umfeld eines Faches, einer Universität und förderlicher gesellschaftlicher Rahmenbedingungen eingebettet sein. Die Evaluierung von Veranstaltungen und der Lehre im Allgemeinen ist für eine umfassende Beurteilung der Lehre an einem Fachbereich notwendig. Hierzu – und im Besonderen bei inhaltlichen Fragen – sollten nicht nur studentische Einschätzungen, sondern auch Selbst-, Peer- und Expertenbeurteilungen sowie quantitative Indikatoren herangezogen werden, die im Falle der Lehrveranstaltungsqualität studentische Angaben aber keinesfalls ersetzen können. Forschung, die solche Merkmale zu Ausbildungserfolg in Beziehung setzt, ist aber rar.

2.2 Urteilskompetenz der Studierenden, Objektivität, Reliabilität, Generalisierbarkeit und Validität

Von Kritikern studentischer Lehrevaluationen wird oft die Urteilskompetenz der Studierenden in Frage gestellt. Konkret wird die Heterogenität der Urteile oder mangelnde Vergleichbarkeit von Veranstaltungskritik aus Veranstaltungen mit verschiedenen Themen und vor allem der Einfluss durch Drittvariablen wie Thema, Besuchsgrund und Benotungsstrenge betont (Kromrey 1994). Es wird befürchtet, Studierende würden nur nach Beliebtheit des Dozenten oder des Themas eindimensional urteilen.

3 Merkmalslisten, Forderungen, Begründungen, Berichte und empirische Untersuchungen finden sich u.a. bei Altrichter/Schratz/Pechar (1997), Cashin (1989), Hochschulrektorenkonferenz (2002), Olbertz/Pasternack/Kreckel (2001), Rindermann/Oubaid (1999), Rindermann/Wagner (2003), Webler (2002) und Wissenschaftsrat (1996).

Tabelle 1: Messgenauigkeit: Ergebnisse aus Studien mit verschiedenen Instrumenten

Verfahren	Einzelurteile	Kursmittel (10/20)	
Interraterreliabilität	.30 [.24]	.80/.89	
Cronbachs- α	.76 [.81]	.86 [.88]	
Retestreliaibilität	.61 [.63]	.70 [.78]	innerhalb von Kursen
		.69 (Lehrerfolg)	zwischen Kursen
Generalisierbarkeit	–	.47 [.49]	(5 Kurse: .80; 10 Kurse: .90)

Anmerkungen: Werte in eckigen Klammern Ergebnisse nationaler und internationaler Studien im Mittel, Werte ohne eckige Klammern Ergebnisse aus Studien mittels des HILVE-I u. II im Mittel; Kursmittel basieren auf 10 oder 20 Urteilen; Lehrerfolg sind die Einschätzungsskalen Interessenantheit, Lernen, Interessenförderung und Allgemeinurteil; Retestreliaibilität kann innerhalb von Kursen mit studentischen Rohwerten und über verschiedene Kurse hinweg innerhalb eines Semesters berechnet werden; Generalisierbarkeit nur für Dozenten- oder Dozenten- und Lehrerfolgsskalen (s. Rindermann 2001).

Ergebnisse verschiedener Untersuchungsansätze (Interraterreliabilität; Reliabilitätsberechnungen; Faktorenanalysen; Generalisierbarkeit; Korrelation verschiedene Aspekte der Lehre abdeckender Skalen mit potentiellen Verzerrungsfaktoren; Vergleich der Resultate mit den Urteilen geschulter Fremdurteiler; Zusammenhang mit Leistungstestergebnissen) sprechen für die Brauchbarkeit von Lehrevaluationen (s. Tabellen 1 und 2). Die Reliabilitäten für Kursmittel sind zufrieden stellend bis gut. In Faktorenanalysen und Generalisierbarkeitsüberprüfungen (dimensionsbezogene Vergleiche der Urteilsstabilität von Evaluationsergebnissen, die aus verschiedenen Veranstaltungen eines Dozenten stammen, mit Stabilitäten der Beurteilungen für Veranstaltungen gleichen Themas, die verschiedene Dozenten hielten) lassen sich das Handeln des Dozenten (z.B. Lehrkompetenz), Lehrerfolgsscharakteristika (z.B. Allgemeinbeurteilung der Veranstaltung und Lernen), studentische Beteiligung und Rahmenbedingungsvariablen (z.B. Anforderungen) unterscheiden. Voraussetzung hierfür ist ein multidimensionales Instrument.

Geben Studierende in ihren Urteilen tatsächlich das Veranstaltungsgeschehen wieder (Validität)? Zwischen Fremdurteilen bzw. Leistungsmaßen als Kriterium und studentischen Urteilen konnten Zusammenhänge in einer Höhe wie zwischen Intelligenztests und Leistungskriterien bekannt festgestellt werden. Beurteilungen durch Studierende zeichnen somit kein unrealistisches Bild der Lehre und sind sensibel für Lerngewinn. Zudem reflektieren Dozenten-Skalen mehr das Lehrverhalten des Lehrenden als Rahmenbedingungen. Deshalb kann von studentischen Lehrevaluationen als einem Maß universitärer Lehrqualität gesprochen werden.

Tabelle 2: Validität: Ergebnisse aus HILVE-Studien und Studien mit anderen Instrumenten

Verfahren	Validitätsmaß	
Zusammenhang mit Fremdraterurteilen	Übereinstimmungsvalidität/Adäquanz	Fremdurteiler-Studenten, $r=.54$ Dozenten-Studenten, $r=.24$ Fremdurteiler-Dozenten, $r=.06$
Generalisierbarkeit	Dozentenvalidität	.47/.49 bei Dozentenskalen über verschiedene Veranstaltungen
Zusammenhang mit Leistungskriterien	Effektivität der Lehre	$r=.52$ zwischen mittlerem Prüfungserfolg und mittlerer Veranstaltungsbeurteilung in Dozenten- und Lehreffektivitätsvariablen
Zusammenhang mit Biasvariablen (unerwünscht)	Verfälschbarkeit	biographische Variablen ohne Zusammenhang (Alter, Geschlecht, Abitur etc.) Popularität $r=.57$ (Kausalitätsproblem!) Thema $r=.43$ (HILVE-I)/.19 (HILVE-II) (Kausalität?), Vorinteresse $r=.21$ Besuchsgrund $r=.31$ (Kausalität?)

Anmerkungen: Zusammenfassung verschiedener Studien unterschiedlicher Autoren mit verschiedenen Instrumenten in mehreren Hochschulsystemen (s. Rindermann 2001); Zusammenhang mit Leistungskriterien: Kursmittel, Vergleich verschiedener Dozenten – ein Kurstyp – ein Thema – eine Prüfung – eine einheitliche Auswertung (keine einzige deutsche Studie); Generalisierbarkeit steht zwischen Reliabilität und Validität.

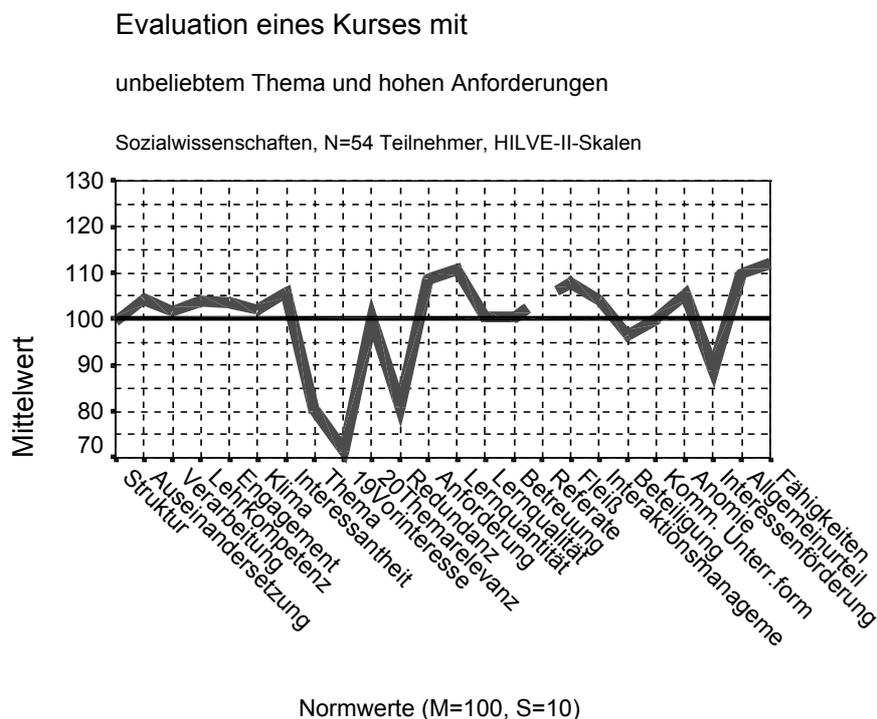
Auch wenn die Frage der Kausalität nicht abschließend geklärt ist⁴, kann ein gewisser Einfluss von Veranstaltungsthema und Besuchsgrund auf die Veranstaltungsbeurteilung und – zwar geringer, aber inhaltlich gravierender – auf die Dozentenbeurteilung nicht ausgeschlossen werden. Generell ist deshalb für die Praxis 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, aber nicht völlig ausgeschlossen werden.

Ein Beispiel der Evaluation einer Veranstaltung mit einem unbeliebten Thema aus einem sozialwissenschaftlichen Studiengang zeigt Abbildung 2. Positive bzw. hohe Werte sind erkennbar in den *dozentenbezogenen Skalen*, z.B. in Lehrkompetenz und Auseinandersetzung. Positive bzw. hohe Werte fallen auch auf in den *Lehreffektivitätsskalen*, z.B. in Interessantheit der Veranstaltungsgestaltung, der Lern-

4 Nach Umformulierung der Themenskala (HILVE-I: „Das Thema der Veranstaltung interessiert mich.“ HILVE-II: „Ich habe mich schon *vor* dem Kurs sehr für die Themen interessiert.“) sinkt der Zusammenhang zwischen Thema und den relevanten Dozenten- und Lehreffektivitätsskalen. Fragt man global nach der Themeninteressantheit im Kurs, lassen sich Vorinteresse und Seminarwirkung nicht trennen: Gute Lehre kann für Themen interessieren oder interessant empfundene Themen erleichtern Lehre und Lernerfolg oder interessant empfundene Themen führen zu positiv verzerrten Beurteilungen. Dringend notwendig sind für diese und andere Fragestellungen (z.B. Effektivität) aktuelle Metaanalysen, die auch unterschiedliche Varianten von Untersuchungsbedingungen und Berechnungen von Koeffizienten beachten.

quantität und im Allgemeinurteil. Negative bzw. geringe Werte sind dagegen in den *Skalen Thema, Redundanz und Interessenförderung* beobachtbar. Zwar sind die Studierenden der Meinung, dass es sich hierbei um ein nicht irrelevantes Thema handelt (Item 20: Das Thema d. Kurses als solches ist relevant [Beruf/Praxis/Prüfung/Gesellschaft].), aber die Studenten geben sehr deutlich kund, dass sie sich vor der Veranstaltung in keinsten Weise für dieses Thema interessiert haben (Item 19: Ich habe mich schon vor dem Kurs sehr für die Themen interessiert.). Thema dient als „Biasskala“ der Erhebung von möglicherweise die Beurteilung des Dozenten und der Veranstaltung beeinflussenden aber von der Lehrleistung des Dozenten weitgehend unabhängigen Rahmenbedingungen. Der niedrige Wert in Redundanz deutet darauf hin, dass es keine thematischen Überschneidungen mit anderen Veranstaltungen gibt und dass eventuell das Vorwissen der Teilnehmer zu gering ist. Hierfür spricht auch der hohe Wert in Anforderungen. Etwas kritisch ist aber, dass es nicht ausreichend gelingt, die Interessen der Studierenden durch den Besuch der Veranstaltung zu fördern (Item 48: Die Veranstaltung fördert mein Interesse am Studium.; Item 49: Der Kurs motiviert dazu, sich selbst mit den Inhalten zu beschäftigen.). Anscheinend sind dies und das Urteil in dieser Skala nicht völlig unabhängig vom Vorinteresse der Studierenden. *Fähigkeiten* basiert auf der Abiturnote und dem Fähigkeitsselbstkonzept der Veranstaltungsteilnehmer.

Abbildung 2: Differenzierte Ergebnisse aus einer Veranstaltung als Beispiel



Fasst man das Ergebnis zusammen so fällt auf, dass trotz völlig fehlenden Vorinteresses und trotz hoher Anforderungen gut durchschnittliche und überdurchschnittliche Dozenten- und Veranstaltungsbewertungen erzielbar sind. Eine solche Differenzierung ist nur mittels eines multidimensionalen Instrumentes möglich. Dozenten dürfen nur auf dozentenbezogenen Skalen bewertet werden, für die Einschätzung des Veranstaltungsgeschehens und für das Feedback sind jedoch Dozenten-, Studenten-, Rahmenbedingungs- und Lehrerfolgsskalen wichtig. Hierfür ist ein multidimensionales Instrument mit der Erhebung unterschiedlicher Determinanten und Merkmale guter Lehre notwendig.

2.3 Auswirkungen und Nutzen studentischer Lehrevaluationen

Studentische Veranstaltungskritik sollte einen Beitrag zur Verbesserung der Lehre leisten, Evaluation ist nicht Selbstzweck. Kann Lehre durch Evaluationsmaßnahmen nachweisbar verändert werden? Stellt Evaluation als solche schon eine Intervention dar (Henninger 2000)? – Die Summe bisheriger Studien zeigt, dass bloße Durchführung von Lehrevaluation nicht ausreichend ist für eine deutliche Verbesserung der Lehre. Alles andere wäre auch unwahrscheinlich und würde Ergebnissen psychologischer Forschung in anderen Bereichen widersprechen: Umweltbedingungen blieben stabil, Routinen sind nur durch Information nicht aufbrechbar, Evaluation stellt zunächst Messung dar, noch keine Intervention!

Tabelle 3: Veränderung durch Lehrveranstaltungsevaluation?

Ansatz	Auswirkung (d)
Sensibilisierung durch Evaluation	0-Effekt
individuelles Feedback an Lehrende	0,2
hochschuldidaktisches Diskursmodell	0-Effekt
Beratung, Training, z.T. Gratifikation	0,64
HILVE-Studie Feedback Heidelberg	$\pm 0,2$
HILVE-Studie Rückmeldung + Beratung, Logopädie	Dozentenskalen 0,65, Lehreffektivität 0,64, studentische Skalen 0,43, Rahmenbedingungen 0,24

Anmerkungen: Überprüfung der Effektivität durch Messwiederholung innerhalb und zwischen Semestern, Lehre – Lehrevaluation (+Treatment) – Lehre – Lehrevaluation; *d* berechnet entweder durch Vergleich mit Kontrollgruppe oder durch Vorher-Nachher-Vergleich in z.T. verschiedenen Berechnungsvarianten (s. Rindermann/Kohler 2003)

Ergebnisse verweisen auf die Notwendigkeit, Feedback studentischer Lehrevaluationen mit mündlicher Erläuterung und persönlicher Beratung zu ergänzen. Dann sind positive wie substantielle Veränderungen im Messwiederholungsdesign oder verbesserte Einschätzungen im Vergleich zu Kontrollgruppen erkennbar. Diese treten vor allem in solchen Bereichen auf, die durch Handeln des Dozenten direkt veränderbar sind (für HILVE: Lehrkompetenz: $d=0,94$; Dozentenengagement: $d=0,78$) und in Skalen, die den Erfolg der Lehre aus Sicht der Studierenden wiedergeben (für HILVE: Allgemeinurteil: $d=0,85$; Rindermann/Kohler 2003). Evaluation dient in solchen Ansätzen der Bestimmung von Defiziten (wo besteht Handlungsbedarf?), sie dient als Handlungsanstoß und der Motivationsänderung (Balk 2000) und sie dient als Informationsgrundlage für Beratung und Training. Hierfür sind

Normwerte unabdingbar. Zusätzlich prüft sie den Erfolg von Rückmeldung, Beratung und Training. Ohne Anreize für gute Lehre bleibt Engagement für die Lehre aber wahrscheinlich auf individuelle Voraussetzungen (Interesse u.ä.) beschränkt (Leitner 1998; Spiel/Fischer 1998).

Evaluationsprojekte, die Lehre und Studium ganzer Fachbereiche untersuchen, haben bislang nicht in einer gängigen wissenschaftlichen Kriterien genügenden Form Veränderungserfolge belegen können (vgl. Übersicht 1). Fachbereichsbezogen müsste sich durch die Evaluation die Lehre verbessern, als Konsequenz sollte sich auch die Qualifikation der Absolventen unter sonst gleichen Bedingungen (*ce-teris paribus*) erhöhen. Das Gleiche trifft für die Hochschulebene zu. Und ein Hochschulsystem, das Evaluation und Akkreditierung einsetzt, sollte bessere Ausbildungserfolge erzielen als eines, das diese Instrumente der Qualitätssicherung nicht nutzt. Ob dem so ist, weiß man aufgrund fehlender Wirkungsmessungen oder Kosten-Nutzen-Analysen nicht. Generell ist eine wissenschaftsbezogene Fundierung und Überprüfung bei dieser Praxis der Evaluation seltener (Westermann/Rockenbauch 1999). „Irgendwelche wissenschaftlichen Methoden zur Erhebung ... oder Auswertung der herangezogenen Daten und Statistiken werden weder gefordert noch gefördert.“ (Westermann 2002: 20) Ähnlich kritisch äußerte sich Weinert (2001: 6): „Stark gefördert von einigen Bundesländern und nachhaltig unterstützt durch Initiativen der Hochschulrektorenkonferenz entsteht ... eine Evaluationspraxis, die relativ unabhängig vom jeweiligen Stand der theoretischen Reflexion ist.“

Übersicht 1: Wissenschaftliche Überprüfung und Ergebnissicherung bei Evaluation auf veranstaltungsübergreifender Ebene? Beispiele

Ein Beispiel ist Teichmanns (1999) Arbeit zur „Wirksamkeit der Evaluation von Studium und Lehre“: Außer einer Beschreibung des Vorgehens der Evaluation im Verbund norddeutscher Universitäten findet sich im Text die Behauptung, dass „nachweisliche Verbesserungen ... festzumachen sind“ (S. 247) – eine deskriptive Aussage, die empirisch aber nirgendwo belegt wird. Die Autorin fordert für zukünftige Evaluationen, dass „systematische Ergebnissicherungen eingeleitet werden ...“, es ist noch nachrücklicher zu fragen, welche Erfolge sich eingestellt haben“ (S. 247) – eine Forderung, der unbedingt auch entsprechende Untersuchungen folgen sollten.

Ähnlich Hackl und Sedlacek, die in Wien Evaluationen durchführen (2001: 119): „Die Frage der Effizienzmessung von Evaluierungsverfahren ist aber auch in der uns bekannten Literatur nicht befriedigend beantwortet.“ – ohne selbst diese Frage durch systematische Untersuchungen zu beantworten zu versuchen. Oder S. 128: „Wichtig für die Zukunft der Evaluierung der Lehre wird es sein, Messinstrumente zur Bewertung der Wirksamkeit und Effizienz der eingesetzten Evaluierungsverfahren zu entwickeln.“

Meyer und Heesen (2001) entwickelten ein Verfahren zum betriebswirtschaftlichen Hochschul-Controlling. Aufwendig werden verschiedene Kosten und Leistungen verrechnet, die Leistung einzelner Einheiten wird bestimmt. Controlling soll „mittels seiner Informations-, Planungs-, Steuerungs- und Kontrollfunktionen im Rahmen der laufenden Haupt- und Serviceprozesse sowie im Bereich der Ressourcen der Hochschule Effizienz und Effektivität steigern.“ (S. 18) Legitimerweise wird das Mehr an Information dem Mehr an Arbeitsaufwand gegenüber gestellt. Völlig im Kontrast dazu bleibt jedoch die von den Autoren beschriebene Lehrevaluation und vor allem ihre mögliche Wirkung diffus, werden Evaluation und die weiteren Maßnahmen (kommunikative Prozesse, Zielvereinbarungen) nicht einer Wirkungskontrolle unterzogen: Ob das „vorrangige Ziel der selbstgesteuerten Qualitätssicherung und -verbesserung der Ausbildung in einem Fachbereich“ (S. 21) erreicht wird, ist nicht thematisiert. Obwohl im Abschnitt „Evaluation“ sechsmal von Verbesserung oder Optimierung gesprochen wird, werden deren Messung und die Ergebnissicherung des ganzen

Verfahrens und die bei Betriebswirtschaftlern doch nahe liegende und im Text behandelte Kosten-Nutzen-Rechnung nicht aufgegriffen.

Lehrveranstaltungsevaluationen können indirekt als Indikator für Veränderungen auf Fachbereichsebene oder eines ganzen Ausbildungsganges herangezogen werden, falls sie repräsentativ die Veranstaltungen abdecken (Rindermann/Kohler 2003). Besser wäre jedoch, bei veranstaltungsübergreifender Evaluationsperspektive entsprechende Qualitätsmerkmale aus unterschiedlichen Perspektiven und mittels quantitativer Indikatoren in Messwiederholungsdesigns direkt zu bestimmen. Günstige Erfahrungsberichte finden sich in dem Bändchen der Hochschulrektorenkonferenz (2002).

3. Praktizismus und Formalismus

Es gibt neben der Psychotherapieevaluation kaum einen Bereich der Evaluationsforschung, der so eingehend untersucht wurde wie den der Lehr(-veranstaltungs)-evaluation. Nach Cashin (1995) sind allein bis 1994 im englischsprachigen Raum 1500 Beiträge zur Lehrevaluationsforschung erschienen. Autoren wie Wilbert McKeachie, Kenneth Feldman oder Herbert Marsh haben schon seit den 70er Jahren aus Perspektive der Psychologie unter Zuhilfenahme oft sehr aufwendiger statistischer Methoden eine große Anzahl von wissenschaftlichen Arbeiten zur Lehrveranstaltungsevaluation publiziert, in Deutschland gibt es seit den 70er Jahren und erneut seit den 90ern ebenso eine Vielzahl von Arbeiten, die sich fundiert mit Fragestellungen der Lehrevaluations- und Hochschulforschung auseinandersetzen. Desto mehr fällt ein dreifaches Defizit an Forschungsrezeption auf: Bei der Bewertung von Evaluationen und der Bewertung der Urteilsfähigkeit der Studierenden durch Hochschullehrer wird der Forschungsstand nicht ausreichend zur Kenntnis genommen (1), von Evaluationsforschern und Praktikern wird aus der Breite der Hochschulforschung nur selektiv rezipiert (2) und bei der Planung und Durchführung von Lehrevaluationsprojekten durch Praktiker fließen Erfahrungen und Arbeiten der Wissenschaft kaum ein, statt dessen dominieren formale oder untersuchungstechnische Orientierungen (3). Es liegt ein Praktizismus vor, ohne dass dieser der Praxis dient.

3.1 Nichtbeachtung von Forschung bei der Bewertung von Lehrevaluation

Mangelnde Kenntnisnahme des Forschungsstandes bei der Bewertung der Lehrevaluation und der Urteilsfähigkeit von Studierenden durch Hochschullehrer wurde auch schon für die Vereinigten Staaten von Cohen (1990: 124-125) beschrieben: „Negative attitudes toward student ratings are especially resistant to change, and it seems that faculty and administrators support their belief in student-ratings myths with personal and anecdotal evidence, which [for them] outweighs empirically based research evidence.“ Wieso neigen aber Wissenschaftler, die in Kritik und systematisch-erfahrungswissenschaftlicher Herangehensweise an Probleme geschult sind, zu einer ihrer sonstigen wissenschaftlichen Einstellung entgegengesetzten Haltung?

In der Hochschul- und Lehrevaluationsforschung werden die Lebenswelten der Wissenschaftler untersucht. Zwischen Praxis und Forschung besteht wie in der Hochschulforschung nur eine vage Trennung (vgl. Teichler 1994). Distanzierung von Alltagserfahrungen und Handlungsentlastetheit stellen aber Bedingungen wissenschaftlicher Rationalität dar. Die Evidenz lebensweltlicher Erfahrungen und der fehlende Bruch mit dieser Primärerfahrung legen dagegen eine perspektivische Interpretation aus Sicht eines Hochschullehrers nahe. Bei der Bewertung von Lehrevaluation und der Urteilsfähigkeit der Studierenden wird der Zugang über die eigene Erfahrung gewählt, nicht über die Wissenschaft. Warum ist dies im Falle der Lehrevaluation bedeutsam? Die Beurteilung von Hochschullehrern durch Studierende dreht die gängige Bewertungsrichtung um, sie führt zu einer (partiellen) Hierarchie-Inversion und verletzt so die traditionelle Struktur einer historisch gewachsenen Institution. Die Einführung eines Lehrevaluations-Systems und die mögliche Berücksichtigung der Lehrqualifikation durch Evaluation bei der Stellenvergabe und Gehaltsbemessung würden die bisherigen Reproduktions- und Kooptationsmechanismen in der Wissenschaft und deren Autonomie gefährden. Mit erfolgreicher Integration in ein soziales System werden aber dessen Regeln und Strukturen stärker verteidigt. Erfolgreiche Sozialisation lässt sich durch ein hohes Maß an Symmetrie von objektiver mit subjektiver Wirklichkeit kennzeichnen. Bei der Untersuchung von Einstellungen zur studentischen Lehrevaluation zeigte es sich dementsprechend, dass insbesondere Professoren skeptisch gegenüber der studentischen Urteilsfähigkeit eingestellt sind (Rindermann 2000, 2001).⁵

3.1 Fachspezifische Rezeption von Forschung

Lehrevaluation ist Gegenstand der Forschung unterschiedlicher Disziplinen und Praktiker entstammen unterschiedlichen Fachrichtungen (Helmke/Krapp 1999). Neben Psychologie, Soziologie, Pädagogik und Verwaltungswissenschaften sind auch innerhalb der genannten Fächer unterschiedliche inhaltliche und methodische Orientierungen erkennbar: An Pädagogischer Psychologie, Unterrichtsforschung, Diagnostik, Methodenlehre, Evaluationsforschung, Hochschulsozialisation, Hochschulforschung, Bildungsforschung usw. Ansätze der eigenen Arbeit und Publikationsorgan dieser, Literatursuche und Rezeption, Verständnis und Wertschätzung der Arbeiten anderer sind abhängig von Mechanismen fachlicher Binnensozialisation und Binnenselektion. Hochschulforschung ist wenig institutionalisiert und im deutschen Sprachraum fehlen Studiengänge wie „Higher Education“. Das wissenschaftliche Feld ist zersplittert, Evaluationsforscher und -praktiker orientieren sich weitgehend an Traditionen ihres eigenen Faches. So wird etwa von Soziologen und Bildungsforschern kaum die stark empirisch-statistisch orientierte angloamerikanische Lehrevaluationsforschung zur Kenntnis genommen, Psychologen wiederum nehmen die stärker pädagogisch, soziologisch, organisations-, bildungs- und ver-

5 Bei Rückmeldung schlechter individueller Evaluationsergebnisse können Strategien der Abwertung (der Fähigkeiten der Studenten, ihrer Urteilskompetenz etc.) auch dem Selbstwertschutz dienen (vgl. die prägnante Darstellung von Wottawa 2001: 153f.; er spricht vom „Prinzip der Hedonistischen Verzerrung“).

waltungswissenschaftlich geprägten Arbeiten der Hochschuldidaktiker und Hochschulforscher kaum auf. Dies obwohl verschiedene Forschungsfelder wie die der Lehrqualität oder der Hochschulsozialisation in Psychologie *und* Pädagogik *und* Soziologie fachlich angesiedelt sind (vgl. Helmke 2001; Tippelt 2002). Um im eigenen Feld zu reüssieren, um publizieren zu können und um gelesen zu werden, werden die fachspezifischen Grenzen nicht überschritten.

3.3 Evaluationspraxis als formaler Verwaltungsakt

Fehlende Theorie-Praxis-Differenzierung und fachliche Zersplitterung machen aber nicht verständlich, dass bei der Durchführung von Evaluationsmaßnahmen der wissenschaftliche Kenntnisstand unabhängig von seiner Herkunft wenig rezipiert und umgesetzt wird. Bei Evaluationsmaßnahmen sind zunächst verschiedene Organisationsformen zu unterscheiden: als eigenständige Abteilungen oder Projekte, besetzt mit hauptamtlich für Evaluation zuständigen Akademikern in entweder ganzen oder halben Dauerstellen oder ganzen oder halben befristeten Projektstellen. Oder Evaluationsmaßnahmen werden durch hauptamtliche Fachwissenschaftler aller an den Universitäten vertretenen Fächer durchgeführt, denen in einer zusätzlichen Nebentätigkeit in der eigenen Fakultät ähnlich der Aufgabe eines Institutsvorstandes und Dekans die Zuständigkeit für Lehrevaluation übertragen wurde oder die für Fachbereiche anderer Universitäten als externer kollegialer Gutachter wie bei der Manuskript- und Projektbewertung üblich angefragt wurden.

Gemeinsam ist diesen Varianten, dass Evaluation in der Praxis als der Universität vom Gesetzgeber oder Ministerium aufgetragene Aufgabe verstanden wird, die gemäß der gegebenen Vorgaben zu erfüllen ist und deren Erfüllung nach innen, in die Universität, nach außen, in die eventuell interessierte Öffentlichkeit, und vor allem nach oben, in die die Auflagenerfüllung überprüfende Verwaltung, zu dokumentieren ist. Diese reicht das Vollzugsergebnis als abgezeichnete Akte dann durch die Instanzen weiter. Evaluation stellt hier einen Verwaltungsakt dar, oft ist die Evaluationsstelle – wie etwa an der Universität Magdeburg – in der Verwaltung selbst angesiedelt.

In traditionellen Formen der Verwaltung hat das Prinzip der Legalität oberste Priorität (Weber 2002/1922). Legalität rangiert vor Funktionalität. Regelorientierung ist wichtiger als Zielorientierung. Juristisches Denken dominiert. Die Begründung von Maßnahmen liegt nicht in der Idee inhaltlicher Richtigkeit, sondern in der Korrektheit des Verfahrens, mit der sie beschlossen und durchgeführt werden. „Das Erfordernis der Rechtsbindung der Verwaltung lässt ein Spannungsverhältnis zur Verfahrenseffizienz nicht zu. Die formelle Rechtsbindung der Verwaltung, das heißt ihre Pflicht, das geltende Recht zu beachten und zu vollziehen, duldet keine Abstriche im Interesse effizienter Gestaltung des Verfahrens.“ (Fuchs 2001: 287). Verwaltungsleistung wird an rechtlicher Richtigkeit festgemacht, auch wenn durch die Durchführung einer Anordnung und insbesondere durch die Art der Durchführung einer Anordnung die Erreichung des ursprünglichen Ziels einer Institution, zu dem die Evaluation als Maßnahme beitragen soll, gehemmt oder gar verhindert wird. Für die Universität stellen die wesentlichen, selbstgesetzlichen Bestimmungen *Gewinnung und Bewahrung von Erkenntnis* sowie *akademische Aus-*

bildung und Bildung von Studierenden dar. Diese stehen vor externen gesellschaftlichen, politischen oder wirtschaftlichen Funktionalisierungen.

Als Erfolg einer Maßnahme wird im Rahmen der Verwaltung das definiert, was die auferlegten Verpflichtungen in für die Verwaltung nachprüfbarer Form erfüllt. Als Beleg hierfür dient bei der Evaluation in der Regel ein „Bericht“. Merkmale und Inhalte des Selbstberichts oder Fremdberichts stehen für die Evaluation. Was sind Merkmale eines Evaluationsberichts, anhand derer die Befolgung der Evaluationsauflage geprüft werden kann? Der Bericht muss den Kriterien einer aus der Sicht der Verwaltung wissenschaftlichen Darstellung genügen bzw. präziser denjenigen Kriterien genügen, die innerhalb der abnehmenden Instanzen als Nachweis seriös geleisteter Arbeit gelten: Dazu zählt etwa die Abgabe innerhalb der vorgegeben Fristen, anfänglicher Verweis auf Gesetz, Erlass und Finanzierungshilfe, Abzeichnung durch die innerhalb der Institutionen relevanten Personen (Direktoren u.ä.). Womöglich kommen auch textformale Kriterien hinzu wie kein zu geringer Textumfang, eine beim Durchblättern bemerkbare sorgfältige Formatierung und saubere Textgestalt, Quantität und Herkunft von Fremdwörtern (vornehmlich Latinitäten und Anglizismen), Armut an erkennbaren Rechtschreibfehlern, etwas komplizierterer Satzbau und ein komplexeres Satzgefüge, Einsatz von numerisch-mathematischer Rhetorik in Form von Zahlen, Formeln und grafischen Darstellungen, die quantitative Größen und Beziehungen demonstrieren. Inhaltliche Aussagen oder gar Bewertungen werden bei kritischen Tatbeständen durch Formen semantischer Relativität vermieden: Hierzu dienen Euphemismen, Mengenunbestimmtheit, Problemdiminuierung, Einerseits-Andererseits-Unbestimmtheiten oder Distanzierungen durch Gänsefüßchen (Heinrich 1994, 2003). Ein Beitrag für die Gestaltung und Verbesserung der Lehrpraxis oder gar eine Antwort auf theoretisch bedeutsame Fragestellungen ist dagegen vergleichsweise unwichtig. Kritik und Infragestellung gegebener Verfahrensabläufe und Strukturen – der evaluierten oder evaluierenden Praxis –, die Veränderungsvorschlägen immer implizit sind, wären nicht angezeigt. So gibt es Evaluationsansätze und Evaluationsberichte – vor allem in den Evaluationsvarianten Selbstevaluation und Lehrbericht –, die die regionale Herkunft der Studierenden oder Lesbarkeit des Tafelanschiebes, nicht aber Unterrichtsqualität, Betreuungsverhältnisse oder Absolventenqualifikation thematisieren.

Übersicht 2: Konventionelle versus moderne Verwaltung?

Auf einer Diskussion dieser Thesen an der Humboldtuniversität im September 2002 wurde angesprochen, dass das hier skizzierte Bild nicht mehr der Selbstsicht der Verwaltung und einer angestrebten modernen Verwaltungspraxis (maximale Funktionalität im Rahmen der Legalität) entspricht. – Verwaltungswissenschaften vertreten in Forschung und Ausbildung ein anderes Ziel als klassische Verwaltung und die Verwaltungsmodernisierung versucht Prinzipien und Verfahren wie Zielerreichung, Performanzorientierung, Kosten-Leistungsrechnung und Effizienz zu etablieren (vgl. Pitschas 2000; Rürup 2000; Wollmann 2002). Die Praxis der Verwaltung an Hochschulen, die wiederum von der Kultusbürokratie und politischen Instanzen abhängig ist, sieht in der Regel aber noch anders aus (Höllinger/Steinbacher 1992). Beispiele: Studenten haben aufgrund von starr gehandhabten Ausschlussfristen Schwierigkeiten, sich an der gleichen Hochschule in einen neuen Studiengang umzumatrikulieren, weil nicht bis zu einem bestimmten Termin über das „Bestanden“ hinaus eine präzise Note vom hochschuleigenen Prüfungsamt geliefert werden konnte. Reisekostenabrechnungen gehen bei einer Abweichung von 50 Cent der Fahrkarte von der billigst möglichen für eine Strecke wieder zurück mit der Konsequenz der Beschäftigung einer Ökonomin, einer Sekretärin und einer Wissenschaftlerin (insgesamt ca. 1,5 Zeitstunden Mehrarbeit; andere Beispiele s. Luh-

mann 1992). Es gibt selbstverständlich auch Gegenbeispiele, um zu einem methodisch abgesicherten Urteil zu kommen, müsste man neben einer *Befragung von Verwaltungskunden* (Studierende, Lehrende) die Effizienz der Verwaltung systematisch untersuchen und mit der der *Privatwirtschaft* oder von *Universitäten anderer Länder und Staaten vergleichen*.

Zentral für diesen Beitrag zur Evaluation ist jedoch nicht, wie Verwaltung oder Hochschulverwaltung heute arbeiten, sondern die *Dominanz formal-verwaltungsmäßiger Orientierung innerhalb der Praxis der Evaluation an Hochschulen*. Das Problem ist nicht die Verwaltung, sondern die Art der Praxis von Evaluation, die als formaler Verwaltung stark ähnelnd beschrieben wurde.

Ziel und Zweck einer Maßnahme – hier von Hochschulevaluation – können im Verlauf ihrer Durchführung verloren gehen. Allein schon die Frage nach dem Wozu und Warum aber ist mit der traditionellen Erfüllung einer Verwaltungsaufgabe und verwaltungsgesteuerter Evaluation nicht kompatibel, denn die Reflexion auf Sinn und Zweck des Handelns ist nicht gefordert, nur die korrekte und dokumentierbare Ausführung der übertragenen Aufgabe. Der Gedanke schon an Sinn und Zweck, an Funktionalität einer Maßnahme für das ursprüngliche Ziel, wäre nicht mehr symmetrisch mit Prinzipien der Verwaltungstätigkeit. Evaluation als solche Tätigkeit sollte möglichst störungsfrei für alle verlaufen und im Sinne einer behördlichen Auflage erfolgreich abgearbeitet werden. Evaluation bleibt so Ausdruck systemimmanenter Denkens und Beharrungsvermögens. Mit von Verwaltung geprägter Praxis geht strukturbedingt die Betonung und Formung bestimmter, sich auch auf Persönlichkeit auswirkender Merkmale einher wie Korrektheit, Genauigkeit, Formalität und im Rahmen des Auftrages notwendiger Kompetenzen wie Organisationsgeschick und Computerbedienfähigkeiten, weniger aber Reflexion und Innovation solcher Praxis (Häußermann 1977). Aufgabe und personenbezogenes Merkmalsmuster, objektive und subjektive Strukturen, Umweltanforderungen und Personendispositionen, die miteinander in Übereinstimmung stehen, sprechen für einen unauffällig-korrekten Verfahrensablauf als Ziel der Verwaltung.

In dieser Form wird die Evaluationspraxis der ihr eigenen Zielsetzung aber nicht gerecht. *Praxis* als tätige Auseinandersetzung mit der Wirklichkeit im Gegensatz zur Theorie ist immer ziel- und zweckbezogen, dient der Verfolgung eines subjektiven Sinns und objektiven Zwecks. Zwar ließe sich Evaluation als formaler Verwaltungsakt ohne Bewertungs-, Entscheidungs- und Verbesserungsperspektive immer noch ein Zweck unterstellen, sei es *institutionell* durch Erfüllung behördlicher Auflagen die Vermeidung von Problemen mit Ministerium und Verwaltungsinstanzen oder sei es *persönlich* die Möglichkeit eine bezahlte Anstellung innezuhaben, sich durch Übernahme der Evaluationstätigkeit anderer Aufgaben zu entledigen, oder sei es *wissenschaftlich* die Datengewinnung für die Forschung.⁶ Der

6 Evaluation als formaler Verwaltungsakt kann auch die Forschung prägen: Oft wird nur mit mehr oder weniger zufällig angefallenen Daten als Nebenprodukt standardisierter Erhebungen ohne ein explizites Forschungskonzept gearbeitet, übliche Forschungsprotokolle wie Korrelation mit Geschlecht werden schulmäßig heruntergerechnet. Bedeutsame Aufgabenstellungen wie systematische Metaanalysen im Bereich der Lehrevaluationsforschung, Prüfung der Effektivität von Evaluation und Beratung u.ä. werden dagegen vernachlässigt. Beiläufig gesammelte Daten sind für verschiedene Anwendungen nicht unbrauchbar (z.B. Erstellung von Normen, Berechnung von Faktorenanalysen, Prüfung der Alltagstauglichkeit), bedürfen aber der Ergänzung durch an Fragestellungen orientierter Forschung.

Zweck der *Evaluation*, durch Informationsgewinnung mittels systematischer Analyse und empirischer Untersuchung von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zu deren Bewertung und vor allem zu deren Verbesserung beizutragen (Hager/Patry/Brezing 2000; Rindermann 2001), geht aber verloren.

Evaluation könnte einer evaluierten Aktivität sogar schaden: Zunehmende Verwaltungstätigkeit – zu der auch Evaluation in dem beschriebenen und kritisierten Sinne gehört – bindet Zeit, Personal und Mittel, die kreativer Forschungs- und Lehrtätigkeit fehlen. Webler (2002: 158) hat einmal am Beispiel Nordrhein-Westfalen für das dort präferierte Peer-Review-Evaluationsmodell eine aktuelle Aufwandsberechnung vorgenommen: Während der Vorlesungszeit müssten die 15 Landesuniversitäten mit im Schnitt jeweils 15 Fachbereichen von je vier Professoren als Peers evaluiert werden, hierzu wären etwa 800 bis 900 Professoren erforderlich, die pro Fachbereich mit Vorbereitung, Besuch und Nachbereitung je acht Tage bräuchten, insgesamt also 6400 Arbeitstage im Semester. In dieser Zeit könnten sie selbstverständlich nicht selbst lehren und forschen.⁷ Die monetären Kosten beliefen sich auf 8 bis 15 Millionen Euro, hinzukommen die Opportunitätskosten durch Wegfall von Arbeitszeit. Zudem könnte eine Verordnung von Lehrqualitätsmessung ohne Personen und Institutionen zu ermöglichen, Qualität zu verbessern, sich kontraproduktiv auswirken. Evaluation kann bei kritisch bewerteten Dozenten und Fachbereichen ohne Verbesserungsperspektive zu Frustration und Entmutigung führen, Cranton und Knoop (1991) sprechen sogar von „*professorial melancholia*“ (S. 100). Diese Form von Evaluation wäre nicht nur dysfunktional, sondern bei personenbezogener Verwendung auch unethisch: „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: 126). Ähnlich Marsh und Roche (1999: 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.“

Lehrevaluation als Verwaltungsaufgabe ohne Qualitätsorientierung könnte genau das Gegenteil dessen erreichen, was sie (im engeren Sinne) bezwecken soll: Statt Verbesserung der Lehre und der Ausbildung deren Verschlechterung. Die Form dieser Praxis an einer Institution, deren zentrale Orientierungen um Rationalität, Erkenntnis und Bildung kreisen, führt an die Grenzen dessen, was noch mit Rationalität von institutionellem Handeln beschreibbar ist.

Vielleicht ist eine Praxis, die zur Erreichung eines bestimmten Zieles etabliert wurde, zu dieser Erreichung aber nicht nur kaum beiträgt, sondern die Zielannäherung sogar beeinträchtigen kann, die man deshalb legitimerweise als höchst widersprüchlich bezeichnen muss, vielleicht ist eine solche Praxis gut mit in literarischen

7 Ein konkretes Beispiel: Der Hochschulforscher Ulrich Teichler meinte im Februar 2003 auf der 4. Arbeitstagung zur Evaluation an Hochschulen der Hochschulrektorenkonferenz an der Universität Potsdam, dass er so viele Anfragen zur Fachbereichsbegutachtung bekäme, dass er damit 150% seiner Arbeitszeit ausfüllen könne.

Vorläufern beschriebenen Handlungsabläufen vergleichbar. Ein solcher Vergleich kann Prozesse pointiert veranschaulichen. So fand die Praxis dysfunktionaler, rein formaler Verwaltungsprozesse einen prägnanten literarischen Ausdruck in den beiden Romanen *Das Schloß* und *Der Prozeß* von Franz Kafka: Emsige, an Regelungen und Weisungen orientierte Betriebsamkeit ohne nachvollziehbaren subjektiven Sinn und objektiven Zweck kennzeichnen das Auftreten der Figuren. Diese charakterisiert die nicht in verständlichen Mittel-Zweck-Beziehungen stehenden, aber von extremer Anstrengung geprägten Tätigkeiten der aus dem für andere unerreichbaren Schloss stammenden Beamten und Gehilfen und diese charakterisiert das geschäftige Treiben der Advokaten und Wächter in einem labyrinthischen und von der Schuldfindung völlig entkoppelten Justizverfahren.⁸

Warum wird Evaluation entgegen ihrer für die Praxis produktiven Zwecksetzung in wenig funktionaler Weise gehandhabt? Zum einen mag die Anbindung an Verwaltungsroutinen eine Rolle spielen. Defizite in Kenntnis und Anwendung der Evaluation und Evaluationsforschung lassen sich nicht ausschließen. Wichtiger ist, dass die spezifische Tradition deutscher Hochschulorganisation eine Orientierung an Effizienzgesichtspunkten verhindert und selbst das Erkennen von Effizienzdefiziten erschwert. Es ist das Zusammenspiel dieser Faktoren und die Möglichkeit, durch eine formal-verwaltungsmäßig orientierte Evaluation Qualitätsbewusstsein zu dokumentieren ohne in einem stabilen System etwas ändern zu müssen. Evaluation wirkt so strukturell konservativ.

4. Schlussfolgerungen für Praxis und Forschung

4.1 Evaluationsforschung

Die gegenwärtige Praxis der Lehrevaluation wurde kritisch beschrieben. Das Reflektionsdefizit und das Effektivitätsdefizit innerhalb der Evaluationspraxis und Evaluationsforschung resultieren in einem nur wenig hilfreichen Beitrag zur Verbesserung der Lehre an Hochschulen. Die näher analysierten Phänomene der Nichtrezeption von Forschung, fächerbezogenen Fragmentierung von Lehrevaluationsansätzen und verwaltungsprinzipiengesteuerten Evaluation wurden auf spezifische Bedingungen der Hochschulorganisation und Hochschulforschung (perspektivische Primärerfahrung, Folge fachlicher Binnensozialisation und Binnenselektion, institutionelle und hochschulpolitische Rahmenbedingungen) sowie auf die stabilisierende Funktion der bisherigen Handhabung und Nichtreflexion von Evaluation zu-

8 Zur Prüfung, ob dieser Vergleich erlaubt ist, wäre einmal zu testen, ob es möglich wäre, einen völlig sinnfreien Bericht zu schreiben, ohne dass die Simulation bemerkt werden würde. So wurde schon in Evaluationsprojekten halb im Spaß und halb im Ernst vorgeschlagen, einfach den Lehrbericht vom letzten Mal mit neuem Datum versehen an die Universitäts- und Kultusinstanzen weiterzureichen, oder die Menge an verbrauchter, aus dem korrekten Titel bezahlter Tafelkreide und die Fläche der im Schnitt pro Stunde vollgeschriebenen Tafel in Quadratzentimetern und verbal angereichert sowie mit Grafiken versehen anzugeben. Anders als bei Sokals Jux wäre die Aufdeckung einer solchen Persiflage aber wahrscheinlich ohne inhaltliche Folgen für die mögliche „Wiedervorlage“.

rückgeführt. Zukünftige Theorie und Forschung von Hochschulunterricht und ihrer Evaluation müssen diese Bedingungen analysieren und für eine veränderte Praxis berücksichtigen. Evaluation ist für die Wissenschaft im doppelten Sinne zurückzugewinnen: Als Tätigkeit, die wissenschaftsgestützt vorgeht und sich wissenschaftlicher Überprüfung stellen muss, und als der Praxis von Wissenschaft und Hochschulen dienendes Instrument zur Qualitätssicherung und -verbesserung eben dieser Praxis.

Die Forschung sollte die bisherigen Erfahrungen aufgreifen, etwa in einer Untersuchung des Zusammenhangs zwischen Lehrqualität, Lehrevaluation und dem Gebrauch der Lehrevaluation in Abhängigkeit von unterschiedlichen institutionellen Organisationsformen und deren Änderung. So zeigte sich etwa, dass es bei privater Trägerschaft einer Ausbildungseinrichtung eher möglich ist, ein effektives und objektiv nachweisbar qualitätsförderndes Evaluationsverfahren zu etablieren (Rindermann/Kohler 2003).

Arbeiten aus Nachbardisziplinen müssen sowohl in der Evaluationsforschung als auch in der Evaluationspraxis aufgegriffen werden, die „fachliche Zersplitterung“ lässt sich so zu den Zielen von Hochschulevaluation dienender Interdisziplinarität zusammenführen. Als Bindeglied für Konzeption, Ausführung und Verwendung von Lehrevaluationen in der Praxis könnten zukünftig Evaluationsstandards dienen (Sanders/Beywl 2000).

Dringend anzuraten ist auch eine Professionalisierung für Evaluationen und Evaluatoren. Evaluationsmaßnahmen müssen ausreichende Mittel und interdisziplinär ausgebildetes Personal zur Verfügung stehen, mit denen aussagekräftige Evaluationen durchgeführt, Beratungen und (hochschuldidaktische) Weiterbildung angeboten und Effekte dieser Maßnahmen überprüft werden können. Evaluation ohne Qualitätsorientierung, ohne Einmündung in Entscheidungen und ohne systematisch-empirische Effektivitätsprüfung – Evaluation ohne Verwendung und ohne Veränderung – sollte grundsätzlich nicht an Hochschulen durchgeführt werden. Evaluation als bloße Rechenschaftslegung (im Sinne einer reinen „Akkreditierung“) ist zu aufwendig und zu teuer und erschwert die Verfolgung der eigentlichen Ziele von Hochschulen: Forschung und Lehre. Evaluation im Gegensatz zur reinen Forschung muss effektiv (wirksam, zielannähernd) und effizient (nützlich) sein und diese Effektivität und Effizienz müssen geprüft und nachgewiesen werden, optimal wären hier klassische Designs wie Vorher-Nachher-Messungen und die Verwendung von Kontrollgruppen zur Ergebnissicherung. Evaluation dient hierbei als niederschwellige Intervention und vor allem als Maß der Wirkung der einer Evaluation vorausgehenden, nachfolgenden oder durch sie angestoßenen Interventionen und Änderungen. Evaluation ist praxisorientiert und muss ihre Relevanz für die Praxis nachweisen.

Lehrevaluation und Lehrveranstaltungsevaluation 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 aus Effektivitätsgründen wie auch aus ethischen Gründen ein Ansatz abzulehnen, Evaluation – die bloße Messung der Lehrqualität – ohne Möglichkeit der Qualitätsverbesserung durchzuführen. Eine Institution zu evaluieren macht nur dann Sinn, wenn Möglichkeiten bestehen, auf Ergebnisse zu reagieren und förderliche Veränderungen einzuleiten. Qualität könnte sonst sogar vermindert werden, weil Zeit für Forschung und Lehre

verloren geht. So sollte Lehrevaluation außer in problematischen Fällen oder bei Interventionsmaßnahmen nicht öfters als alle zwei bis vier Jahre je Dozent und sechs bis zehn Jahre je Fachbereich durchgeführt werden.

Ist man an der Verbesserung der Qualität der Lehre in Form von Lehrveranstaltungen interessiert, sollte die Lehre im Rahmen eines *Beratungs-Rückmelde-Verfahrens* evaluiert werden. Die private Nutzung von Lehrveranstaltungsevaluation kann der Selbstvergewisserung über den eigenen Hochschulunterricht dienen, in günstigen Fällen lässt sich mittels dadurch initiiertes Kommunikationsprozesse auch ohne Beratung die Lehre verbessern.⁹ Ist Verbesserung der lehrveranstaltungsübergreifenden Ausbildungsqualität das Ziel, sind interne Selbst- und externe Fremdevaluationsprozesse durch Kollegen (Peers) sowie studentische Einschätzungen in *effektive Kommunikations- und Entscheidungsprozesse* zu überführen, an deren Ende jeweils Anreize für gute Lehre stehen müssen. Moderne Instruktions-, Management- und Organisationskonzepte sowie Ergebnisse der Hochschul- und Lehrevaluationsforschung liegen hierfür bereit (z.B. Theall/Franklin 1991; Webler 2002; Wottawa 2001). Eine Unterstützung durch externe Experten der Hochschul- und Evaluationsforschung ist anzuraten. Schweizer Hochschulen gehen Schritte in dieser Richtung, systematische Effektivitätsprüfungen stehen aber noch aus. Die Effektivität eines solchen Verfahrens für Lehrqualität muss über Erfahrungsberichte hinausgehend in wissenschaftlicher Form geprüft werden. Wirksamkeitsmaße für Lehrveranstaltungsevaluationen wären studentische Einschätzungen und die objektiven Leistungen in Prüfungen. Für veranstaltungsübergreifende Evaluationen sind die Kriterien schwieriger zu bestimmen, z.B. studenten- und peergestützte Urteile, Absolventenfähigkeiten und objektive Qualitätsmaße im Vergleich.

4.2 Möglichkeiten und Grenzen der Steuerung von Hochschulen durch Evaluation

Nicht übersehen werden sollte in der Diskussion um Evaluation an Hochschulen, dass Universitäten als Institution und insbesondere ihr Auftrag der Erkenntnisgewinnung und Ausbildung wie Bildung nur eingeschränkt effektivitäts- und effizienzorientiert steuerbar sind: Die Grenzen eines an Effektivität, Effizienz und Meritokratie orientierten Steuerungssystems liegen an der inhaltlich schwierigen *Messbarkeit* von Lehr-, Forschungs- und Hochschulqualität – quantitative Indikatoren sagen etwas über Qualität aus, bilden diese aber nicht vollständig ab, quantitative Indikatoren und Qualitätsurteile können durch Störfaktoren beeinflusst werden, lassen sich bei Verbindung mit Gratifikationen einseitig maximieren (z.B. bei Publikationszählung durch Publikationszersplitterung Erhöhung der Publikationsrate), sie sind abhängig von Beziehungsmustern zwischen Urteilenden und Beurteilten, deren wissenschaftlichen Orientierungen usw. Studentische Lehrveranstaltungsevaluation ist hierbei vergleichsweise aussagekräftig und diese ist gut erforscht. Die Grenzen liegen neben Bestimmungsproblemen aber auch in *Umsetzungsdefiziten* und im Be-

9 In der Regel gehen eher die motivierteren Dozenten und vor allem diejenigen Dozenten diesen Weg, die gute bzw. gut beurteilte Lehre halten.

harrungsvermögen einer behördlich organisierten Institution, in *ethischen Problemen* (können Personen, Fächer oder Universitäten mit unterschiedlichen Eingangsvoraussetzungen einheitlich hinsichtlich ihrer Leistung bewertet werden?) und in Grenzen, die durch eine *ingeschränkte Funktionalität des meritokratisch-steuern-den Modells* für wissenschaftliche Arbeit und Universitäten bedingt sind: Freiheits-spielräume sind zum Gelingen wissenschaftlicher Arbeit notwendig. Forschende und unterrichtende Tätigkeit, die nicht im wesentlichen selbstmotiviert und selbst-gesteuert durchgeführt werden würden, würden als nur extern gesteuerte zur lang-fristig unproduktiven Auftragsarbeit. Die universitäre Freiheit vor äußeren Eingrif-fen sichert einen gewissen Raum vor externen staatlichen, gesellschaftlichen und kollegialen Eingriffen durch Ministerien, Moden, externe Geldgeber, Wissen-schaftsverbände und Kollegien, der sowohl privatisierenden Missbrauch ermöglicht als auch unkonventionellen wissenschaftlichen Wegen ihre Freiheit lässt. Vielfalt kann entstehen, neue Ideen entwickeln sich. Allerdings ist unverkennbar, dass ein Hochschulsystem, das gänzlich auf die Mittel der Steuerung durch Evaluation und Qualitätsmanagement, durch Qualifizierung und Gratifikation verzichtet, sich damit auch wirksamer Möglichkeiten enthält, Qualität in Forschung und Lehre zu erhöhen.

4.3 Beziehung zwischen Evaluationsforschung und Fachdisziplinen

Welche Bedeutung kommt auf die Fachwissenschaften im Rahmen einer überzeu-genderen Evaluationspraxis und Evaluationsforschung zu? Die *pädagogischen Dis-ziplinen* (Pädagogische Psychologie, Pädagogik usw.) können Erkenntnisse aus der Unterrichtsforschung bereitstellen, die darlegen, wie gute Lehre erreicht und (Hochschul-)Lehren als praktische Kompetenz vermittelt werden können. Die *Dia-gnostik* hilft geeignete Instrumente zu entwickeln, um die Qualität von Lehre und Forschung zu bestimmen. *Organisations- und Verwaltungsforschung* untersuchen, wie Institutionen und Arbeitsprozesse organisiert werden sollten, damit optimale Ergebnisse und ein förderliches Arbeitsklima erreicht werden. Die *Sozialpsycholo-gie* kann helfen, Interaktionsprozesse innerhalb des Hochschulunterrichts und inner-halb der Hochschule als Institution besser zu verstehen. Beide Unterdisziplinen sollten auch dazu beitragen, die Durchführung von Evaluation, den Umgang mit und die Anwendung von Evaluationsergebnissen zu optimieren. Die *Differentielle Psychologie* kann helfen, Persönlichkeitsmerkmale, die für gutes Unterrichten not-wendig sind, zu bestimmen. Die *Allgemeine Psychologie* ist vor allem hinsichtlich der Aufklärung von Lernvorgängen und Motivation relevant. Und *Methodenlehre, Statistik* und *Evaluationsforschung* können in ihrer Breite quantitativer und qualita-tiver Verfahren Hilfsmittel zur Durchführung und Auswertung von Evaluation und für Evaluationsforschung und den Umgang mit erzielten Ergebnissen bereitstellen. Einzelne Disziplinen dürfen aber nicht absolut gesetzt werden. Evaluation und Evaluationsforschung sollten immer interdisziplinär orientiert sein, um tatsächlich Qualität bestimmen und verbessern zu können.

An wissenschaftlichen Standards orientierte Lehrevaluationsforschung, ihre Me-thoden und Ergebnisse, können wiederum für verschiedene Disziplinen bedeutsam sein, vor allem für die *Unterrichtsforschung* (in Psychologie, Pädagogik und Hoch-schuldidaktik), für die *sozialwissenschaftliche Methodenlehre*, die *Lernpsychologie*,

Organisationsforschung und für die *Wissenschaftsforschung* (z.B. Daniel/Fisch 1988; Krampen/Montada 2002; Teichler 2002). Dies kann aber nur dann gelingen, wenn in der Lehrevaluationsforschung theorie- und methodenorientiert gearbeitet wird. Zu oft wurden ohne Bezug auf eine Fragestellung gesammelte Daten für ad hoc generierte „wissenschaftliche“ Studien herangezogen, die dafür aber nur bedingt geeignet waren bzw. die Fragestellungen wurden dann der gegebenen Datennlage angepasst. Statistisch aufwendige Analysen als Indikator für „Wissenschaftlichkeit“ können in solchen Fällen an relevanten Fragen orientierte Forschung verdrängen. Lehrevaluation bedarf der Theorie und der interdisziplinären wissenschaftlichen Rezeption nicht nur um der Praxis willen. Es liegen aber schon jetzt für Wissenschaft, Fachdisziplinen und Praxis bedeutsame Ergebnisse vor: Beispielsweise weisen Resultate der Lehrevaluationsforschung in Regressionsanalysen, Befragungsstudien und Zusammenhangsprüfungen mit externen Kriterien auf die zentrale Rolle didaktischer und sozialer Kompetenzen des Dozenten für den Erfolg der Lehre hin: Lehrkompetenz (z.B. verständliche Sprache, Kompliziertes erklären können), Struktur, Verarbeitung (Anregung zum Mitdenken, kritische Beleuchtung), Auseinandersetzung/Breite (Verwendung von Beispielen, Vermittlung der Bedeutung von Themen, Theorie-Praxis-Bezug), Engagement (Vermittlung von Begeisterung und Motivierung der Teilnehmer) und Klima (Freundlichkeit, Kooperativität) sind die Merkmale, anhand derer sich der multikriteriale Erfolg einer Veranstaltung (in Kursbeurteilungen, Lerngewinn, Kompetenzerwerb und Einstellungsänderungen) am besten vorhersagen lässt. Lehrqualität kann in Beziehung zu Forschungsqualität und -quantität stehen. Effektivitätsanalysen prüfen die Bedeutung von Evaluation als Intervention (Metaevaluation) innerhalb gegebener Organisationsstrukturen und zeigen die Grenzen eines rein diagnostischen Ansatzes. Interventionsstudien demonstrieren die Veränderbarkeit von beruflichem Handeln unter verschiedenen Bedingungen. Lehrevaluation und Lehrevaluationsforschung, aber auch Grundlagen- und Anwendungsforschung unterschiedlicher Disziplinen können durch eine stärker wechselseitige Rezeption gewinnen: Lehrevaluation für die Praxis, Forschung für den Erkenntnisprozess, dessen Ergebnisse der Praxis wieder zugute kommen; Lehrevaluationsforschung hat beiden zu dienen, sowohl der Praxis wie der Erkenntnisgewinnung.

6. Literatur

- Altrichter, H./Schratz, M./Pechar, H. (Hg.) (1997): Hochschulen auf dem Prüfstand. Innsbruck: Studienverlag.
- Balk, M. (2000): Die Evaluation von Lehrveranstaltungen. Die Wirkung von Evaluationsrückmeldung. Frankfurt a. M.: Peter Lang.
- Beywl, W. (1988): Zur Weiterentwicklung der Evaluationsmethodologie: Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Frankfurt: Peter Lang.
- Cashin, W.E. (1989): Defining and evaluating college teaching. Manhattan: Center for Faculty Evaluation and Development (IDEA No. 21).
- Cashin, W.E. (1995): Student ratings of teaching: the research revisited. Manhattan: Center for Faculty Evaluation and Development (IDEA No. 32).
- Cohen, P. A. (1990): Bringing research into practice. *New Directions for Teaching and Learning*, 43, S. 123-132.

- Cranton, P./Knoop, R. (1991): Incorporating job satisfaction into a model of instructional effectiveness. In: Theall, M./Franklin, J. (Hg.): *Effective practices for improving teaching*. San Francisco: Jossey-Bass, S. 99-109.
- Daniel, H.-D. (1996): Evaluierung der universitären Lehre durch Studenten und Absolventen. In: *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 16(2), S. 149-164.
- Daniel, H.-D./Fisch, R. (Hg.) (1988): *Evaluation von Forschung: Methoden – Ergebnisse – Stellungnahmen*. Konstanz: Universitätsverlag.
- Diehl, J.M. (2001): Studentische Lehrevaluation in den Sozialwissenschaften. Fragebögen, Normen, Probleme. In: Keiner, E. (Hg.): *Evaluation (in) der Erziehungswissenschaft*. Weinheim: Beltz, S. 63-90.
- Doyle, K.O. (1991): Report on a trip downtown. In: Theall, M./Franklin, J. (Hg.): *Effective practices for improving teaching (New directions for teaching and learning 48)*. San Francisco: Jossey-Bass, S. 123-126.
- Fuchs, E. (2001): Rechtsbindung der Verwaltung als Politikproblem. Aspekte eines Missverständnisses. In: *Verwaltung und Management*, 7(5), S. 285-293.
- Gold, A./Souvignier, E. (2001): Referate in Seminaren. Warum man sie beibehalten und verbessern sollte. In: *Das Hochschulwesen*, 3, S. 70-74.
- Greimel, B. (2002): Lehrevaluation durch Beurteilungen der Lernenden – eine Analyse des Standes der Evaluationsforschung. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 98, S. 197-224.
- Hackl, P./Sedlacek, G. (2001): Evaluierung als Chance zur kontinuierlichen Verbesserung der Lehre. Das Beispiel der Wirtschaftsuniversität Wien. In: Spiel, Ch. (Hg.): *Evaluation universitärer Lehre*. Münster: Waxmann, S. 111-129.
- Hager, W./Patry, J.-L./Brezing, H. (Hg.) (2000): *Evaluation psychologischer Interventionsmaßnahmen*. Bern: Huber.
- Häußermann, H. (1977): *Die Politik der Bürokratie. Einführung in die Politik der staatlichen Verwaltung*. Frankfurt: Campus Verlag.
- Heinrich, P. (1994): *Sprache als Instrument des Verwaltungshandelns*. Berlin: Hitit.
- Heinrich, P. (2003): *Materialien zum Workshop „Probleme der Herausbildung eines ‚Evaluationscodes‘: Bluff oder Erkenntnis?“*. Universität Potsdam: 4. Arbeitstagung zur Evaluation an Hochschulen der Hochschulrektorenkonferenz.
- Helmke, A. (2001): Hochschulsozialisation. In: Rost, D. H. (Hg.): *Handwörterbuch Pädagogische Psychologie*. Weinheim: PVU, S. 255-260.
- Helmke, A./Krapp, A. (1999): Lehren und Lernen in der Hochschule. In: *Zeitschrift für Pädagogik*, 45(1), S. 19-24.
- Henninger, M. (2000): Evaluation: Diagnose oder Therapie. In: Harteis, C./Heid, H./Kraft, S. (Hg.): *Kompodium Weiterbildung. Aspekte und Perspektiven betrieblicher Personal- und Organisationsentwicklung*. Opladen: Leske + Budrich, S. 249-260.
- Hochschulrektorenkonferenz (Hg.) (2002): *Evaluation, was nun? Erfahrungen mit der Umsetzung von Evaluationsergebnissen*. Bonn: HRK, Beiträge zur Hochschulpolitik 1/2002.
- Höllinger, S./Steinbacher, W. (1992): Selbstorganisation und Management gegen Bürokratie. In: Altrichter, H./Schratz, M. (Hg.): *Qualität von Universitäten*. Innsbruck: Österreichischer StudienVerlag, S. 43-57.
- Kafka, F. (2002/1912): *Kritische Ausgabe*. Frankfurt: Fischer.
- Krampen, G./Montada, L. (2002): *Wissenschaftsforschung in der Psychologie*. Göttingen: Hogrefe.
- Kromrey, H. (1994): Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In: Mohler, P.Ph. (Hg.): *Hochschullehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Münster: Waxmann, S. 91-114.
- Leitner, E. (1998): The pedagogical qualification of the academic teaching staff and the quality of teaching and learning. In: *Higher Education in Europe*, 23(3), S. 339-349.
- Luhmann, N. (1992/1981): Zwei Quellen der Bürokratisierung in Hochschulen. In: Kieserling, A. (Hg.): *Universität als Milieu*. Bielefeld: Haux, S. 74-79.

- Marsh, H. W. (1987): Students' evaluations of university teaching. Research findings, methodological issues, and directions for future research. In: *International Journal of Educational Research*, 11, S. 253-388.
- Marsh, H.W./Roche, L. A. (1997): Making students' evaluations of teaching effectiveness effective. In: *American Psychologist*, 52(11), S. 1187-1197.
- Marsh, H.W./Roche, L. A. (1999): Reply upon SET research. In: *American Psychologist*, 54(7), S. 517-518.
- Meyer, H./Heesen, H. (2001): Von der Hochschul-Kosten- und Leistungsrechnung zum Hochschul-Controlling. In: *Verwaltung und Management*, 7(1), S. 17-23.
- Olbertz, J.-H./Pasternack, P./Kreckel, R. (Hg.) (2001): *Qualität – Schlüsselfrage der Hochschulreform*. Weinheim: Beltz, Deutscher Studienverlag.
- Perry, R.P./Smart, J.C. (Hg.) (1997): *Effective teaching in higher education: Research and practice*. New York: Agathon Press.
- Pitschas, R. (2000): Wirkungsorientierung des Verwaltungsmanagements im internationalen Vergleich. In: *Verwaltung und Management*, 6(6), S. 329-334.
- Rindermann, H. (1999): Was zeichnet gute Lehre aus? Ergebnisse einer offenen Befragung von Studierenden und Lehrenden nach Merkmalen guter Dozenten und Veranstaltungen. In: *Zeitschrift für Hochschuldidaktik*, 23(1), S. 136-156.
- Rindermann, H. (2000): Das Selbstobjektivierungsproblem im akademischen Milieu (I): Forschung an Hochschulen über die Qualität der Lehre an Hochschulen. In: *Das Hochschulwesen*, 48(3), S. 74-82.
- Rindermann, H. (2001): *Lehrevaluation – Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Mit einem Beitrag zur Evaluation computerbasierter Unterrichts*. Landau: Empirische Pädagogik.
- Rindermann, H./Kohler, J. (2003): Lässt sich die Lehrqualität durch Evaluation und Beratung verbessern? Überprüfung eines Evaluations-Beratungs-Modells. In: *Psychologie in Erziehung und Unterricht*, 50(1), S. 71-85.
- Rindermann, H./Oubaid, V. (1999): Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. In: *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20(3), S. 172-191.
- Rindermann, H./Wagner, A. (2003): Studienabbruch und Fachwechsel: Ursachen und Interventionsmöglichkeiten. In: Krampen, G./Zayer, H. (Hg.): *Psychologiedidaktik und Evaluation IV*. Bonn: Deutscher Psychologen Verlag, S. 232-249.
- Ringler, V. (2002): Der Feind im Inneren. Studenten als Spitzel, schwarze Listen politisch verdächtigter Professoren. In: *Die Zeit*, 46, S. 35.
- Rürup, B. (2000): Effizienzrevolution in der öffentlichen Verwaltung. In: *Verwaltung und Management*, 6(5), S. 265-268.
- Sanders, J. R./Beywl, W. (Hg.) (2000): *Handbuch der Evaluationsstandards*. Opladen: Leske + Budrich.
- Spiel, Ch. (Hg.) (2001): *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck*. Münster: Waxmann.
- Spiel, Ch./Fischer, U. (1998): Evaluierung eines Weiterbildungsangebots für Hochschullehrende. In: *Zeitschrift für Hochschuldidaktik*, 22(1), S. 83-99.
- Teichler, U. (1994): Hochschulforschung – Situation und Perspektiven. *Das Hochschulwesen*, 4, S. 169-177.
- Teichler, U. (2002): Die Zukunft der Hochschulen in Deutschland. Was sich aus der Perspektive der Hochschulforschung dazu sagen lässt. In: *Die Hochschule*, 11(1), S. 29-45.
- Teichmann, S. (1999): Wirksamkeit der Evaluation von Studium und Lehre. In: Olbertz, J.-H./Pasternack, P. (Hg.): *Profilbildung, Standards, Selbststeuerung*. Weinheim: Deutscher Studienverlag, S. 241-248.
- Theall, M./Franklin, J. (Hg.) (1991): *Effective practices for improving teaching*. San Francisco: Jossey-Bass (New directions for teaching and learning 48).
- Tippelt, R. (Hg.) (2002). *Handbuch Bildungsforschung*. Opladen: Leske + Budrich.

- Weber, M. (2002/1922): *Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*. Tübingen: J.C.B. Mohr.
- Webler, W.-D. (2002): Vergleich von fünf wichtigen Evaluationsmodellen in Deutschland – Grundlagen einer Meta-Evaluation. In: Neumann, K./Osterloh, J. (Hg.): *Gute Lehre in der Vielfalt der Disziplinen*. Weinheim: Beltz, S. 141-172.
- Weinert, F. (2001): *Die evaluierte Universität*. Heidelberg: Universitätsrede am 25. Januar 2001.
- Westermann, R. (2002): Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. In: *Zeitschrift für Psychologie*, 210(1), S. 4-26.
- Westermann, R./Rockenbach, B. (1999): Evaluation der Lehre zwischen angewandter Wissenschaft und kollegialem Ratschlag. In: Hacker, W./Rinck, M. (Hg.): *Zukunft gestalten. Bericht über den 41. Kongreß der DGPs*. Lengerich: Pabst, S. 583-591.
- Wissenschaftsrat (1996): *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Berlin: Bericht des Wissenschaftsrates.
- Wollmann, H. (2002): Verwaltungspolitik und Evaluierung. Ansätze, Phasen und Beispiele im Ausland und in Deutschland. In: *Zeitschrift für Evaluation*, 1(1), S. 75-99.
- Wottawa, H. (2001): Qualitätsmanagement durch Zielvereinbarung. In: Spiel, Ch. (Hg.): *Evaluation universitärer Lehre*. Münster: Waxmann, S. 151-164.