

Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität: Entwicklung, Begründung und Überprüfung

Heiner Rindermann

Immer mehr studentische Hochschulgruppen, Fachbereiche und universitäre Institutionen beschäftigen sich seit Beginn der 90er Jahre mit Lehrevaluationen. Mit unterschiedlicher Zielsetzung verwenden sie meist selbstkonstruierte Fragebögen, die verschiedene Aspekte des Lehrgeschehens erfassen sollen. Die Instrumententwicklung orientierte sich jedoch oft nicht an den Kriterien wissenschaftlicher Testkonstruktion. Zudem wird der Veranstaltungskritik eine fehlende stringente theoretische Basis vorgeworfen. Deshalb soll hier ein multifaktorielles hochschuldidaktisches Konzept vorgestellt werden, welches universitäre Lehre als Ergebnis von Bedingungsvariablen auf seiten des Dozenten, der Studierenden und der Rahmenbedingungen auffaßt. Das *Handeln des Dozenten* wird dabei beispielsweise in didaktische und soziale Komponenten unterschieden, *studentisches Handeln* in Veranstaltungsbeteiligung und Arbeitshaltung, *Rahmenbedingungen* in Veranstaltungstyp und -themen. Diese Bedingungsvariablen sollten in einem Inventar zur Bestimmung der Lehrqualität thematisiert werden. Zusätzlich empfiehlt es sich, als Produktvariablen der *Lehreffektivität* Lernerfolg und eine allgemeine Kursqualität zu erheben. Dieses Konzept wird auf seine theoretische Stringenz und empirische Bewährung überprüft.

1 Generierung von Skalen zur Veranstaltungskritik: verschiedene Ansätze

Infolge des öffentlichen Interesses an der Qualität der Lehre – man denke an die Artikel von Spiegel, Stern, Focus, Profil, Forbes – und als Antwort auf ministerielle Erlasse oder gesetzliche Vorgaben, wie z.B. in Nordrhein-Westfalen oder in Österreich, sind in den letzten Jahren an deutschsprachigen Universitäten Instrumente zur Beurteilung der Lehre entwickelt worden (Beispiele: *Daniel 1994; Diehl 1994; Reischmann 1995; Spiel/Gössler 1995; Todt/Götz 1997*; Dokumentation von *Reissert 1992*, und *Hage 1996*). Nur selten stellen die Instrumente Fortentwicklungen älterer überprüfter Verfahren dar (z.B. *Diehl/Kohr 1977; Müller-Wolf 1977*), in der Regel orientierten sich neue Verfahren mehr an Anforderungen der Praxis und weniger an theoretischen Modellen (positiv dazu: *Heger 1991; Hage 1996*; kritisch: *Astleitner/Krumm 1996; Nigsch 1994; Schmidt et al. 1995*).

Unter den Itemgenerierungsverfahren für Fragebögen lassen sich prinzipiell folgende Methoden unterscheiden:

1. **Befragung:** Offene Befragung von Studierenden, Dozenten oder Experten (Hochschuldidaktiker, Ausbilder von Lehrern oder Erwachsenenbildnern) nach relevanten Kriterien guter Lehre. Aus den eruierten allgemeinen oder fach- und veranstaltungsspezifischen Kriterien lassen sich Items formulieren und ein Fragebogen zusammenstellen.
2. **Synkretistisch:** Bislang vorhandene Verfahren und Ergebnisse der Lehrevaluationsforschung dienen der Auswahl und Kombination geeigneter Items und Skalen.
3. **Gremienbeschluß:** Hochschulgremien (Fakultätsrat, Fachschaften etc.) handeln intern oder in Absprache mit anderen Gruppen und Personen wie Dekan, Direktor, Mittelbauvertretungen, Frauenbeauftragte etc. einen Fragebogen und dessen Items aus.
4. **Theoretisch:** Die Itemgenerierung basiert auf schon existierenden oder für die Fragebogenkonstruktion entwickelten Theorien und Lehrzielkonzeptionen.
5. **Autor:** In Eigenregie entwickelt eine Person ein Instrument anhand impliziter oder expliziter Annahmen.

Jedes der skizzierten Verfahren zeichnet sich durch spezifische Vor- und Nachteile aus. Während im **Befragungs-Modell** die Generierungskriterien explizit sind (vgl. *Rindermann 1996b*), bleiben die letztendlichen Auswahlverfahren

durch einen oder mehrere Autoren oft unklar, weil bei Befragungen zu viele und nicht *itemfertige* Nennungen anfallen, die noch zu Items destilliert werden müssen. Hierzu müßte wie im **synkretistischen Modell** ein strukturierendes Konzept herangezogen werden – und sei es nur eine Kombination nach Ähnlichkeit und anschließende Auszählung –, um aus der Fülle der Anregungen die relevantesten Aspekte herausfinden zu können. Vorteil des ersten Verfahrens ist die Nähe zur Hochschulpraxis und zu den Instrument-Nutzern, beim zweiten – falls man sich auf überprüfte Verfahren verläßt – die hohe Wahrscheinlichkeit, auf inhaltlich bedeutsame wie auch empirisch bewährte Items zu stoßen. Das synkretistische Verfahren zeichnet sich durch den geringen Aufwand aus (vgl. Berk 1979), allerdings sind die Fragebögen nicht immer direkt übertragbar (US-amerikanische Bögen auf deutsche Verhältnisse, Post-68er-Entwicklungen¹ auf die heutige Hochschulsituation, verschiedene Studienfächer). In **Gremien** entwickelte Instrumente stoßen in der Praxis auf große Akzeptanz, wenn sich alle relevanten Entscheidungs- und Nutzergruppen im fertigen Inventar wiederfinden können. Allerdings kann das Ergebnis eines solchen Verfahrens ein unstrukturierter, weder an theoretischen noch empirischen Kriterien orientierter Kompromißbogen sein (vgl. Marsh/Roche 1997). Im Extremfall sind die Ergebnisse unverwendbar, weil alle Items, die ein kritisches Ergebnis zu Tage fördern mögen, im Aushandlungsverfahren eliminiert wurden – übrig bleiben dann nur äußerliche Aspekte des Kurses wie Anzahl der Ausfälle und Länge der Referate. Auch sind die Auswahlkriterien unklar, da jede Gruppe ihre eigenen einbringt und womöglich als einziges Kriterium das der Durchsetzungsfähigkeit übrig bleibt. Anhand **theoretischer Modelle** erstellte Instrumente halten den Vorteil eines in sich geschlossenen Konzepts bereit, darüber hinaus verfügen sie über einen höheren wissenschaftlichen Verteidigungswert. Aus grundlegenden theoretischen Annahmen und Lehrzielkonzeptionen werden einzelne, in einem Fragebogen einsetzbare Items abgeleitet, die inhaltlich durch die Theorien begründet werden. Der theorienbasierte Fragebogen eignet sich jedoch nur dann für den Einsatz, falls das Konzept dem Hochschulunterricht gerecht wird, für den das Instrument entwickelt wurde. Die Eignung des Instruments steht und fällt mit der Qualität und empirischen Absicherung der zugrundeliegenden Theorien sowie den daran ansetzenden Itemableitungsverfahren. Die Gefahr, *sich zu verirren*, ist bei einer theoretischen Orientierung größer als unter den vorher genannten Generierungsmethoden. Diese Gefahr trifft im übrigen auch auf das letzte Modell (**Eigenregie**) zu. Für das Autorenverfahren sprechen Unkompliziertheit und Freiheitsgrade des Konstruktionsprozesses, dagegen die Intransparenz der Kriterien und eine fehlende Orientierung an theoretischen Modellen oder empirischen Befragungen.

1 Beispiel: Müller-Wolf (1977) mit Skalen *demokratisches Lehrverhalten (vs. autoritäres)* oder *ideologiekritisch-progressive wissenschaftstheoretische Haltung*.

In der Praxis der Lehrinventarkonstruktion überwiegen jedoch Mischverfahren. Beispielsweise beginnt man mit Befragungen von Studierenden und Lehrenden (1.), nimmt Items aus anderen Fragebögen hinzu (2.), wählt aus subjektiven Annahmen die relevantesten Items aus (5.), zieht zwei oder drei theoretische Modelle zur Ergänzung oder Strukturierung des Itempools heran (4.) und handelt am Ende in Hochschulgremien unter Veränderung oder Ergänzung einzelner Fragen das endgültige Instrument aus (3.). Vorteil eines solchen Verfahrens ist, daß nichts völlig falsch gemacht werden kann, weil die einzelnen Schritte einander die größten Fehler korrigieren. Allerdings könnte je nach Schwäche eines strukturierenden Konzepts oder je nach Stärke spezifischer Interessengruppen in den Gremien ein eventuell an Überlänge leidendes *Mixinstrument* ohne klare Struktur das Ergebnis sein. Ein solcher Konstruktionsweg liegt jedoch einem optimalen Vorgehen dann nahe, wenn Studenten- und Dozentenbefragungen, Items anderer Bögen und Ergebnisse der Lehrevaluationsforschung sowie der Instruktionspsychologie als heuristische Quelle für Skalen innerhalb eines multifaktoriellen Modells der Lehrqualität dienen (vgl. Abschnitt 3).

2 Theoretische Modelle in der Lehrevaluationsforschung

Unter Theorien werden Gebilde begrifflicher Art in Form von wenn-dann-Aussagen verstanden (*Opp 1976, S. 73 ff.*), die untereinander durch Ableitungsbeziehungen verbunden sind. Davon zu unterscheiden sind präskriptive Sätze, die wünschenswerte Aspekte und Ziele der Lehre benennen (Verständlichkeit, Interaktivität, Einstellungsänderung etc.). Meist werden nur dem gängigen oder als gängig angenommenen Hochschulunterricht entgegengesetzte Lehrziele explizit präskriptiv formuliert: statt Vermittlung deklarativen Wissens auch Vermittlung von Verständnis und prozeduralem Wissen, Förderung sozialer Kompetenzen, Einstellungs- und Persönlichkeitsänderungen (z.B. *Standhardt/Löhmer 1995*). Umfassender als einzelne Theorien oder Wertungen sind Konzeptionen und Modelle, die den Hochschulunterricht auf höherer Generalitätsebene zu beschreiben versuchen: Beispielsweise stellt das multifaktorielle Modell der Lehrqualität verschiedene Bedingungsvariablen wie Dozenten- und Studentenhandeln als Determinanten des Lehrerfolgs heraus. Modelle enthalten meist Annahmen über komplexe (multifaktorielle und multivariate) Kausalbeziehungen zwischen Variablensätzen. *Konzeption* ist begrifflich weitgefaßter, darunter können begriffliche Definitionen, Theorien, Modellannahmen in Form von Determinationsgefügen oder Grundannahmen und ganze Forschungsprogramme (Paradigmen) fallen. Als Konzeption wäre z.B. die Annahme der *Interaktivität der Lehre* zu bezeichnen: In dieser wird Lehre nicht als monokausaler Prozeß (guter Dozent führt zu Lehrerfolg o.ä.), sondern als reziproker Prozeß mehrerer Determinanten aufgefaßt.

Die meisten Theorien enthalten verdeckt normative Bestandteile. Beispielsweise ist der Hypothese *Wenn der Dozent klar spricht, dann kann man ihn verstehen und etwas lernen* indirekt entnehmbar, daß *verstehen* und *lernen* und damit auch *klar sprechen* als wünschenswerte Aspekte des Hochschulunterrichts gelten. Solche kryptonormativen Gehalte (Groeben 1986, S. 417) sind nicht nur theoretischen Grundlagen, sondern auch Fragebögen und deren Items entnehmbar. Die in Lehrinventaren behandelten Aspekte formulieren implizit Annahmen, wie der Hochschulunterricht aussehen sollte (z.B. gut strukturiert, motivierender Dozent; ausf. Wilson 1988). Da die Items Zielvorstellungen und Anforderungen an die Lehre widerspiegeln, können sie durch ihre Verwendung in Inventaren auf die Lehre rückwirken, während nicht thematisierte Aspekte möglicherweise ungerechtfertigt marginalisiert werden. Solange diese Zielvorstellungen nicht des Status unklarer Vorannahmen entkleidet werden, bleiben sie der wissenschaftlichen Analyse vorenthalten.

Die meisten sich an Theorien orientierenden Generierungsverfahren bedienen sich mehrerer unabhängiger theoretischer Grundannahmen und kombinieren diese. Drei Beispiele sollen hier besprochen werden: das von Westermann et al. (1993, 1996) und von Kramis (1990) sowie das Modell von Helmke (1996), das bislang aber noch nicht als Grundlage für Lehrinventare herangezogen wurde.

2.1 Der Ansatz von Westermann et alii

Die Forscher(innen)gruppe um Westermann (*et al.* 1993; *et al.* 1996; *Spies/Westermann/Heise/Schiffner* 1996) orientierte sich bei der Konstruktion eines Fragebogens zur Erhebung der Zufriedenheit mit der Lehre an verschiedenen Theorien:

- Konzept der Arbeitszufriedenheit: Die Einstellung einer Person zu ihrer Arbeit ist Folge einer Linearkombination der Valenz der Objektattribute und der Stärke der subjektiven Überzeugungen, daß die Objekte diese Attribute besitzen.
- Person-Environment-Fit-Theorie: Die Zufriedenheit einer Person ist abhängig von der subjektiven Passung zwischen ihren Bedürfnissen, Einstellungen, Fähigkeiten und den Merkmalen der studienbezogenen Umwelt.
- Zusätzlich werden für Teilaspekte humanistische Bedürfnishierarchien, Lernzieltaxonomien sowie die Job-Characteristics-Theorie und Handlungsphasentheorie herangezogen.

Die Fragebogenentwicklung richtete sich an den Konzepten der (Arbeits-/Studien-) Zufriedenheit und der Person-Umwelt-Passung aus. Die Zufriedenheit gilt als interaktives Produkt von Personen- und Umweltmerkmalen.¹ Zwar wurden die theoretischen Grundlagen der Skalenkonstruktion für Lehr- und Lehrveranstaltungsinventare dargelegt. Allerdings bleibt unklar, inwieweit diese Theorien für den tatsächlichen Skalengenerierungsprozeß ausschlaggebend waren. Die späteren Items und Dimensionen des Lehrveranstaltungsinventars lassen sich zumindest nicht den einzelnen theoretischen Annahmen zuordnen. Da die Autoren zusätzlich andere, bereits existierende Fragebögen als Generierungsquelle erwähnen, ist es schwierig zu entscheiden, welche Rolle die theoretische Konzeption und welche der bereits vorhandene Itempool gespielt hat. Die Theorien dienen eher als konzeptioneller Rahmen für die Erhebung der Zufriedenheit mit der Lehre in Form von Kursen und veranstaltungübergreifenden Studienmerkmalen.

Aber an dieser Konzeption ist grundsätzlich Kritik zu üben: Nicht mehr die Qualität der Lehre stellt bei Westermann et alii das zu analysierende Merkmal dar, sondern die Zufriedenheit der Studierenden mit der Lehre. Es fehlen didaktische Theorien, die Hypothesen über einzelne Lehrformen und Unterrichtstechniken in Zusammenhang mit verschiedenen Lehrerfolgskriterien formulieren. Streng genommen geht es nicht mehr um gute, schlechte oder optimale Lehre, wie sich diese beschreiben, erklären und verbessern läßt, sondern um die (subjektive) Zufriedenheit der Studierenden (*Da nur der subjektive P-E-Fit interessiert, wurde auf eine Erhebung von objektiven Merkmalen verzichtet.*; Spies et al. 1996, S. 382). Objektmerkmale, wie etwa das Unterrichtsgeschehen, gehen nur noch als studentische Attributvalenzen ein, nicht mehr als Merkmale zur Beschreibung und Bestimmung von Lehrqualität. Konsequenz durchgeführt würde dies bedeuten, daß

- die Erfassung des Lehrgeschehens nur noch bedeutsam ist in Form von Rezeptionsmerkmalen als Indikator der Zufriedenheit,

¹ Allerdings stützen diejenigen empirischen Resultate, die zur Thematik *Passung* aus der Lehrevaluationsforschung vorliegen, nicht diese Konzeption: Die von Studierenden *wahrgenommene* Passung zwischen Dozenten- und Studentenmerkmalen korreliert nur schwach mit der Veranstaltungsbeurteilung ($r = .23$; Abrami/Mizener 1985, 1983; s.a. Tollefson/Chen/Kleinsasser 1989). Die *tatsächliche* Ähnlichkeit in Einstellungen (Dozenten- und Studentenselbsteinschätzungen) korreliert nur zu $r = .06$ bis $r = .14$ (Good/Good 1973a, 1973b). Passung in diesem Sinne kann deshalb kaum Lehrevaluationsresultate beeinflussen. Die Einschätzung der Persönlichkeit des Dozenten durch Studierende hängt zudem von der Beurteilung der Lehrkompetenz des Dozenten ab (Abrami et al. 1982, S. 123). Eventuell kann für die abhängige Variable *Zufriedenheit* ein höherer Varianzanteil durch Passung in anderen Merkmalen der Lehre wie in Lerneinstellung und Veranstaltungstypus erklärt werden.

- Rückmeldungen an den Dozenten von geringer Informationshaltigkeit und praktischem Nutzen sind (Lehre wird nicht beschrieben),
- Aussagen über die Qualität der Lehre weder möglich noch erstrebenswert sind und
- die empirische Validität studentischer Bewertungen nicht mehr analysierbar ist, da Außenkriterien für die Zufriedenheit fehlen und konzeptionell irrelevant sind. Gleichzeitig entledigt man sich hiermit einiger brisanter Fragestellungen, z.B. die der Urteilskompetenz der Studenten sowie die der Validität, Vergleichbarkeit und Verwendung der Resultate, die im Zufriedenheitsmodell nicht mehr analysierbar sind.

Allerdings wird dieser konzeptionelle Rahmen von Westermann et alii selbst nicht konsequent durchgehalten, bei näherer Betrachtung finden sich (1993) in dem *Instrument zur Beurteilung von Lehrveranstaltungen* (sic!) viele Items (ca. 28 von 55), die sich direkt auf das Lehrverhalten des Dozenten beziehen (Beispiel: *LV18: Der/Die Lehrende gestaltet Tafelbild oder Folien leserlich und übersichtlich*). Diese Skalen werden nicht mehr durch die Zufriedenheits-Konzeption abgedeckt. Tatsächlich liegt der Mehrzahl der Items eine implizite didaktische Theorie mit dem zumindest partiell vorhandenen Anspruch zugrunde, das Lehrgeschehen und nicht nur die Zufriedenheit der Studierenden zu beschreiben.

Westermanns theoretischer Ansatz kann deshalb nicht völlig überzeugen: Zunächst bleibt unklar, inwieweit Theorien eine itemgenerierende Funktion bei der Lehrskalenkonstruktion aufwiesen, das Konzept der Zufriedenheit/Passung wird nicht stringent durchgehalten und schließlich stellt das Zufriedenheits-Paradigma ein theoretisch wie auch praktisch unzureichendes Konzept zur Beschreibung, Erklärung und Verbesserung universitärer Lehre dar. Als ein Kriterium des Lehrerfolges (neben anderen wie dem Lerngewinn) wäre die studentische Veranstaltungszufriedenheit jedoch eine praktisch und wissenschaftlich bedeutsame Größe: Läßt sich diese eher durch dozentenbezogene Skalen oder durch Rahmenbedingungen wie Vorinteresse und Anforderungen prognostizieren?

2.3 Der Ansatz von Kramis

Kramis' Konzeption (1990) zeichnet sich aus durch die Orientierung an drei postulierten *grundlegenden Gütekriterien für Unterricht*¹ (S. 279), die er in Anlehnung an theoretische didaktische Modelle sowie instruktionspsychologische Metaanalysen konstruiert hat. Diese sind:

- Bedeutsamkeit der gewählten Unterrichtsinhalte und Ziele,
- Effizienz der gewählten Lernorganisation, Lernaktivitäten und Medien und
- gutes Lernklima.

Den genannten fundamentalen und notwendigen Gütekriterien ordnet Kramis 36 didaktische Prinzipien zu und leitet von diesen wiederum Items ab. Faktorenanalytisch bestätigen sich die drei Unterrichts-Gütekriterien, die Skalen sind homogen (Trennschärfe und Cronbach- α). Da die Analysen nur auf einer Stichprobe von $N = 110$ Lehramtsstudierenden basieren (sieben Veranstaltungen), sind für weitergehende Schlußfolgerungen Replikationen an größeren und anders zusammengesetzten Stichproben abzuwarten.

Bei Kramis sind einzelne Items den didaktischen Theorien genau zuordenbar. Die empirische Abstützung ist jedoch sowohl auf seiten der Stichprobe (relativ klein und nur Lehramtsstudium) als auch auf seiten der verwendeten statistischen Verfahren (Stabilitäts- und Validitätsanalysen stehen aus) schwach. Seine Konzeption stellt eher ein präskriptives Modell (*Leitbilder*) für die Lehre und Lehrende dar, als ein empirisch beschreibendes und Kausalbeziehungen aufzeigendes Bedingungsmodell guten Unterrichts (vgl. Mayr 1997, S. 244 ff.). Unterschieden werden zwar Inhalte, Lehrstrategien (die die Effizienz sichern sollen) und soziale Komponenten, allerdings wird nicht präzise zwischen Rahmenbedingungen (Skala *Bedeutsamkeit* umfaßt sowohl vorgegebene Themen als auch Unterrichtsformen) und dem Dozenten- oder Studentenhandeln differenziert. Das Vorinteresse am Thema, die studentische Arbeitshaltung, die Höhe der Anforderungen und der selbsteingeschätzte Lerngewinn bleiben unthematisiert. Exemplarisch bleibt bei Kramis sicherlich die enge Verzahnung von Theorie, Fragebogenkonstruktion und Lehrpraxis.

Beiden Ansätzen gemeinsam ist die Berücksichtigung interaktiver Komponenten der Lehre. Westermann konzeptualisiert diese Interaktion in Form einer Passung von studentischen und Umweltmerkmalen als Bedingung für Zufriedenheit mit

¹ Kramis spricht von „Unterricht“ im tertiären und sekundären Bildungssektor. Im vorliegenden Beitrag wird „Lehre“ und „Unterricht“ synonym gebraucht (vgl. a. den einheitlichen englischen Begriff „instruction“).

der Lehre, Kramis als die notwendige gleichzeitige Erfüllung dreier Grundbedingungen für guten Unterricht.

2.4 Helmke: Angebots-Nutzungs-Modell des Studienerfolgs

Helmke (1996) stellt ein allgemeines Modell der Studienbedingungen und des Studienerfolgs vor. Er unterscheidet hierbei Lehrkontext, Lehrangebot, Lehrperson, studentische Determinanten und individuelle Lernvoraussetzungen als Bedingungen der nominalen und aktiven Nutzung des Lehrangebots. Auf der Effektseite wird zwischen kognitiven und nichtkognitiven Wirkungen und langfristigen Effekten wie Berufserfolg und Lebensbewältigung differenziert. Diese stehen in einem Wechselwirkungsprozeß zur Perzeption und Evaluation.

Bislang wurde dieses Modell noch nicht als Grundlage für Lehrevaluationen oder Skalenkonstruktionen herangezogen. Für ein Evaluationsmodell der universitären Lehre wäre das Angebots-Nutzungs-Modell zu allgemein, hier haben der Wissenschaftsrat (1996) und Webler (1996a, b) detailliertere Programme vorgelegt. Es könnte aber als theoretische Grundlage solcher Programme dienen. Modelle, die viele mögliche Bedingungsfaktoren ohne Gewichtung einzelner Pfade in Beziehung setzen, können heuristisch zur Formulierung einzelner Hypothesen genutzt werden. Widerlegbar oder empirisch prüfbar sind sie aber nicht. Die Auswahl der Faktoren und Pfade bei Helmke ist fragwürdig: Warum wirkt in dem veranstaltungübergreifenden Modell die Nutzung des Lehrangebots nicht auf die individuellen Lernvoraussetzungen zurück, oder sollte die Lehre etwa keine Lerneffekte haben? Wodurch wird begründet, daß die Attraktivität einer Lehrperson als Bedingungsvariable aufgeführt wird, nicht aber Charakteristika der Veranstaltungs- und Studienthemen (Anforderungsniveau, Relevanz, etc.)? Warum fehlen Studienabbruch und -dauer als Studienerfolgskriterien?

3 Synthetisches Verfahren der Skalenkonstruktion

Ein synthetisches Konstruktionsverfahren berücksichtigt verschiedene Quellen bei der Itementwicklung, zur Generierung und anschließenden Überprüfung wird ein theoretisch-empirisches Vorgehen gewählt. Ein multifaktorielles Modell der Lehrveranstaltungsqualität bildet den konzeptionellen Rahmen des Entwicklungsverfahrens, ein Modell, welches die Qualität der Lehre (und nicht die Zufriedenheit!) als Ergebnis des Zusammenspiels mehrerer, sich gegenseitig beeinflussender Faktoren ansieht: des Handelns des Dozenten, des Handelns der Studierenden und der Wirkungen der Rahmenbedingungen. Die Vernachlässigung einer dieser drei Komponenten würde weder der Beschreibung des Lehrgeschehens noch der Erklärung des Lehrerfolges gerecht.

Ergebnisse der Befragungen von Studierenden und Lehrenden sowie Auswertungen bisheriger Inventare und der Lehr-Lern-Forschung sollten im Rahmen eines solchen Modells als heuristische Quellen für einzelne Skalen aufgenommen werden. Am Schluß sollte die Entscheidung über den Einsatz des entwickelten Instruments durch universitäre Gremien stehen.¹

Die empirische Komponente in der Generierungsphase besteht in der Befragung Lehrender und Lernender und in der empirischen Abgesichertheit der herangezogenen Theorien. Auch die Überprüfung des entwickelten Verfahrens unterliegt einem kombiniert theoretisch-empirischen Vorgehen. Statistische Analysen können die Urteilsübereinstimmung, Reliabilität, faktorielle Struktur, Generalisierbarkeit, Validität und in Meßwiederholungsdesigns den Nutzen des Einsatzes untersuchen. Befragungen der Instrument-Rezipienten nach Relevanz der Items sowie Geeignetheit der Rückmeldung sichern die Praxistauglichkeit. Dies sollte jedoch auf der Basis eines multifaktoriellen Modells der Lehrqualität, einer anzustrebenden Differenziertheit des Feedbacks und der Multidimensionalität der Lehre geschehen.

4 Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität

4.1 Die Struktur des Modells

Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität basiert auf verschiedenen empirischen Studien und unterschiedlichen theoretischen Annahmen (vgl. *Rindermann/Amelang 1994a, b; Rindermann 1996a, b, 1997a*): Zu diesen zählen Befragungen von Studierenden und Lehrenden nach Kriterien zur Beurteilung der Lehre, Auswertungen bisheriger Verfahren, Orientierung an Lehrzielen sowie Berücksichtigung von unterrichtsdidaktischen Theorien. Zur empirischen Überprüfung ließen sich Resultate statistischer Verfahren (Faktoren-, Reliabilitäts- und Validitätsanalysen etc.) nutzen, die an Stichproben unterschiedlicher Universitäten und Studiengänge gewonnen wurden.

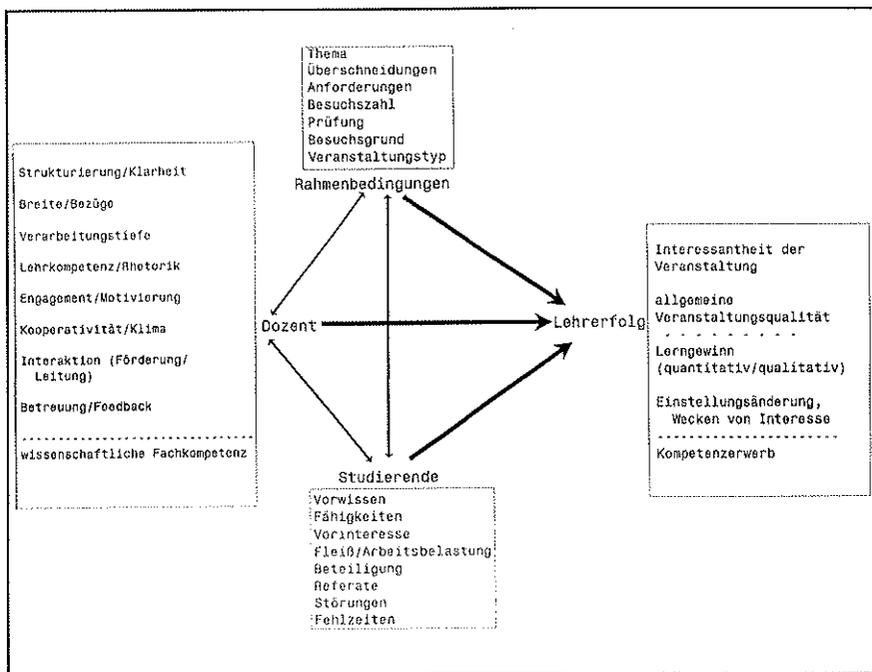
Lehre stellt eine intentionale interpersonale Aktivität mit dem Ziel einer Modifikation von kognitiven und nichtkognitiven Strukturen dar. Zum Gelingen dieses interaktiven Prozesses müssen die Dozenten und Studenten entsprechende Vor-

¹ In der Evaluationsforschung hat sich die Trennung von Auftraggeber und durchführenden Wissenschaftlern bewährt (*Wottawa/Thierau 1990, S. 25 ff.*). Die Lehrevaluation(-forschung) steht jedoch wie die Hochschulforschung insgesamt vor dem Problem der fehlenden Differenzierung von Lebenswelt und Wissenschaft und ihrer unterschiedlichen Rollenerwartungen. Dies hat spezifische wissenschaftstheoretische, -soziologische und -praktische Probleme zur Folge (*Rindermann 1997 d*).

aussetzungen erfüllen. Dieser Prozeß ist in gesellschaftliche, curriculare und institutionelle Rahmenbedingungen eingebettet.

Deshalb werden im Modell die Komponenten *Dozent*, *Studierende*, *Rahmenbedingungen* und *Lehrerfolg* unterschieden (s. Abbildung). Erst das Zusammenspiel eines fachlich, didaktisch und kommunikativ kompetenten Dozenten, vorgebildeter, interessierter und engagierter Studenten sowie günstiger Rahmenbedingungen in Form von ausreichend Zeit, Platz, adäquater Anforderungshöhe und inhaltsangemessenen Veranstaltungsformen (vgl. *Sturm 1994*), kann Lehrerfolg mit hoher Wahrscheinlichkeit erzielen.

Abbildung: Multifaktorielles Modell der Lehrveranstaltungsqualität



Diese drei Hauptfaktoren werden als Bedingungs- und Prozeß-Variablen mit der Produkt-Variable Lehrerfolg in Beziehung gesetzt (vgl. *Dunkin/Barnes 1986*; *Meredith 1975*). Sie bewirken den Erfolg der Lehre in Form guter und günstig beurteilter Veranstaltungen und in Form intendierter Veränderungen auf seiten der Veranstaltungsteilnehmer. Einzelne Skalen werden zu Faktoren in einem

Determinationsgefüge universitärer Lehre gebündelt. Allerdings wirken Bedingungsfaktoren nicht isoliert voneinander auf den Lehrerfolg (vgl. *Weinert/Schrader/Helmke 1989*). Beispielsweise kann der Dozent die Studierenden motivieren und dadurch deren Beteiligung erhöhen. Ebenso steigert eine veranstaltungsbezogene Prüfung den Fleiß der Studenten. Die didaktische, fachliche und soziale Kompetenz der Dozenten wirkt auf die Besuchszahl. Umgekehrt beeinträchtigen schlechte Rahmenbedingungen wie große Besuchszahlen die Betreuung, Störungen behindern die Strukturierung und fehlendes Vorwissen oder geringe studentische Kompetenzen lassen durchschnittliche Anforderungen als hohe erscheinen. Die Kausalbeziehungen verlaufen nicht nur zwischen den Hauptfaktoren und den Effektivvariablen, sondern auch zwischen und innerhalb der einzelnen Faktoren. Beispielsweise kann sich das Vorwissen der Studenten auf das Auftreten von Störungen auswirken (*Weinert et al. 1989, S. 907*).

Der Lehrerfolg wird als zeitlich nachgeschaltete Produktvariable aufgefaßt. Allerdings wirken die Erwartungen der Studierenden über den möglichen Lehrerfolg Ende des Semesters und die während der Vorlesungszeit vorgenommenen Bilanzierungen über bisherige Lehrqualität und Lernerfolg auf die Bedingungsvariablen zurück, beispielsweise auf die Besuchszahl.

Im **Gegensatz zu Westermann und Kramis** wird nicht von Interaktivität in Form von Passung oder gleichzeitiger Erfüllung fundamentaler Kriterien als Bedingung von Zufriedenheit oder guten Unterrichts ausgegangen, sondern in Form von wechselseitiger Beeinflussung und Bedingung dreier Hauptfaktoren guter Lehre und Lehreffektivität. Neben unidirektionalen Effekten sind ebenso auch kompensatorische und wechselseitige, sich gegenseitig fördernde oder hemmende Beeinflussungen zu konstatieren. So hat gleiches Dozentenverhalten vor verschiedenen Studentengruppen oder bei unterschiedlichen Veranstaltungsformen andere Auswirkungen (aptitude-treatment-interaction). Lehrkompetenz zeichnet sich deshalb auch durch adaptives Einsetzen didaktischer Methoden aus (*Klauer 1985; McKeachie 1997*).

Dies hat zur Konsequenz, daß der Veranstaltungserfolg statistisch ausgedrückt nie zu 100% durch den Einfluß des Dozenten aufgeklärt werden kann, oder daß die Konstanz seines Lehrverhaltens über verschiedene Veranstaltungen, Themen und Studentengenerationen nie $r = 1$ erreicht, sondern in einer Mittelposition zu vermuten ist – genaueres ergeben empirische Studien (z.B. *Frey 1978; Rindermann 1997b*).

Im Vergleich zu dem von Helmke (1996) vorgestellten Angebots-Nutzungs-Modell des Studienerfolgs fokussiert das Münchner Modell auf die Qualität und Effektivität der Lehre in Form von Lehrveranstaltungen. Allgemeine Rahmenbedingungen wie gesellschaftliche und materielle Faktoren (Zulassungssystem, Finanzierung, Alter der Studierenden, Vorbildung wie durch eine Lehre oder ein vorheriges Studienfach) erwiesen sich in bisherigen Analysen der Lehrevaluationsforschung als nicht prognosekräftig für individuelle oder aggregierte Lehrevaluationsresultate (Rindermann, 1997 a). Diese sollten nur in ein allgemeines Modell der Studienbedingungen und des Studienerfolges wie von Helmke aufgenommen werden. In Analogie zur *Veranstaltungszufriedenheit* im Ansatz von Westermann et al. (1993, 1996) wird ein generelles Erfolgskriterium in der Form der *Veranstaltungsqualität* implementiert. Als Bedingungsvariablen dienen jedoch wie bei Kramis (1990) Dimensionen, die das Lehrgeschehen thematisieren. Bei den Dozentenskalen werden hierzu didaktische Skalen verwendet.

Das Münchner Modell stützt sich vor allem auf Erfahrungen mit dem Heidelberger Inventar zur Lehrveranstaltungsevaluation (HILVE, Rindermann/Amelang 1994b) und stellt eine Weiterentwicklung dieser Konzeption dar (Rindermann 1997a). Modifizierte Formulierungen unterschieden stärker zwischen den drei Hauptfaktoren Dozent-Studenten-Rahmenbedingungen, auch zählen die Skalen Interaktion, Betreuung/Feedback, Vorwissen, Störungen und Fehlzeiten nun zum Grundkanon des revidierten Instruments.

Dem Ziel, die wesentlichen Bedingungen der Lehre in verschiedenen Studienrichtungen und Veranstaltungsformen zu formulieren, folgt nun die Beschreibung der Skalen und deren Begründung durch Theorien und Resultate der Evaluations- und Lehr-Lern-Forschung.

4.2 Bedingungsfaktor Dozent (Handlungen, Kompetenzen und Einstellungen)

Strukturierung/Klarheit

Ein für die Teilnehmer nachvollziehbarer Aufbau zählt zu den zentralen Merkmalen guten Unterrichts (Bolton et al. 1976; Feldman 1976; Weinert et al. 1989; Winteler/Schmolck 1983). Eine klare Struktur fördert den Verständnisprozeß, erleichtert die kognitive Strukturierung der zu lernenden Inhalte und deren späteren Abruf. Die Skala Struktur korreliert in HILVE-Studien mit allen Lehrerfolgsskalen zu $r \geq .50$. Land (1981) sowie Cruickshank und Kennedy (1986) wiesen einen positiven Zusammenhang zwischen klarer Strukturierung und objektiverer studentischer Leistung in der Höhe von $r = .40 - .60$ nach (vgl. a. Weinert 1996). Aus Studenten- und Dozentsicht in verschiedenen Hochschul-Systemen und Unterrichtstraditionen stellt die Struktur das wichtigste

Kriterium zur Beurteilung der Lehre dar (*Rindermann 1996a, S. 43; Shatz/Best 1986*).

Breite/Bezüge

Beispiele und Darlegungen, die den zu vermittelnden Stoff in verschiedene Kontexte einbetten, erleichtern wie die Strukturierung den Lernprozeß, die Aufnahme, Verarbeitung, Abrufung und aktive Anwendung von Wissen (vgl. *Renkl 1996a, b*). Zu diesen Kontexten zählen Praxisbezüge, Beispiele aus anderen Bereichen und Erläuterungen, welche die Relevanz und Bedeutung der Inhalte aufzeigen.¹

Verarbeitungstiefe

Lehrziel ist nicht nur die Vermittlung fertigen Wissens, sondern auch die Erreichung höherer kognitiver Ziele (Verständnis, Anwendung, Bewertung). Der Stoff wird als Problem oder Aufgabe, die des eigenen Mit- und Durchdenkens bedarf, dargestellt. Passive Stoffaufnahme beschleunigt Vergessensprozesse, während aktive Verarbeitung die Enkodierung und Speicherung auf tieferem Niveau ermöglicht. Zeigt der Dozent z.B. verschiedene Perspektiven auf und nutzt verschiedene Kontexte, fördert er den Verständnisprozeß und die Wissensflexibilisierung. Zusätzlich enthält diese Bedingungsvariable noch eine interaktive Komponente: Studierende werden nicht nur als Rezipienten, sondern als am Erkenntnisprozeß Beteiligte betrachtet (*Wilson 1988*).

Lehrkompetenz/Rhetorik

Unter Lehrkompetenz wird weniger eine einzelne Unterrichtstechnik, sondern eine allgemeine Fähigkeit zu lehren verstanden. Diese äußert sich etwa in guter Vorbereitung, in verständlicher und anregender Sprache (vgl. *Döring 1995, S. 66 ff.*) oder in der Fähigkeit, Kompliziertes erklären zu können. In angloamerikanischen Verfahren dienen *teaching effectiveness* und *overall instructor* oft als übergeordnete Kategorie des auf den Dozenten attribuierbaren Varianzanteils der

¹ Oft wurde kritisch angemerkt, daß in Fächern wie Mathematik, Geschichte oder Philosophie Kontextualisierung oder Praxisbezug nicht möglich seien. In Mathematik oder Statistik sind jedoch Beispielaufgaben üblich, in Geschichte können Beziehungen zwischen verschiedenen Zeitepochen bis zur Gegenwart hergestellt werden, in der Philosophie sind dem Alltag entnommene Probleme der Erkenntnis, der Moral oder der Ästhetik oft Ausgangspunkt philosophischer Erörterungen. Ähnliches trifft auch für Sprach- und Naturwissenschaften zu.

Lehre. In diesem Sinne könnten alle unter dem Dozenten subsumierbaren Bedingungsfaktoren unter Lehrkompetenz gebündelt werden. Diese wird nach Marques et al. (1979) über verschiedene Fachbereiche und Urteilergruppen weitgehend invariant konzeptionalisiert (z.B. Wissensvermittlung, Stimulation von Interesse). Hildebrand (1973) kategorisierte Charakteristika von Lehrenden hinsichtlich ihrer Differenzierungsfähigkeit zwischen den besten und schlechtesten Dozenten, die als Indikatoren der Lehrkompetenz aufgefaßt werden können. Als besonders differenzierungsfähig erwiesen sich: *Kontrastiert die Implikationen verschiedener Theorien, ist ein exzellenter Sprecher, macht Schwieriges einfach zu verstehen, zeigt Interesse an der Lehre.*

Engagement/Motivierung

Engagement des Dozenten und Motivierung der Teilnehmer gehören wie die Lehrkompetenz zu den wichtigen Attributen guter Hochschullehrer (Feldman 1976; Klauer 1985; Marques et al. 1979; Marsh/Roche 1992; Müller-Wolf 1977). Desinteresse des Dozenten überträgt sich auf die Teilnehmer einer Veranstaltung, senkt die Aufmerksamkeit und motiviert nicht zu eigener Beschäftigung mit den Themen. Dies beeinträchtigt den Lernprozeß, wie die Korrelation zwischen Engagement und Lerngewinn bei HILVE-Daten verdeutlicht ($r = .57$; Rindermann 1996a, S. 39).

Kooperativität/Klima

Lehre als interpersonale Aktivität erfordert nicht nur fachliche und didaktische Kompetenz, sondern auch soziale Fähigkeiten und Verhaltensweisen wie Freundlichkeit, Kooperativität und Kommunikation (vgl. Moses 1985; Slotnick/Durkovic 1975). Dies wirkt sich positiv auf die studentische Beteiligung und Motivation aus.

Interaktion (Förderung, Leitung)

Studentische Beteiligungen und Diskussionen bedürfen neben der Förderung (vgl. Basler et al. 1995; Skala *Motivation zur aktiven Teilnahme*) auch der Leitung: Durch eine freundliche Atmosphäre und offene wie divergente Fragen (Dunkin/Barnes 1986) werden studentische Beiträge angeregt, die Moderation von Diskussionen verhindert deren thematischen Zerfall. Interaktionsvariablen hängen stärker von Kontextbedingungen der Lehre ab als die Lehrkompetenz (Frey 1978).

Betreuung/Feedback

Studierende benötigen neben Rückmeldungen auf Fragen besonders bei der Erstellung von Referaten oder Hausarbeiten Betreuung und Feedback durch den Dozenten. Feedback durch den Dozenten ist auch in Diskussionen zwischen Teilnehmern wichtig (*Feldman 1976*). Betreuung sichert die Qualität der Beiträge und stützt die Eigeninitiative. Nach Spies et al. (1996) und Multrus (1995) stellen Rückmeldungen und die Kontaktdichte zwischen Dozenten und Studierenden wichtige Determinanten der Studienzufriedenheit und der studentischen Bewertung der Studienqualität dar.

Wissenschaftliche Fachkompetenz

Die Fachkompetenz des Dozenten gehört neben der didaktischen, motivationalen und sozialen Kompetenz zu den zentralen Voraussetzungen universitärer Lehre (vgl. *Moses 1985; Wilson et al. 1973*). Diese kann in der Regel nicht durch Studenten beurteilt werden; wenn sie dies könnten, besuchten sie aufgrund der Redundanz die falsche Veranstaltung (*Marques et al., 1979*). Die Sicherung der fachwissenschaftlichen Qualität stellt eine Aufgabe der Peer-Evaluation dar, ähnlich wie die Gutachtertätigkeit bei zur Publikation eingereichten Manuskripten (s. *Merton/Zuckerman 1985*).

4.3 Bedingungsfaktor Studierende (Kompetenzen, Einstellungen und Handlungen)

Das Handeln des Dozenten stellt jedoch nur einen – wenn auch wahrscheinlich den wichtigsten (*Müller-Wolf 1977, S. 115*) – Bedingungsfaktor des Erfolges universitärer Lehre dar. Hinzu müssen auf studentischer Seite spezifische und allgemeine Lernvoraussetzungen sowie veranstaltungsinterne und -externe Handlungen kommen:

Vorwissen

Das Vorwissen der Studenten ist das Ergebnis vorheriger Bildungserfahrungen durch Schule, Universität und Selbststudium. Verfügen die Studierenden über zu wenig, zu viel oder fehlerhaftes Vorwissen, schadet dies dem Lernprozeß und der Lehreffektivität (*Renkl 1996b; Schmidt et al. 1995*). Bei zu wenig Wissen müssen erst die Grundlagen erarbeitet werden oder die Themen bleiben bei unzureichender Vor- und Nachbereitung durch die Teilnehmer unverständlich. Bei zu viel Vorbildung (z.B. aufgrund von inhaltlichen Überschneidungen mit

anderen Kursen) tritt Langeweile auf, die – wie Überforderung – zu Störungen oder Besuchsabbruch führen kann. Ist dem Dozenten das Vorwissen bekannt, kann er die Inhalte und Formen der Lehre in gewissem Umfang adaptieren.

Fähigkeiten

Der Lehrprozeß kann nur bei adäquat vorhandenen kognitiven Kompetenzen zu günstigen Lernresultaten führen (*Klauer 1985*). Neben diesen Fähigkeiten können auch nichtkognitive Kompetenzen, z.B. Kommunikationskompetenz, Organisationsgeschick oder Frustrationstoleranz, relevant für den Lern- und Studien-erfolg sein. Bislang fand jedoch nur die kognitive Leistungsfähigkeit, meist in Form der Abiturnote oder der allgemeinen Intelligenz, als Prädiktorvariable Eingang (*Rindermann/Oubaid 1998*), sie korrelierte in bisherigen Studien allerdings nicht mit der Beurteilung der Lehre oder des selbst eingeschätzten Lernerfolges. Prüfungsergebnisse lassen sich aber damit sehr wohl prognostizieren (vgl. *Rindermann 1996a, S. 121; Daniel 1995*).

Vorinteresse

Vorinteresse oder Interesse am Thema, d.h. selbstintentionale, positive, wertschätzende Zuwendung, zählt mit Besuchsgrund zu den wichtigsten Bedingungs- und Biasvariablen (*Rindermann 1997 a*). Unter Biasvariablen sind solche Faktoren zu verstehen, die die Beurteilung der Lehre verzerren können, ohne daß sie mit den tatsächlichen Geschehnissen einer Veranstaltung zusammenhängen. Sie führen zu einer unzutreffenden Beschreibung der Lehre, die besonders in den Dozentenskalen zu kontrollieren sind. Das Interesse fördert jedoch auch die Lernmotivation und den Lernprozeß (s. *Klauer 1985; Rheinberg 1996*), die Beteiligung und den Fleiß der Studierenden ($r = .34$, *Rindermann 1996a, S. 110*). In der Unterrichtsforschung wird Interesse am Thema als mögliche, aber nicht notwendige Bedingung des Interesses am Unterricht konzipiert (z.B. *Todt/Händel-Mattes 1990, S. 135*).

Fleiß/Arbeitshaltung/Arbeitsbelastung

Unter Fleiß oder Arbeitsbelastung fallen die Vor- und Nachbereitung einer Veranstaltung außerhalb der Kurssitzung. Diese sind nicht unabhängig von Rahmenbedingungen wie der Existenz einer Prüfung und der Höhe der Anforderungen (*Rindermann 1996a*). Mit anderen Skalen ist Fleiß im HILVE unkorreliert (ähnliche Resultate bei Dozenten- und Fremdurteilen). Auch zur Abiturnote und zum Klausurergebnis (individuelle Rohdaten) bestehen keine

Zusammenhänge. Fleiß in Form individueller Rohdaten innerhalb einer Veranstaltung gibt die persönliche Arbeitshaltung wieder, als Veranstaltungsmittelwert bildet sie eher die von Rahmenbedingungen abhängige Arbeitsbelastung im Kurs ab.

Beteiligung

Ähnlich wie Interaktion und Referate ist die veranstaltungsinterne Beteiligung der Studierenden in Seminaren für den Erfolg der Veranstaltung mitverantwortlich. In Vorlesungen treten mit der Allgemeinbeurteilung Korrelationen um Null auf, in Seminaren wird jedoch $r \approx 50$ erreicht (HILVE-Daten). Die Beteiligung variiert beträchtlich mit Rahmenbedingungen (Veranstaltungstyp, Prüfung) **und** einzelnen didaktischen Skalen (Verarbeitungstiefe; s. *Rindermann 1996a*). Nach Dunkin und Barnes (1986) ist die studentische Partizipation eine Bedingung für die Förderung höherer kognitiver Prozesse wie kritisches Denken. Durch aktive Einbindung und kooperative Prozesse der Wissensaneignung können Lernprozesse effektiver gestaltet werden, die über den Erwerb assoziativer Inhalte hinausgehen (*Seel 1997*).

Referate

Studentische Referate bilden in vielen Seminaren und Tutorien das Fundament der Lehre. Referate dienen als Nachweis aktiver Einarbeitung in ein Themengebiet, vermitteln anderen Studierenden Stoff (bei gleichzeitiger Entlastung des Dozenten) und stellen Lerngelegenheiten zum Erwerb von Präsentationskompetenz dar (*Bromme/Rambow 1993*). Da die Referate den wesentlichen Teil der Inhalte weitergeben und oft die überwiegende Veranstaltungszeit beanspruchen, ist die inhaltliche und didaktische Qualität dieser Beiträge zentral für den Veranstaltungserfolg. In Seminaren korreliert die Qualitätsbeurteilung der Referate mit der Allgemeineinschätzung des Kurses zu $r = 60$! Diese Referatqualität ist nicht unabhängig vom Dozenten. Er gibt einerseits den Veranstaltungsrahmen vor (Zeiteinteilung etc.), andererseits ist er über Betreuung und Feedback für die Referatqualität mitverantwortlich. Gerade an der Qualität der Referate wird von Studierenden jedoch häufig Kritik geübt, sie werden schlechter beurteilt als das Lehrverhalten des Dozenten (*Gold/Mayring 1997; Rindermann 1996a, S. 30 f.*). Anhand des HILVE (*Rindermann/Amelang 1994b*) lassen sich Referate summativ bewerten (alle vergangenen der Veranstaltung). Zur einzelnen Rückmeldung müßte wie durch Gold und Mayring (1996) oder Preiser (1995) ein spezielles, oft einsetzbares Referatfeedbackverfahren angewendet werden.

Störungen

Unruhe und Gespräche zwischen Studenten, Essen und Trinken (*Symposion*), Zeitung lesen und andere Nebentätigkeiten oder fehlende Pünktlichkeit (zu spät kommen oder früher gehen) stören den Veranstaltungsablauf. Störungen sind nicht unabhängig von Rahmenbedingungen (Besuchszahl, Besuchsgrund, Veranstaltungstyp, Anforderungen) und von der didaktischen wie sozialen Kompetenz des Dozenten, sie können deshalb als ein Indikator für Defizite in diesen Bereichen gelten.

Fehlzeiten

Auch Fehlzeiten der Studierenden können als Hinweis für Defizite in der Gestaltung der Rahmenbedingungen oder im Lehrverhalten des Dozenten genutzt werden. Allerdings sind diese ebenso von externen Faktoren abhängig, die nicht unter die Kontrolle der Hochschule fallen (Berufstätigkeit, Krankheit).

4.4 Rahmenbedingungen

Thema

Interessantheit und Relevanz des Themas werden in der vorgestellten Konzeption als eine dem Dozenten vorgegebene Rahmenbedingung des Kurses aufgefaßt.¹ Ein von den Studierenden als relevant betrachtetes Thema erleichtert dem Dozenten die Motivierung und fördert die Mitarbeit. Die Beurteilung des Themas korreliert in bisherigen Untersuchungen vor allem mit der Interessantheit der Veranstaltung. Bislang war jedoch die Kausalitätsrichtung nicht feststellbar, da nicht vor der 1. Veranstaltungssitzung anhand eines Seminarplanes das Thema unabhängig vom Kursgeschehen beurteilt wurde (vgl. *Rindermann 1996a*).

¹ Diehl und Kohr (1977) und Kramis (1990) betrachten sie eher als Aspekt der Lehrqualität und des Lehrerfolges, nach Klauer (1985) zeichnet sich gute Lehre durch adäquate Vermittlung (wie-Aspekt) bedeutsamer Inhalte (was-Aspekt) aus. Themen sind dem Dozenten in der Regel curricular vorgegeben. Die Lehre – nicht der Dozent – kann hinsichtlich der Themen bewertet werden.

Überschneidungen

Inhaltliche Überschneidungen zwischen Veranstaltungen sind ein Zeichen unzureichender Curriculumsorganisation. Sie können zu geringeren Besuchszahlen oder in Pflichtveranstaltungen zu Nebenbeschäftigungen und Störungen führen.

Anforderungen

Die Höhe der Anforderungen in Form von Stoffschwere, Stoffumfang und Tempo stellt ein zentrales Merkmal zur Beschreibung von Lehrveranstaltungen dar (vgl. *Banz/Rodgers 1985; Bolton et al. 1976; Burdsal/Bardo 1986*). Sie korreliert in HILVE-Studien negativ mit dem selbsteingeschätzten Lernerfolg und positiv mit Fleiß. Allerdings ist ebenso eine negative Beziehung zu Breite/Bezüge erkennbar (*Rindermann 1996a, S. 39*). In Analysen zur Generalisierbarkeit studentischer Beurteilungen (*Rindermann 1997b*) ließ sich beobachten, daß die Beurteilung der Anforderungen anhand des HILVE in bedeutendem Maße mit der Person des Dozenten variiert und somit auch auf sein Lehrverhalten (Anforderungsmanagement, Erklärungen etc.) zurückzuführen ist.

Besuchszahl

Steigende Besuchszahl erschwert kommunikative Veranstaltungsformen, z.B. Seminare mit Diskussion und Referaten. Auch Vorlesungen können bei schlechten räumlichen Bedingungen (nicht genug Sitzplätze, Akustik) unter großen Besuchszahlen leiden. Die Beteiligung nimmt in beiden Veranstaltungstypen mit zunehmender Teilnehmerzahl ab (*Rindermann 1996 a, S. 112; Marsh/Roche 1992*). Andererseits können geringe Besuchszahlen ebenso Folge schlechter Lehre sein. Die Besuchszahl sollte auch aus statistischen Gründen erhoben werden: Sie dient der Ermittlung der Teilnahmequote an der Lehrevaluation durch Relativierung der Fragebogenzahl an der Besuchszahl (Repräsentativitätsmaß).

Prüfung

In Veranstaltungen mit anschließender Prüfung beteiligen sich die Studierenden weniger, sie sind aber fleißiger. Prüfungsvorbereitungskurse zeichnen sich durch hohe Effektivitätsorientierung im Sinne von *Erwerb prüfbaren Wissens pro Zeiteinheit* aus.

Besuchsgrund

Besuchsgrund gilt neben dem Vorinteresse (s. Abschnitt 4.3) als wichtigste Biasvariable: Kurse, die wegen eines Scheines, einer Prüfung oder schlichtweg wegen Anwesenheitskontrolle zu besuchen sind, werden kritischer beurteilt als Veranstaltungen, die aus Interesse belegt werden (*Rindermann 1997a*). Die retrospektive Einschätzung des Besuchsgrunds gegen Semesterende bezieht sich möglicherweise auf das Thema und die Veranstaltungsgestaltung: Studierende, die aus Pflichtgefühl, aus Angst etwas zu verpassen oder um Freunde zu treffen (*Heger 1991*) freiwillig eine für sie langweilig gehaltene Veranstaltung besuchen, können kaum Interesse angeben. Um eine unabhängige Einschätzung zu erhalten, müßte der Besuchsgrund wie die Themenrelevanz vor Veranstaltungsbeginn erhoben werden.

Veranstaltungstyp

Hier werden Vorlesungen, Seminare und praxisorientierte Kurse (Praktika, Übungen) unterschieden. Für Vorlesungen und Seminare, nicht aber für Laborpraktika oder einübende Veranstaltungen (Sport, Sprachübungen), ist das gewählte Modell geeignet. In Vorlesungen wird der fakultative Block Referate freigelassen, bei Beteiligung, Interaktion und Betreuung treten angemessen niedrige Werte auf. Die Information Veranstaltungstyp ist für die Normierung der interaktiven Skalenresultate wichtig. Einstellungsänderungen und höhere kognitive Ziele wie Problemlösen sollen nach *Dunkin und Barnes (1986)* in sozial-interaktiven Lehrformen wie Seminaren eher erreichbar sein als in Vorlesungen (vgl. auch *Klauer 1985; Winteler 1974*).

4.5 Zielgröße Lehrerfolg

Zu den Zielgrößen erfolgreicher Lehre zählen neben veranstaltungsbezogenen Kriterien im engeren Sinne, wie **Interessantheit der Veranstaltungsgestaltung** und **allgemeine Veranstaltungsqualität**, auch Effekte bei den Teilnehmern in Form von quantitativem und qualitativem **Lerngewinn**, **Einstellungsänderungen** und **Kompetenzerwerb** (vgl. *Dunkin/Barnes 1986*). Die ersten beiden Effektvariablen auf Seiten der Studierenden können wie die Skalen der Bedingungsfaktoren sowohl durch Selbst- als auch durch Fremdbeurteilung erhoben werden. In der hier gewählten Begrifflichkeit wird Lerngewinn als Rating-Variable konzipiert. Als adäquate Maße für den Kompetenzerwerb gelten dagegen Prüfungen und Tests.

Interessantheit der Veranstaltungsgestaltung

Die Interessantheit der Veranstaltungsgestaltung wird als Effektvariable des Dozentenhandelns und im geringeren Maße des studentischen Vorinteresses konzipiert (vgl. empirische Resultate in *Rindermann 1996a, 1997b; Todt/Händel-Mattes 1990, S. 23, 126 ff.*). Manche Forscher heben deren Prozeß-Charakter hervor, da sie die allgemeine Veranstaltungsqualität, den Lernerfolg und Einstellungsänderungen positiv beeinflusst (s. *Schiefele/Krapp/Schreyer 1993*). Interessantheit läßt sich in Regressionsanalysen (HILVE-Daten) am besten durch die Skalen Thema, Verarbeitung, Struktur und Lehrkompetenz vorhersagen.

Allgemeine Veranstaltungsqualität

Wie in den meisten angloamerikanischen Verfahren wird ein *overall course* aufgenommen (vgl. a. *globale Lehrveranstaltungszufriedenheit* von *Westermann et al. 1996*). Dieser kann nicht die Ergebnisse in den einzelnen Bedingungsfaktoren ersetzen, vor allem sollten zur Beurteilung des Dozenten primär die Dozentenskalen herangezogen werden und in keinem Falle diffuse Zufriedenheitseinschätzungen (*Weiss 1991*). Die allgemeine Veranstaltungsqualität kann jedoch als Kriterium zur Bestimmung der Relevanz einzelner Skalen genutzt werden. Die Skalen *Interessantheit* und *Lernen* prognostizieren am besten Allgemeinbeurteilung, ohne andere Lehrerfolgsskalen sind Verarbeitungstiefe, Thema, Strukturierung und Lehrkompetenz die bedeutendsten Prädiktoren (zuf. *Rindermann 1996b*).

Lerngewinn (quantitativ/qualitativ)

Der quantitative und qualitative Lerngewinn gilt als wichtigstes Merkmal guter Lehre. Teilnehmer sollten mehr als vorher wissen und sie sollten über ein grundlegendes Verständnis in als wertvoll erachteten Bereichen verfügen (*Prosser/Trigwell 1991*). Mit Lerngewinn ist eher die Wissenszunahme im Vergleich zum Veranstaltungsbeginn konzipiert als der interindividuelle Leistungsstand im Vergleich zu anderen Veranstaltungsteilnehmern. Dies läßt sich durch den geringen Zusammenhang zwischen Klausurergebnissen und Lernselbsteinschätzungen empirisch bestätigen ($r = 18$, *Rindermann 1996a, S. 120*; $r = 13$, *Rosemann/Schweer 1996, S. 178*). Der Lerngewinn ist regressionsanalytisch (im HILVE) am besten durch die Faktoren Verarbeitung, Breite/Bezüge, Strukturierung, Thema und Fleiß vorhersagbar. Nach Marsh und Roche (*1992*) gelten Enthusiasmus und Struktur als die am besten differenzierenden Skalen für Lernerfolg.

Einstellungsänderungen

Gesteigertes Interesse gegenüber den Studieninhalten und eine gewachsene Motivation, sich mit diesen selbständig zu beschäftigen, kennzeichnen diese Skala. Gute Lehre weckt Neugier und intrinsisches Interesse (s. *Schiefele et al. 1993, S. 123; McKeachie 1997*). Diese Skala bestand in früheren Versionen des HILVE nur aus einem Item (Interesse gefördert), hierbei wurden regressionsanalytisch Thema (Kombination von Interesse und Relevanz, ohne Vorinteresse), Verarbeitungstiefe und Strukturierung als relevante Prädiktoren ermittelt.¹

Kompetenzerwerb

Im engeren Sinne bedeutet er den in Prüfungen nachweisbaren Lernerfolg. Im weiteren Sinne gehören auch nicht-kognitive Fähigkeiten und Fertigkeiten dazu (z.B. soziale Kompetenzen, Präsentationsmethodik, Lernstrategien; vgl. *Teml 1997*). Als Effektivvariablen wurden bislang aber nur kognitive Leistungen in Form von Klausurergebnissen herangezogen. Bei Heranziehung **individueller** Daten erwies sich die vorherige Kompetenz (gemessen an der Abiturnote) als wichtigstes Prognosemerkmal für Prüfungserfolg ($r = 48$; vgl. *Rindermann 1996 a, S. 120*). Veranstaltungs- und Dozentenskalen haben im Roh-Datensatz (Vergleich der Studenten untereinander) keine Prognosewirkung. Im Vergleich zwischen Dozenten (gleiche Kurse und Prüfungen, Kursmitteldaten; s. *Rindermann 1997 a*) korrelieren Dozentenskalen und Klausurresultate mit $r = 48$. Bei individuellen Daten kann das Leistungsniveau relativ gut das Klausurergebnis vorhersagen, zwischen Dozenten (bei egalisierten studentischen Stichproben) zeigt sich jedoch die bedeutsame Wirkung des Dozentenverhaltens auf Kursdurchschnittsergebnisse.

Die fünf vorgestellten Lehreffektivitätskriterien, die Ziele universitären Unterrichts wiedergeben, konkurrieren in bestimmten Konstellationen. Beispielsweise ist in *Paukkursen* mit hohen Arbeitsanforderungen und häufiger Leistungskontrolle kein positiver Einstellungswechsel zu erwarten. Nicht mit jeder Veranstaltungsform sind alle Ziele äquivalent erreichbar, Dozenten sollten je nach Zielpräferenz entsprechende Unterrichtsmethoden einsetzen.

¹ Falls bei Erhebung vor Veranstaltungsbeginn mit der veränderten Skala Thema und Vorinteresse ähnliche und kausal eindeutig interpretierbare Ergebnisse zu beobachten sind, liegt es näher, statt von Einstellungsänderung von Steigerung (intrinsischer) Motivation zu sprechen. Unter Einstellungsänderung wäre in einer weiteren Konzeption auch eine veränderte Einstellung gegenüber wissenschaftlicher Rationalität und ihren Erkenntnisverfahren oder gegenüber ethisch und gesellschaftlich bedeutsamen Fragestellungen zu verstehen.

Über diese Skalen hinaus können **fach- und veranstaltungsspezifische Items** die Angemessenheit eines Inventars für verschiedene Lehrziele und Unterrichtsformen sichern (vgl. die Liste freier Items in *Rindermann/Amelang 1994 b*). Erweiterungsmöglichkeiten bestehen z.B. bei den Rahmenbedingungen (Termin, Dauer, Räumlichkeit und Überfüllung), einzelnen Unterrichtsaspekten (Skript, Medien, Hausaufgaben, Arbeit mit Patienten, Prüfungen/Bewertungen) oder der Beurteilung des Fragebogen selbst (*Greenwood et al. 1973; Hage 1996; Marsh 1987*).¹ Kooperative Arbeitsformen wie Gruppen- und Partnerarbeit sollten hinsichtlich ihrer Gestaltung und Effektivität (z.B. hinsichtlich Prüfungsängstlichkeit und sozialer Integration; *Rotering-Steinberg 1995*) bewertet werden. Ein Grunditemsatz deckt die zentralen Unterrichtsprinzipien vieler Veranstaltungen in verschiedenen Studiengängen ab, für spezifische Aspekte ist jedoch eine Adaptation durch hinzufügbare Skalen notwendig.

5 Möglichkeiten der empirischen Überprüfung

Die vorgestellte Konzeption läßt sich mittels verschiedener Methoden überprüfen. Dazu zählen im weiteren Sinne alle statistischen Verfahren, die der Qualitätskontrolle entwickelter Fragebögen dienen (vgl. für den HILVE *Rindermann 1996 a, 1996 c, 1998*): Berechnungen von Urteilerübereinstimmungen zur Einschätzung der Objektivität, von Skalenhomogenitäten und Stabilitäten für die Meßgenauigkeit der Skalen und Items, von Beziehungen zwischen Skalen und externen Kriterien für die Validität und von Meßwiederholungsanalysen für den Nutzen des Einsatzes. Zusätzlich können Relevanzbefragungen bei Dozenten und Studenten die Praxistauglichkeit sichern.

Im engeren Sinne wird das Modell durch Verfahren überprüft, die

- die Attribuierbarkeit der einzelnen Skalen zu den Komponenten Dozenten, Studenten und Rahmenbedingungen über verschiedene Veranstaltungen und Dozenten hinweg untersuchen (Generalisierbarkeit, *Rindermann 1997 b*),
- die Beziehung zwischen Bedingungs- und Effektvariablen analysieren (Korrelationsstudien mit Einschätzungsdaten wie Selbst-, Studenten und Beobachterratings auf unterschiedlichen Inferenzniveaus oder experimentelle Designs),

¹ Oder Ergänzung durch ephemere Political-Correctness-Items, z.B.: „Der/Die Dozent/in benachteiligt Frauen gegenüber Männern nicht.“ „In der Veranstaltung wurden Themen zum Geschlechterverhältnis angemessen berücksichtigt.“ (Quelle: Münchner Inventar zur Lehrveranstaltungsevaluation, MILEVA). S. auch Seite 191, Fußnote 1.

- die Zuordenbarkeit der Items zu Skalen und der Skalen zu den Komponenten überprüfen (Faktorenanalysen),
- die Prädiktionskraft für den Lehrerfolg bestimmen (Regressionsanalysen) und
- die ein Lehrmodell als Ganzes testen (Strukturgleichungsmodelle).

Ergebnisse der Objektivitäts-, Reliabilitäts- und Validitätsanalysen wurden für den HILVE schon an anderer Stelle ausführlich geschildert, Beziehungen zwischen Bedingungs- und Effektvariablen wurden in Abschnitt 4 empirisch belegt. Im folgenden soll deshalb auf Verfahren eingegangen werden, die empirische Indikatoren zur Beurteilung der inhaltlichen Konzeption liefern.

5.1 Faktorenanalysen

In Faktorenanalysen der **Items** (N = 239 Veranstaltungsmittel aller bisherigen HILVE-Studien) fanden sich sieben Faktoren. Auf dem ersten Faktor laden Items der Dozentenskalen (Verarbeitung, Lehrkompetenz, Engagement, Kooperativität/Klima) und Allgemeinurteil, auf Faktor 2 Items der Lehreffektivitätsskalen (Interessantheit, Lernen, Allgemeinurteil) sowie Thema und Verarbeitung, auf Faktor 3 Items der Beteiligungs- und Diskussionsskala, auf Faktor 4 Anforderungs- und Fleißitems, Faktoren 5 (Breite/Bezüge), 6 (Referate) und 7 (Struktur) bündeln die gleichnamigen Skalenitems.

In Faktorenanalysen der **Skalen** (N = 239 Veranstaltungsmittel aller bisherigen HILVE-Studien) lassen sich vier Faktoren 2. Ordnung extrahieren: Der erste Faktor *Dozent* umfaßt Dozentenskalen (Struktur, Verarbeitung, Lehrkompetenz, Engagement, Kooperativität/Klima) sowie Interessantheit, Lernen und Allgemeinurteil, Faktor 2 *Lehrerfolg* Lehreffektivitätsskalen (Interessantheit, Lernen, Allgemeinurteil) sowie Thema und Breite, Faktor 3 *studentische Beteiligung* Referate, Beteiligung und Diskussion, Faktor 4 *Anforderung* Anforderungen und Fleiß.

Dozenten-, Studenten- und Rahmenbedingungs-skalen laden auf verschiedenen Faktoren. Die Lehrerfolgsskalen bilden entweder einen eigenen Faktor oder einen mit Dozentenskalen und Thema gemeinsam. Die Resultate erwiesen sich über verschiedene Stichproben und Variablensätze (Items vs. Skalen) als relativ konstant.

Kritisch ist anzumerken, daß die gewählte theoretische Dimensionierung feiner als die empirische ist, d.h. Studierende differenzieren zwar zwischen verschiede-

nen Aspekten der Lehre, innerhalb der Komponenten findet jedoch keine Unterscheidung zwischen allen einzelnen Skalen statt. Innerhalb der Dozentenskalen ließe sich zwischen dem Dozenten als Person (Lehrkompetenz, Engagement und Klima) und einzelnen Unterrichtstechniken (Breite/Bezüge, Struktur und Verarbeitung) unterscheiden. Um jedoch die Informationshaltigkeit des Feedbacks nicht zu reduzieren, wurde beim HILVE die Unterskalierung anhand inhaltlicher Kriterien beibehalten. Zudem läßt sich beobachten, daß studentische Skalen nicht immer von Rahmenbedingungsskalen trennbar sind. Dies ist jedoch insofern kein gravierendes Problem, als der Fleiß als angemessene Antwort auf die Höhe der Anforderungen gewertet werden kann. Empirische Zusammenhänge zwischen den verschiedenen Komponenten spiegeln nur das reale interdependente Lehr-geschehen wider.

Festzuhalten bleibt, daß die Multidimensionalität universitärer Lehre und ihrer Beurteilung durch Studierende nachgewiesen werden konnte (auch in der Beurteilung durch Lehrende und Fremdurteiler; *Rindermann 1996a*). Das Kriterium der allgemeinen Veranstaltungsqualität korreliert sowohl mit dem Handeln des Dozenten als auch mit den anderen LehrerfolgsvARIABLEN und stellt somit eine repräsentative Größe der Lehrqualität dar.

5.2 Regressionsanalysen

Durch schrittweise Regressionsanalysen wird die Prognosekraft von Prädiktorvariablen für Kriteriumsvariablen bestimmt. Prädiktorvariablen sind im vorgestellten Modell Skalen des Dozentenverhaltens, des studentischen Verhaltens und der Rahmenbedingungen. Kriterien sind die einzelnen Lehreffektivitätsskalen (s. Abschnitt 4.5). Prädiktorskalen werden nach ihrem inkrementellen Beitrag in ein statistisches Prognosemodell aufgenommen, d.h. als erstes geht die Skala mit der höchsten Korrelation zum Kriterium in die Gleichung ein. Weitere Skalen berücksichtigen Regressionsanalysen nur dann, wenn sie statistisch signifikant die Restvarianz aufklären (solche Anteile des Kriteriums, die noch nicht durch die vorherigen Variablen erklärt wurden).

Die Regressionsanalysen an HILVE-Daten (zusf. *Rindermann 1996b*) konnten die Ausprägungen in Lehreffektivitätsskalen durch Bedingungsvariablen gut aufklären: In Interessantheit wurden 78 % der Kriteriumsvarianz aufgeklärt, in Allgemeinbeurteilung 85 %, in Lernen 77 % und in Einstellungsänderung 76 %. Im Vergleich zu Neumann und Finaly (1989), die zwischen 71 % und 74 % der Lehrerfolgsvarianz durch Prozeßvariablen aufklären konnten, stellt dies eine hohe Quote dar. Die vorgestellten Bedingungsvariablen können demzufolge als relevante Determinanten des Lehrerfolgs bezeichnet werden.

5.3 Strukturgleichungsmodelle

Strukturgleichungsmodelle dienen der Analyse von Zusammenhangsstrukturen und komplexen Beeinflussungshypothesen auf der Ebene nicht direkt beobachtbarer latenter Variablen. Mittels Pfadanalysen, Kausal- oder Strukturgleichungsmodellen kann jedoch wie durch Regressionsanalysen kein Kausalitätsnachweis erbracht werden (*Eid 1996*). Diese können Kovarianzen/Korrelationen zwischen manifesten Variablen bzw. latenten Faktoren, die den Variablenbeziehungen zugrunde liegen, als eine notwendige Bedingung für Kausalität aufzeigen. Kovarianzen lassen sich nur auf Grundlage theoretischer Annahmen als kausale Beziehungen interpretieren. Multidirektionale Modelle wie das vorgestellte können mittels gängiger Strukturgleichungsmodelle nicht überprüft werden, da jene nur einseitige und nicht reziproke Beeinflussungsrichtungen annehmen (Multidirektionalität wäre nur bei mehreren Beobachtungsintervallen oder bei der Einführung exogener Variablen möglich; *Hodapp 1984, S. 33 ff. u. 94 ff.*). Legt man sich jedoch auf unidirektionale Hauptpfade fest, was – wie jedes Modell – die tatsächliche Systemkomplexität reduziert, sind die Modellannahmen mit den bisherigen Daten prüfbar.

Solche Modelltests liegen bislang nur in rudimentärer Form vor. In Pfadanalysen am HILVE-Datensatz (Heidelberger Daten) trat zwischen den latenten Variablen **Dozent und Lehreffektivität** ein bedeutsamerer Pfadkoeffizient auf ($\beta = 57$) als zwischen **Studenten und Lehreffektivität** ($\beta = 29$). Die **Anforderung** (bestehend aus Anforderungshöhe und Fleiß) hat keinen Einfluß ($\beta = 04$). Die Modellpassung als Ganzes ist jedoch noch nicht ausreichend (GFI-Index = 75). In bisherigen Analysen erwies sich die Heterogenität der Skalen innerhalb der Rahmenbedingungs- und Studierenden-Faktoren als problematisch. Zwar lassen sich die studentischen Skalen auf die Studierenden und ihr Verhalten attribuieren, untereinander besteht jedoch keine theoretisch notwendige wie empirisch beobachtbare Beziehung. Festhalten läßt sich jedoch der wesentlich bedeutendere Einfluß der unter dem Dozenten subsumierten Skalen gegenüber den studentischen Skalen für den Erfolg der Lehre.

6 Ausblick

Das Münchner multifaktorielle Modell der Lehrveranstaltungsqualität stellt eine Weiterentwicklung des Heidelberger Modells dar (*Rindermann 1997a*). In der revidierten Form wird der auf diesem Modell basierende veränderte Fragebogen in einer laufenden Studie überprüft. Bisher konnten alle wesentlichen Grundannahmen (Dimensionierung, Gütekriterien, Validität, Beziehung zwischen Bedingungs- und Effektvariablen), die im Heidelberger und Münchner Modell identisch sind, empirisch untersucht werden. Hierbei zeigten sich recht günstige

Ergebnisse, die weitgehend mit Resultaten der internationalen Lehrevaluationsforschung übereinstimmen. Im Gegensatz zu der primär methodisch ausgerichteten angloamerikanischen Forschung liegt dem Münchner Modell ein theoretischer Rahmen zugrunde, welcher die pädagogische Unterrichts-, hochschuldidaktische Lehr- und diagnostisch-methodische Evaluationsforschung zu integrieren versucht.

Über diesen Rahmen hinaus greifen Fragen der Implementation der Lehrevaluation an der Universität und die der Problematik der Lehrevaluationsforschung als wissenschaftliche Analyse eines an der Hochschule selbst beheimateten Bereichs. Soll die Lehre an der Universität überhaupt evaluiert werden? Falls ja, durch wen und wie? Sollen Lehrveranstaltungen oder die Studiensituation als Ganzes an einem Fachbereich oder beides beurteilt werden? Wenn ja, wie oft? Welche Konsequenzen sollen die Ergebnisse haben (vgl. *Rindermann 1997c*)? Dies sind organisationssoziologische, hochschulpolitische und nicht zuletzt auch ökonomische Fragen, die auf den jeweiligen Ebenen untersucht und politisch entschieden werden müssen.

Für die Forschung stellt die Lehrevaluation eine besondere Herausforderung dar, da hier Wissenschaftler untersuchen, die gleichzeitig Hochschullehrer sind, wie Professoren und Dozenten und deren Lehre durch Studierende beurteilt werden können (*Rindermann 1997d*). Dies stellt hohe Anforderungen an das Differenzierungsvermögen zwischen wissenschaftlichen Erkenntnisfragen und statusbezogenen Rollenerwartungen.

Lehrevaluation sollte auch deshalb nicht nur als universitäre Praxis, sondern ebenso als Gegenstand wissenschaftlicher Arbeit behandelt werden. Wer wie *Preißer (1994)* die sozialwissenschaftliche Forschung auf diesem Gebiet – im Vergleich zur Verbesserung der Lehre durch Lehrveranstaltungsbewertungen – als nachrangig erklärt, enthält sich der Möglichkeit, Resultate einer solchen Forschung für die universitäre Evaluationspraxis zu nutzen. Diese Forschung kann dazu beitragen, Einwände, wie sie immer wieder von Professoren hinsichtlich der Urteilsfähigkeit der Studierenden vorgebracht werden, zu entkräften und ein effektives, zur Verbesserung der Lehrqualität beitragendes Einsatzverfahren zu entwickeln.

Danksagung:

Frau Dipl.-Kffr. Katrin Milde und Herrn Prof. Dr. Andreas Gold möchte ich für hilfreiche Ideen und Anmerkungen zu Vorformen des Manuskripts danken.

Literaturverzeichnis

Abrami, Ph. C.; Mizener, D. A. (1983): Does the attitude similarity of college professors and their students produce "bias" in course evaluations? *American Educational Research Journal*, 20 (1), 123-136

Abrami, Ph. C.; Perry, R. P.; Leventhal, L. (1982): The relationship between student personality characteristics, teacher ratings, and student achievement. *Journal of Educational Psychology*, 74 (1), 111-125

Astleitner, H.; Krumm, V. (1996): Dimensionen von Lehrverhalten: Faktorenstrukturen 1. und 2. Ordnung mit Kreuzvalidierung. *Empirische Pädagogik*, 10 (1), 7-26

Banz, M. L.; Rodgers, J. L. (1985): Dimensions underlying student ratings of instruction: a multidimensional scaling analysis. *American Educational Research Journal*, 22 (2), 267-272

Basler, H.-D.; Bolm, G.; Dickescheid, Th.; Herda, Ch. (1995): Marburger Fragebogen zur Akzeptanz der Lehre. *Diagnostica*, 41 (1), 62-79

Berk, R. A. (1979): The construction of rating instruments for faculty evaluation. A review of methodological issues. *Journal of Higher Education*, 50 (5), 650-669

Bolton, B.; Bonge, D.; Hinman, S. (1976): Dimensions of students' evaluation of instruction: an empirical synthesis. *Psychological Reports*, 38, 119-123

Bromme, R.; Rambow, R. (1993): Verbesserung der mündlichen Präsentation von Referaten: Ein Ausbildungsziel und zugleich ein Beitrag zur Qualität der Lehre. *Das Hochschulwesen*, 6, 289-295

Burdsal, Ch. A.; Bardo, J. W. (1986): Measuring student's perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurement*, 46, 63-79

Cruickshank, D. R.; Kennedy, J. J. (1986): Teacher clarity. *Teaching & Teacher Education*, 2 (1), 43-67

Daniel, H.-D. (1994): Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung. *Empirische Pädagogik*, 8 (2), 109-129

Daniel, H.-D. (1995): Der Berufseinstieg von Betriebswirten. *Personal*, 47 (10), 492-499

Diehl, J. M. (1994): VBVOR und VBREF. Fragebögen zur studentischen Evaluation von Hochschulveranstaltungen. Gießen: hektographiertes Manuskript

- Diehl, J. M.; Kohr, H.-U. (1977):* Entwicklung eines Fragebogens zur Beurteilung von Hochschulveranstaltungen im Fach Psychologie. *Psychologie in Erziehung und Unterricht*, 24, 61-75
- Döring, K. W. (1995):* Lehren in der Weiterbildung. Ein Dozentenleitfaden. Weinheim
- Dunkin, M. J.; Barnes, J. (1986):* Research on teaching in higher education. In M. C. Wittrock (Hg.), *Handbook of research on teaching* (S. 754-777). New York
- Eid, M. (1996):* Lineare Strukturgleichungsmodelle. In Strauß, B. u.a. (Hrsg.), *Techniken der Datenbearbeitung in der Sportwissenschaft*. Schorndorf
- Feldman, K. A. (1976):* The superior college teacher from the students' view. *Research in Higher Education*, 5, 243-288
- Frey, P. W. (1978):* A two-dimensional analysis of student ratings of instruction. *Research in Higher Education*, 9, 69-91
- Gold, A.; Mayring, P. (1996):* Wie gut sind studentische Referate: Eine Studie zur Evaluation der Lehre. München: Poster auf dem 40. Kongreß der DGPs
- Gold, A.; Mayring, P. (1997):* Dimensionen studentischer Seminarbewertung. Frankfurt: Poster auf der 6. Tagung Pädagogische Psychologie
- Good, K. C.; Good, L. R. (1973a):* Assumed attitude similarity and instructor evaluation. *Journal of Social Psychology*, 91, 285-290
- Good, K. C.; Good, L. R. (1973b):* Attitude similarity and attraction to an instructor. *Psychological Reports*, 33, 335-337
- Greenwood, G. E.; Bridges, Ch. M.; Ware, W. B.; McLean, J. E. (1973):* Student evaluation of college teaching behaviors instrument: A factor analysis. *Journal of Higher Education*, 44, 596-604
- Groeben, N. (1986):* Handeln, Tun, Verhalten als Einheiten einer verstehend-erklärenden Psychologie. Tübingen
- Hage, N. (1996):* Lehrevaluation und studentische Veranstaltungskritik. Bonn: Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie
- Heger, M. (1991):* Lehrveranstaltungskritik als angewandte Hochschuldidaktik. *Zeitschrift für Hochschuldidaktik*, 15 (1-2), 43-69
- Helmke, A. (1996):* Studentische Evaluation der Lehre - Sackgassen und Perspektiven. Anmerkung zum Beitrag von Rosemann & Schweer (1996). *Zeitschrift für Pädagogische Psychologie*, 10 (3/4), 181-186

- Hildebrand, M. (1973):* The character and skills of the effective professor. *Journal of Higher Education*, 44, 41-50
- Hodapp, V. (1984):* Analyse linearer Kausalmodelle. Bern
- Klauer, K. J. (1985):* Framework for a theory of teaching. *Teaching & Teacher Education*, 1 (1), 5-17
- Kramis, J. (1990):* Bedeutsamkeit, Effizienz, Lernklima. Grundlegende Gütekriterien für Unterricht und Didaktische Prinzipien. *Beiträge zur Lehrerbildung*, 8, 279-296
- Land, M. L. (1981):* Combined effect of two teacher clarity variables on student achievement. *Journal of Experimental Education*, 50 (1), 14-17
- Marques, T. E.; Lane, D. M.; Dorfman, P. W. (1979):* Toward the development of a system for instructional evaluation: Is there consensus regarding what constitutes effective teaching? *Journal of Educational Psychology*, 71 (6), 840-849
- Marsh, H. W. (1987):* Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388
- Marsh, H. W.; Roche, L. (1992):* The use of student evaluations of university teaching in different settings: The applicability paradigm. *Australian Journal of Education*, 36 (3), 278-300
- Marsh, H. W.; Roche, L. (1997):* Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52 (11), 1187-1197
- Mayr, J. (1997):* Evaluieren. In Buchberger, F. u.a. (Hrsg.), *Seminar didaktik* (S. 224-256). Innsbruck
- McKeachie, W. J. (1997):* Student ratings. *American Psychologist*, 52 (11), 1218-1225
- Meredith, G. M. (1975):* Toward a systems approach to student-based ratings of instruction. *Journal of Psychology*, 91, 235-246
- Merton, R. K.; Zuckerman, H. (1985 [1971]):* Institutionalisierte Bewertungsstrukturen in der Wissenschaft. In R. K. Merton (Hrsg.), *Entwicklung und Wandel von Forschungsinteressen* (S. 172-216). Frankfurt
- Moses, I. (1985):* High quality teaching in a university: identification and description. *Studies in Higher Education*, 10 (3), 301-313
- Müller-Wolf, H.-M. (1977):* Lehrverhalten an der Hochschule. München

- Multrus, F. (1995):* Zur Lehr- und Studienqualität. Dimensionen, Skalen und Befunde des Studierendensurveys. Konstanz
- Neumann, Y.; Finaly-Neumann, E. (1989):* An organizational behavior model of students' evaluation of instruction. *Higher Education*, 18, 227-238
- Nigsch, O. (1994):* Pflichtevaluierung oder das Ei des Ministeriums. *Zeitschrift für Hochschuldidaktik*, 3/4, 255-280
- Opp, K.-D. (1976):* Methodologie der Sozialwissenschaften. Hamburg
- Preiser, S. (1995):* Feedback nach Referaten. Ein Beitrag zur Verbesserung der Lehre. *Das Hochschulwesen*, 2, 114-116
- Preißer, R. (1994):* Die Außeralltäglichkeit der deutschen Hochschulen: Hochschule als paradoxe Bürokratie. *Beiträge zur Hochschulforschung*, 2, 151-186
- Prosser, M.; Trigwell, K. (1991):* Student evaluations of teaching and courses: Student learning approaches and outcomes as criteria of validity. *Contemporary Educational Psychology*, 16, 293-301
- Reischmann, J. (1995):* Bamberger Seminar-Evaluierungsbogen BEVA. Beiheft, Kopiervorlagen, Auswertungsprogramm. Bamberg: hektographiertes Manuskript
- Reissert, R. (1992):* Evaluation der Lehre - Aktuelle Aktivitäten an deutschen Hochschulen. Hannover
- Renkl, A. (1996a):* Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47, 78-92
- Renkl, A. (1996b):* Vorwissen und Schulleistung. In Möller, J. u.a. (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 175-190). Weinheim
- Rheinberg, F. (1996):* Von der Lernmotivation zur Lernleistung: Was liegt dazwischen? In Möller, J. u.a. (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 23-50). Weinheim
- Rindermann, H. (1996a):* Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen. Landau
- Rindermann, H. (1996b):* Zur Qualität studentischer Lehrveranstaltungsevaluationen: Eine Antwort auf Kritik an der Lehrevaluation. *Zeitschrift für Pädagogische Psychologie*, 10 (3/4), 129-145

Rindermann, H. (1996c): Verbesserung der Lehre durch Veranstaltungsevaluation? In Witruk, E. u.a. (Hrsg.), *Pädagogische Psychologie im Streit um ein neues Selbstverständnis*. Bericht über die 5. Fachgruppentagung Pädagogische Psychologie in Leipzig (S. 147-156). Landau

Rindermann, H. (1997a): Die studentische Beurteilung von Lehrveranstaltungen: Forschungsstand und Implikationen für den Einsatz von Lehrevaluationen. In Jäger, R. S. u.a. (Hrsg.), *Tests und Trends (Jahrbuch der Pädagogischen Diagnostik, 11)* (S. 12-53). Weinheim

Rindermann, H. (1997b): Generalisierbarkeit studentischer Veranstaltungsbeurteilungen: Sind Lehrevaluationsresultate generalisierbar auf andere Veranstaltungen eines Dozenten oder auf inhaltsgleiche Parallelveranstaltungen verschiedener Dozenten? *Psychologie in Erziehung und Unterricht*, 44(3), 216-234

Rindermann, H. (1997c): Lehrveranstaltungsbewertungen: Erfahrungen und ihre Nutzung für die Fakultätsentwicklung. In Altrichter, H. u.a. (Hrsg.), *Hochschulen auf dem Prüfstand*. (S. 179-196). Innsbruck

Rindermann, H. (1997d): Das Selbstobjektivierungsproblem im akademischen Milieu: Forschung an Hochschulen über die Qualität der Lehre an Hochschulen. München: Unveröff. Manuskript

Rindermann, H. (1998): Übereinstimmung und Divergenz bei der studentischen Beurteilung von Lehrveranstaltungen: Methoden zu ihrer Berechnung und Konsequenzen für die Lehrevaluation. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19 (2), 73-93

Rindermann, H.; Amelang, M. (1994a): Entwicklung und Erprobung eines Fragebogens zur studentischen Veranstaltungsevaluation. *Empirische Pädagogik*, 8 (2), 131-151

Rindermann, H.; Amelang, M. (1994b): Das Heidelberger Inventar zur Lehrveranstaltungs-Evaluation (HILVE). Handanweisung. Heidelberg

Rindermann, H.; Oubaid, V. (1998): Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. München & Halle: Unveröff. Manuskript

Rosemann, B.; Schweer, M. K. W. (1996): Evaluation universitärer Lehre und der Wissenszuwachs bei den Studierenden. *Zeitschrift für Pädagogische Psychologie*, 10 (3/4), 175-180

Rotering-Steinberg, S. (1995): Kooperative Formen des Lehrens und Lernens in der Erwachsenenbildung. *Unterrichtswissenschaft*, 23 (4), 332-346

- Schiefele, U.; Krapp, A.; Schreyer, I. (1993):* Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25 (2), 120-148
- Schmidt, H. G.; Dolmans, D.; Gijssels, W. H.; Marchais, J. E. (1995):* Theory-guided design of a rating scale for course evaluation in problem-based curricula. *Teaching and Learning in Medicine*, 7 (2), 82-91
- Seel, A. (1997):* Handlungsorientiertes Lernen. In Buchberger, F. u.a. (Hrsg.), *Seminardidaktik* (S. 98-113). Innsbruck
- Shatz, M. A.; Best, J. B. (1986):* Selection of items for course evaluation by faculty and students. *Psychological Reports*, 58, 239-242
- Slotnick, H. B.; Durkovic, R. G. (1975):* Dimensions of medical students' perception of instruction. *Journal of Medical Education*, 50, 662-666
- Spiel, Chr.; Gössler, M. (1995):* Evaluierung von Lehrveranstaltungen (un)möglich? Erste Ergebnisse einer repräsentativen Untersuchung an der Universität Wien. Bremen: Vortrag auf der 2. Fachgruppentagung Methoden
- Spies, K.; Westermann, R.; Heise, E.; Schiffler, A. (1996):* Diskrepanzen zwischen Bedürfnissen und Angeboten im Studium und ihre Beziehungen zur Studienzufriedenheit. *Empirische Pädagogik*, 10 (4), 377-409
- Standhardt, R.; Löhmer, C. (Hrsg.) (1995):* Lebendiges Lernen in toten Räumen. Zur Verbesserung der Lehre an der Hochschule. Gießen
- Sturm, M. (1994):* "Die Reise ins Ich ...". Selbstevaluation im Hochschulunterricht. *Handbuch Hochschullehre*, 1-28 (D 1.2)
- Teml, H. (1997):* Kurse und Blockveranstaltungen. In Buchberger, F. u.a. (Hrsg.), *Seminardidaktik* (S. 267-274). Innsbruck
- Todt, E.; Götz, Ch. (1997):* Veranstaltungs-Rückmeldung (VR). Programm zur Evaluation von Veranstaltungen an der Universität. Gießen: Fachbereich Psychologie
- Todt, E.; Händel-Mattes, B. (1990):* Motivation und Motivierung im Unterricht. Gießen: Arbeitsbericht, Fachbereich Psychologie
- Tollefson, N.; Chen, J. S.; Kleinsasser, A. (1989):* The relationship of students' attitudes about effective teaching to students' ratings of effective teaching. *Educational and Psychological Measurement*, 49, 529-536
- Webler, W.-D. (1996a):* Qualitätssicherung in Fachbereichen - ein erster Modellvergleich. *Das Hochschulwesen*, 1, 16-27

- Webler, W.-D. (1996b):* Qualitätssicherung in Lehre und Studium an deutschen Hochschulen. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 16 (2), 119-148
- Weinert, F. E. (1996):* Für und Wider die "neuen Lerntheorien" als Grundlagen pädagogisch-psychologischer Forschung. *Zeitschrift für Pädagogische Psychologie*, 10 (1), 1-12
- Weinert, F. E.; Schrader, F.-W.; Helmke, A. (1989):* Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13, 895-914
- Weiss, R. (1991):* Ziele und Probleme einer Lehrveranstaltungskritik. *Zeitschrift für Hochschuldidaktik*, 15 (1-2), 35-42
- Westermann, R.; Heise, E.; Spies, K.; Hagen, M.; Rickert, M.; Schiffler, A.; Trautwein, U.; Wollburg, S. (1993):* Zufriedenheit mit Studium und Lehrveranstaltungen: Theorieorientierte Konzeptualisierung von Erhebungsverfahren. Göttingen: Psychologisches Institut, Arbeitsbericht Nr. 1
- Westermann, R.; Heise, E.; Spies, K.; Trautwein, U. (1996):* Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie in Erziehung und Unterricht*, 43, 1-22
- Wilson, R. C.; Dienst, E. R.; Watson, N. L. (1973):* Characteristics of effective college teachers as perceived by their colleagues. *Journal of Educational Measurement*, 10 (1), 31-37
- Wilson, T. C. (1988):* Student evaluation-of-teaching forms: A critical perspective. *The Review of Higher Education*, 12 (1), 79-95
- Winteler, A. (1974):* Determinanten der Wirksamkeit akademischer Lehrveranstaltungen. (*Psychologia Universalis*, Band 30). Meisenheim
- Winteler, A.; Schmolck, P. (1983):* Überprüfung eines Schätzverfahrens zur Beurteilung von Lehrveranstaltungen. *Schweizerische Zeitschrift für Psychologie*, 42 (1), 56-79
- Wissenschaftsrat (1996):* Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation. Berlin: Bericht des Wissenschaftsrates
- Wottawa, H.; Thierau, H. (1990):* Evaluation. Bern

Anschrift des Verfassers

Dr. Heiner Rindermann
Institut für Pädagogische Psychologie und Empirische Pädagogik
der Ludwig-Maximilians-Universität München
Leopoldstr. 13

80802 München