



## Probleme der externen Evaluation wirtschaftswissenschaftlicher Studiengänge

Von Alfred Kieser, Erich Frese,  
Detlef Müller-Böling, Norbert Thom<sup>1</sup>

### Überblick

- Eine Reihe von Pilotprojekten zur Evaluation von Fachbereichen zeigt an, daß die zuständigen Ministerien entschlossen sind, Evaluationen in Deutschland zu institutionalisieren.
- Eine Übersicht über verschiedene Ansätze der Evaluationstheorie macht deutlich, daß zu Konzepten und Methoden sehr kontroverse Ansichten existieren. Von einer gesicherten theoretischen Basis kann nicht ausgegangen werden.
- Die meisten Pilotprojekte zur Evaluation von Fachbereichen in Deutschland haben sich das niederländische Modell zum Vorbild genommen, in dem sich Fachbereiche zunächst selbst evaluieren und eine Peer-Gruppe, auf diesem Bericht aufbauend, eine externe Evaluation erstellt.
- Evaluationen dieses Typs haben vor allem mit drei Schwierigkeiten zu kämpfen: (1) Es gibt keinen Konsens darüber, was unter guter Lehre zu verstehen ist. (2) Evaluationen haben verschiedene Ziele, die untereinander in Konflikt stehen. (3) Die Methoden der Evaluation sind weder valide noch zuverlässig. Aus diesen Schwierigkeiten ergeben sich einige Konsequenzen für die zukünftige Durchführung von Evaluationen.

<sup>1</sup> Prof. Dr. Alfred Kieser, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Organisation, Universität Mannheim.

Prof. Dr. Erich Frese, Seminar für Allgemeine Betriebswirtschaftslehre und Organisationslehre, Universität zu Köln.

Prof. Dr. Detlef Müller-Böling, Centrum für Hochschulentwicklung, Gütersloh.

Prof. Dr. Norbert Thom, Institut für Organisation und Personal, Universität Bern.

„Evaluation is a very peculiar breed of cat“ (Scriven 1983, S. 256)

## A. Problemstellung

In den letzten Jahren schwappte eine Welle von Pilotprojekten zur Evaluation der Lehre über die deutschen Hochschulen. Vieles deutet darauf hin, daß weitere Wellen folgen werden, daß Evaluationen von Fachbereichen institutionalisiert werden (HRK 1995) und daß intendiert ist, die Mittelvergabe an Indikatoren zu knüpfen, wie sie auch in Evaluationen zur Anwendung kommen.

In diesem Aufsatz geben wir einen Überblick über einige Grundprobleme der Evaluation von Fachbereichen, wir skizzieren in Deutschland durchgeführte Pilotprojekte und wir versuchen, die spezifischen Probleme herauszuarbeiten, die durch die Pilotprojekte u.E. deutlich wurden. Es ist unser Ziel, die Sensitivität gegenüber Problemen einer Evaluation der Lehre an Fachbereichen innerhalb der verschiedenen Interessengruppen zu fördern, insbesondere den Angehörigen wirtschaftswissenschaftlicher Fachbereiche Informationen und Argumente zur Verfügung zu stellen, auf die sie in der zu erwartenden intensiveren Diskussion um Lehrevaluationen gegebenenfalls zurückgreifen können.

## B. Grundlegendes zur Evaluation der Lehre

### I. Definition

Evaluation ist ein vielschichtiger Begriff, was sich nicht zuletzt an der Vielzahl von Definitionen zeigt (Wottawa/Thierau 1990, S. 9f.). Es gibt, wie wir noch sehen werden, unterschiedliche Grundpositionen in der Evaluationsforschung, und diese schlagen auf die Bestimmung dessen durch, was unter Evaluation zu verstehen ist. Als Ausgangsbasis mag folgende Definition genügen: Evaluation ist „the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action“ (Beeby 1977, S. 68). Sie bringt zum Ausdruck, daß eine *Beurteilung* erfolgt, und betont, daß diese immer auf Werten beruht. Und sie weist darauf hin, daß dazu Evidenz in systematischer Weise zu sammeln ist und daß diese Evidenz nicht für sich selbst spricht, sondern zu interpretieren ist, bzw. daß in der Erhebung von Evidenz und deren Präsentation schon ein Gutteil Interpretation steckt (Wolf, R. M. 1990). Schließlich macht die Definition deutlich, daß Evaluation nicht zum Selbstzweck erfolgt, sondern Bemühungen um die Behebung festgestellter Mängel auslösen soll.

Diese Definition enthält keine Aussagen darüber, was evaluiert wird. Normalerweise geht es um „Programme zu sozialen Interventionen“, um Aktivitätenbündel, durch die Menschen auf Menschen einwirken, um bei ihnen eine Änderung in eine bestimmte Richtung zu bewirken: Ausbildungsprogramme; Programme zur Beseitigung von Arbeitslosigkeit, wirtschaftlicher Rückständigkeit von Regionen, Armut; Programme zur Suchtbekämpfung usw.

## II. Grundpositionen

In den USA ist die Evaluationsforschung seit den frühen 60er Jahren als eigenständige Disziplin etabliert. Sie hat sich in unterschiedliche Richtungen aufgespalten (eine Übersicht bieten Madaus et al. 1983). Die Hauptströmungen lassen sich nach der Chronologie ihrer Entstehung in folgende Kategorien zusammenfassen (in Anlehnung an Scriven 1993, S. 5 ff., dem Nestor der Evaluationsforschung):

- (1) *Evaluation als zielorientierte, objektive Entscheidungsunterstützung.* Dieser Ansatz geht von der Annahme aus, daß Programme objektiv im Hinblick auf ihre Zielsetzung evaluiert werden können, so daß als Ergebnisse der Evaluation eindeutige Empfehlungen zu erwarten sind. Ein bedeutender Vertreter dieser Richtung ist Stufflebeam (1983), der Begründer des CIPP-Ansatzes. Um ein Programm zu evaluieren, so die zentrale Aussage dieses Ansatzes, müssen Kontext (context: welche Zielgruppen mit welchen Bedürfnissen sind zu berücksichtigen?), Input (vor allem: Human- und Sachressourcen, Problemlösungsstrategien), Prozeß (Aktivitäten der Einwirkung) und Produkt (Ergebnisse) analysiert werden. Dazu sollen sich die Evaluatoren eines breiten Arsenal sozialwissenschaftlicher Methoden bedienen von Hearings über Delphi-Befragungen, Dokumentenanalyse und Beobachtungen bis hin zu quantitativen Tests, die etwa zur Feststellung eingesetzt werden, ob die Leistungen von Absolventen von Ausbildungsprogrammen den Lehrzielen gerecht werden.
- (2) *Die relativierende Evaluation.* Diese Richtung problematisiert die Ziele, an denen Programme gemessen werden sollen. Die expliziten Ziele von Programmen sind häufig vage oder widersprüchlich. Auch können verschiedene, in die Programme involvierte Interessengruppen durchaus unterschiedliche Ziele oder zumindest unterschiedliche Gewichtungen der Ziele an das Programm herantragen. Hinzu kommt, daß sich viele Wissenschaftler einer wertfreien Wissenschaft verpflichtet fühlen und Skrupel haben, die Ziele des Auftraggebers umstandslos ihrer Evaluation zugrunde zu legen. Als Ausweg aus diesen Dilemmata schlagen die Vertreter dieses Ansatzes vor: (a) Die Evaluation soll nur Daten generieren und die Bewertung den potentiellen Entscheidungsträgern überlassen (für diese Position steht etwa Alkin 1972). (b) Der Evaluator soll die Ziele des Auftraggebers zugrunde legen und sich einer eigenen Bewertung enthalten (diese Position nehmen bspw. Rossi et al. 1988 ein).
- (3) *Der Prozeß-Ansatz.* Ihm geht es weniger darum, Unterstützung für Entscheidungen bestimmter Interessengruppen zu liefern; vielmehr will er allen Interessengruppen Einsichten in die Interdependenzen der Programmparameter vermitteln, so daß diese ohne weiteres Zutun der Evaluatoren in der Lage sind, Maßnahmen zu ergreifen, die zur Erhöhung von Effektivität und Effizienz des Programms beitragen (Vertreter dieser Richtung sind Cronbach (Cronbach et al. 1980) und Scriven (1993)). Dies heißt jedoch nicht, daß sich die Evaluatoren wertender Aussagen enthalten. Im Gegenteil: Scriven fordert, daß die Evaluation zu eindeutigen Aussagen über die Qualität der Leistungen des Programms gelangen soll. Es ist ihm aber wichtig, daß sich die Evaluatoren nicht irgendwelche vorgegebenen Ziele von Interessengruppen zu eigen machen, sondern aus der Perspektive des öffentlichen Interesses heraus Stellung beziehen:

„It is crucial to see that the evaluation point of view is not the manager's point of view, and is not simply the consumer's point of view; it is a point of view which should stand above identification with either of these parties, but make clear to each the importance of the other.“ (Scriven 1980, S. 103)

Der Evaluator soll folglich eine „Goal-free-Evaluation“ betreiben, d. h., Bewertungskriterien nicht von bestimmten Interessengruppen übernehmen, sondern aus dem Programm und seinem Kontext heraus selbst erarbeiten (Shadish/Cook/Leviton 1991, S. 73 ff.).

- (4) *Der konstruktivistische Ansatz.* Die Vertreter dieses Ansatzes, allen voran Guba und Lincoln (1989), gehen von der Annahme aus, daß eine Evaluation niemals eine „objektiv-zutreffende“ Beschreibung des Zustands, der Effektivität und der Effizienz eines Programms zum Ergebnis haben kann, weil solche Beschreibungen immer soziale Konstruktionen der Wirklichkeit sind. Jede Gruppe konstruiert aus ihrer Interessenslage und ihrer Sicht der Dinge heraus eine Wirklichkeit. Auch den Evaluatoren gelingt es nicht, eine objektive neutrale Rolle einzunehmen. Demzufolge kann ihre Aufgabe nur darin bestehen, die Wirklichkeitskonstrukte der verschiedenen Interessengruppen zutage zu fördern, eine Diskussion unter den Interessengruppen in Gang zu bringen und zu moderieren, mit dem Ziel, so weit wie möglich überlappende Konstruktionen der Wirklichkeit herzustellen und, aufbauend auf diesen, Aktionsprogramme zur Reform zu vereinbaren.

### III. Ziele

#### I. Ziele in der Evaluation

Wie wir bei der kurzen Übersicht über die Grundpositionen der Evaluationsforschung bereits gesehen haben, ist die Frage, an welchen Zielen das zu evaluierende Programm gemessen werden soll, nicht einfach zu beantworten. Häufig werden Programme durchgeführt, ohne daß die Akteure präzisere Vorstellungen über die Ziele herausbilden, z. B.: Ein zentrales Ziel wirtschaftswissenschaftlicher Fachbereiche ist sicherlich die Qualifizierung für anspruchsvolle Führungstätigkeiten in der Wirtschaft. Was jedoch ist darunter genau zu verstehen? „Berufsfertigkeit“ oder „Berufsfähigkeit“? Gehört zur wirtschaftswissenschaftlichen Ausbildung nicht vielleicht auch die Vermittlung von Wissen zur kritischen Reflexion gesellschaftlicher Strukturen und der eigenen Rolle in der Gesellschaft? Befragungen von Hochschullehrern, Vertretern von Ministerien oder Praktikern fördern dazu höchst unterschiedliche Antworten zutage.

In den Darstellungen der Grundpositionen sind die Möglichkeiten zur Lösung dieses Problems bereits angeklungen: Die Evaluatoren veranlassen entweder den Auftraggeber – in unserem Fall das Ministerium – oder zentrale Akteure – die Professoren –, Ziele zu formulieren, und legen dann ihrer Evaluation die jeweiligen Ziele zugrunde. Oder der Evaluator nimmt die „echte“ Position des Konsumenten ein, die nicht unbedingt mit der von diesem dargestellten übereinstimmen muß. Er eruiert die Bedürfnisse des Konsumenten – der Studierenden – oder die Anforderungsprofile des Beschäftigungssystems und prüft, inwieweit das Programm diesen Bedürfnissen bzw. Anforderungen gerecht wird.

## 2. Ziele der Evaluation

Lehrevaluationen sollen Informationen für Entscheidungen liefern. Für verschiedene Entscheider und für verschiedene Entscheidungssituationen ergeben sich unterschiedliche Informationsanforderungen und damit unterschiedliche Evaluationsziele. So kann sich bspw. eine Evaluation der Lehrleistungen von Dozenten auf die Leistungsdeterminanten „Fähigkeiten/Wissen“, „Rollenperzeption“ und „Leistungsanstrengung (Motivation)“ beziehen und Informationen zu Entscheidungen über Qualifizierungs- und Versetzungsmaßnahmen, über die Vermittlung anforderungsrelevanten Wissens oder über Anreizmaßnahmen liefern.

In Evaluationen von Fachbereichen stehen vor allem folgende Evaluationsziele zur Diskussion (so in den Niederlanden, s. Richter 1991 a):

- (1) *Informationen für hochschulinterne Maßnahmen.* Durch die Evaluation soll Handlungsdruck erzeugt werden, der im Wege der Selbstorganisation zur Sicherung und Verbesserung der Lehrqualität beiträgt. Außerdem kann durch einen hochschulübergreifenden Informationsaustausch im Sinne des *organizational learning* der Wissensstand über die Alternativen der Gestaltung der Lehraktivitäten erhöht werden.
- (2) *Informationen für staatliche Maßnahmen.* Ministerien als Träger der Hochschulen sollen in die Lage versetzt werden, zur Beseitigung von Schwachstellen auf die Hochschulen einzuwirken.
- (3) *Informationen für potentielle Nachfrager nach Lehrleistungen.* Studierende sollen ihre Entscheidungen für eine Hochschule auf der Grundlage veröffentlichter Lehrevaluationen treffen können. Entsprechendes gilt für Einstellungsentscheidungen von Unternehmungen und Institutionen.
- (4) *Informationen für die Öffentlichkeit in Erfüllung einer gesellschaftlichen Rechenschaftspflicht.*

Angesichts der verschiedenartigen Zwecke, die mit der Evaluation verfolgt werden, sind Konflikte bei der Gestaltung des Evaluationssystems unvermeidlich. Das Ziel, Handlungsdruck zur Selbstorganisation zu erzeugen, steht bspw. teilweise im Gegensatz zum Ziel, Ministerien Planungsdaten zu liefern. Ein Fachbereich ist weniger bereit, interne Probleme aufzudecken, wenn dies negative Konsequenzen für die Finanzierung haben kann; es wird „gemauert“, und darunter leidet dann auch die Bereitschaft, sich selbst Mängel einzugestehen und an ihrer Behebung zu arbeiten.

Die Ziele der Evaluation haben Auswirkungen auf die Ziele in der Evaluation: Wenn die Evaluation von Fachbereichen Daten für die Allokation von Mitteln zur Verfügung stellen soll, so erwarten die Entscheidungsträger in den Ministerien eindeutige, auf Zielkriterien, die ihre Interessen widerspiegeln, bezogene Urteile. Um internen Reformdruck auszulösen, genügt u. U. auch eine Evaluation, die lediglich auf Probleme aufmerksam macht.

## IV. Evaluatoren

Die verschiedenen Grundpositionen der Evaluation weisen den Evaluatoren unterschiedliche Rollen zu: vom Technokraten über den „Prozeßberater“ bis hin zum Moderator. Re-

lativ unabhängig von ihrer Grundposition sind sich die Evaluationsexperten darin einig, daß Evaluatoren über einige Charakteristika verfügen müssen, die über umfassende Methodenkenntnisse hinausgehen (Scriven 1993, S. 87f.), etwa die Fähigkeit, unvoreingenommen zuhören zu können, Erfahrungen im Umgang mit Konflikten, mit Widerstand gegen Evaluationen und gegen politische Einflußnahme sowie Sensitivität gegenüber politischen Manövern im evaluierten System.

Einig sind sich die professionellen Evaluierer (natürlich!) auch darin, daß Evaluation eine Angelegenheit für „professionals“, nicht für „Amateure“ ist – auch nicht für Amateure, die Experten in dem zu evaluierenden Feld sind (Peers also). So meint Scriven (1993, S. 50) zur Peer-Evaluation im Hochschulbereich (wenn auch im Hinblick auf die Evaluation von Forschungsprogrammen):

„The panel lacks training in the skills that are relevant, notably, the skills of looking past the halo of high research ability to evaluate the soundness of management plans, the past track record in project completion, the critical competitors in science, mathematics, engineering, and technology ... education, the needs assessment, the generalizability, and the many other matters that come into proposal evaluation. ... Of course, most of us think that we can evaluate program, and so we serve on panels to do it from a sense of duty, or a lust for power. But most of us are just amateurs at it, as any competent evaluator serving on (or serving as program officer for) the panels soon sees. As one might expect, the interpanel reliability is poor.“

Als weitere Mängel von Peer-Reviews werden moniert (Scriven 1983, S. 251): Renommiertere Kollegen, die bevorzugt als Peers ausgewählt werden, präferieren konventionelle Ansätze, manchmal auch wissenschaftliche Moden (Scriven 1993, S. 85), sie sind negativ eingestellt gegenüber Kollegen, die etablierte Paradigmen in Frage stellen, legen ungerechtfertigt hohe Standards an die Ausstattung an und majorisieren häufig die Evaluatorengruppe.

## V. Methoden/Validität/Regeln der Fairneß

Alle Methoden von quantitativen Tests (denen etwa Absolventen unterzogen werden, um die Qualität der Ausbildung zu überprüfen) über qualitative Beobachtungen (etwa von Studenten-Professoren-Interaktionen) bis hin zu Hearings oder „gerichtsähnlichen“ Verfahren (Wolf, R. L. 1990) sind in Evaluationen zugelassen (Cook/Reichardt 1979). Jede Methode hat ihre eigenen Validitäts- und Reliabilitätsprobleme, die zu diskutieren hier zu weit führen würde. Da der Evaluator bestimmt, welche Methoden eingesetzt und welche Daten erhoben werden, also aus vielen Möglichkeiten selektiert, ohne daß die Evaluationsforschung dazu immer eindeutige Begründungen liefert, sind neben den methodischen Regeln auch noch *professionelle Standards oder Standesnormen* von Bedeutung, die vor allem am Kriterium der *Fairneß* orientiert sind. Sie sollen sicherstellen, daß der Evaluator seine Position nicht mißbraucht und daß der Prozeß transparent für die Betroffenen bleibt und ihnen auch Interventionschancen bietet. Solche Regeln hat etwa das *Joint Committee on Standards for Educational Evaluation* (Sanders 1994) entwickelt. Einige Beispiele:

„The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.“

Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of information can be assessed.

The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.“

## VI. Metaevaluation

Evaluation ist keine exakte Wissenschaft. Und wie die Psychiatrie ist sie reflexiv: Man kann die Evaluation einer Evaluation unterziehen. Vieles spricht dafür, dies auch zu tun (Scriven 1993, S. 51 f.), denn eine Evaluation ist nicht gefeit gegen Voreingenommenheiten, Wahl inadäquater Methoden und fehlerhafte Interpretationen. Die Evaluierten und die Auftraggeber sind nicht in der Lage, die Qualität der Evaluation zu beurteilen. Eine Supervision – eine Metaevaluation – hält die Evaluatoren zu einer ständigen kritischen Überprüfung ihrer Verfahren und Ergebnisse an, liefert den Klienten eine unabhängige Überprüfung der Qualität der Evaluation und kann eine Akkumulation von Erfahrungen und Verbesserungen der Methoden sicherstellen (Straw/Cook 1990; Schwandt/Halpern 1988):

## VII. Substitute der Evaluation

Wir haben uns bisher auf die klassische externe Evaluation konzentriert, in der ein Evaluator oder ein Evaluatorenteam den Einsatz von Methoden konzipiert, auf der Basis dieser Methoden Daten erhebt, interpretiert und einen Bericht erstellt. Alternativen zu diesem Vorgehen sind die Institutionalisierung der Erhebung bestimmter *Kennzahlen* und die Durchführung von *Rankings*. Kennzahlen bringen Zielerreichungen von Systemen zum Ausdruck, erlauben u. U. also eine direkte Interpretation durch Interessengruppen, ohne daß die Hilfe eines interpretierenden Evaluators in Anspruch genommen werden muß. Werden Kennziffern und andere Daten wie bspw. Befragungen von Experten herangezogen, um verschiedene Systeme zu vergleichen und in eine Rangfolge zu bringen, so wird ein Ranking erstellt. Selbstverständlich kann eine Evaluation neben anderen Methoden auch Kennzahlen und Rankings einbeziehen.

Zur Beurteilung von Fachbereichen werden sowohl Kennzahlen eingesetzt (Hummel 1988) als auch Rankings durchgeführt. Wesentlich seriöser als die von deutschen Zeitschriften „veranstalteten“ Rankings sind die differenziert nach Studiengängen durchgeführten Hochschulrankings in den USA, die seit Jahrzehnten auf der Basis von im wesentlichen unveränderten Methoden durchgeführt werden. Sie bauen auf fünf Skalen auf (Kellermann 1992):

- (1) *Academic Reputation*. Diese Skala erfaßt die Wertschätzung, die der betreffende Studiengang unter Hochschullehrern genießt. Sie wird ermittelt auf der Basis einer Be-

- fragung aller Präsidenten, Dekane und Leiter der Studentenverwaltungen der in die Auswahl einbezogenen Hochschulen.
- (2) *Student Selectivity*. Diese Skala wird aus vier Komponenten gebildet: der Rate der aus der Gruppe aller Bewerber angenommenen Studienanfänger; dem Prozentsatz derer, die sich dann tatsächlich immatrikuliert haben; den Schulleistungen sowie der Rate jener Studenten, die in zwei alternativen, landesweit abgehaltenen Tests unter die besten zehn Prozent kamen.
  - (3) *Faculty Resources*. Diese Skala stellt ein Aggregat aus vier Komponenten dar: dem Zahlenverhältnis von Vollzeitstudenten zu Vollzeitlehrenden; dem Anteil an promovierten Professoren an allen Lehrenden; dem Anteil an Lehrenden mit Teilzeitverträgen sowie dem Durchschnittsgehalt der Professoren ohne Zeitverträge (also mit „tenure“). Unter „faculty resources“ ist mithin der Grad der Bereitschaft der jeweiligen Hochschule zu verstehen, hochqualifizierte Vollmitglieder der Fakultäten in ausreichender Anzahl anzuwerben und finanziell zu unterstützen.
  - (4) *Financial Resources*. Diese Skala erfaßt die allgemeinen und auf die Ausbildung bezogenen Ausgaben im Verhältnis zur Anzahl immatrikulierter Studenten.
  - (5) *Student Satisfaction*. Mit dieser Skala, die eigentlich den Studienerfolg mißt, wird der durchschnittliche Prozentsatz an Erstimmatrikulierten erfaßt, der innerhalb fünf Jahren graduiert.

Indem diese Rankings auf einem *standardisierten Verfahren* beruhen und zu einem hohen Grad auf „hard facts“ aufbauen, unterscheiden sie sich wesentlich von den Hochschulrankings deutscher Zeitschriften. Daneben unterziehen sich die Departments amerikanischer Universitäten noch „Internal Reviews“ und „Peer Reviews“ (Banta 1992; Ewell 1992; Kellermann 1992; Lenn 1992).

Als Instrumente zur Förderung des Wettbewerbs zwischen Universitäten weisen auf strukturierten Verfahren beruhende und sich auf „objektive“ Daten stützende Rankings gewisse Vorteile gegenüber Peer-Evaluationen auf. Das System von Rankings in den USA realisiert bspw. in einem hohen Maße eine *Bewertung durch den Markt*, die Vorurteilen von Evaluatoren nicht unterworfen ist.

## C. Projekte zur Lehrevaluation an deutschen Hochschulen

### I. Pilotprojekt in den Fächern Germanistik und Biologie an den Universitäten Hamburg und Kiel sowie Evaluation im Verbund norddeutscher Universitäten

Nachdem Konzepte der Evaluation von Lehrveranstaltungen Anfang der 90er Jahre verstärkt diskutiert worden waren und zu verschiedenen Projekten von Seiten der Studierenden, Hochschulen und Ministerien (Daniel 1994) geführt hatten (Holtkamp/Schnitzer 1992), wurde das Interesse auf Konzeptionen einer umfassenden Evaluation der Lehre an Fachbereichen gelenkt. Eines der ersten Pilotprojekte, das die Evaluation ganzer Fachbereiche zum Gegenstand hatte, war die 1993 von der Behörde für Wissenschaft und Forschung in Hamburg und dem schleswig-holsteinischen Ministerium für Wissenschaft, For-

schung und Kunst in Zusammenarbeit mit den Universitäten Hamburg und Kiel sowie HIS gestartete Evaluation der Fächer Germanistik und Biologie (Reissert 1994; Christian-Albrechts-Universität zu Kiel/Universität Hamburg 1995). Das von HIS konzipierte Verfahren lehnt sich eng an das holländische Konzept der Hochschulevaluierung an (Richter 1991 a, b; Sluis 1992; Frederiks et al. 1994), das wiederum stark durch die Peer-Reviews, wie sie im Rahmen der Akkreditierung von Ausbildungsinstitutionen in den USA durchgeführt werden (Scriven 1983, S. 248), beeinflusst ist. Dieses Pilotprojekt verfolgte folgende vertraglich fixierten Ziele (Reissert 1994, S. 1): Verbesserung der Transparenz des Lehr- und Ausbildungsbetriebes, Erkennen von Ursachen für unterschiedliche Studienzeiten, Schaffung einer Informationsbasis für Entscheidungen und Planungen, Unterstützung der Evaluation von Lehrveranstaltungen, Intensivierung des Diskurses über das Lehren und Lernen innerhalb der Fächer, Erarbeitung von Vorschlägen zur Verbesserung der Lehre und des Studiums. Es ging darum, „die Fähigkeit der evaluierten Fächer zur Selbststeuerung und zur Studienreform zu fördern“. Die staatliche Seite leistete die Finanzierung, nahm aber auf den Ablauf des Projekts keinen Einfluß.

Dieses Projekt umfaßte wie das niederländische eine Selbstevaluation der Fachbereiche und eine darauf aufbauende Evaluation durch externe Gutachter. Im einzelnen wurden folgende Stufen durchlaufen:

- (1) *Vorbereitung, insbesondere Erstellung eines Leitfadens für die Selbstevaluation.* Der Leitfaden wurde weitgehend vom niederländischen Modell übernommen. Er enthält Fragen zur Qualität der Lehre – Beispiele: „Sind allgemeine Ausbildungs- und Qualifikationsziele für den Studiengang formuliert? Welche fachlichen und beruflichen Ziele der Ausbildung hinsichtlich des Wissens, der Fertigkeiten und Fähigkeiten sowie Werthaltungen werden angestrebt? Was soll ein Absolvent können? Gibt es Vorstellungen über das Verhältnis von Fachwissen, beruflichem Wissen und akademischer Bildung/Persönlichkeitsbildung? . . . Gibt es ein besonderes Profil des Studienganges gegenüber anderen vergleichbaren Studiengängen? . . . Gibt es eine Studienordnung? . . . Wann ist diese zum letzten Mal überprüft worden; entspricht sie noch dem aktuellen Stand des tatsächlich durchgeführten Studienprogramms?“ (Reissert 1994, Anhang 1). Daneben wird auch nach Indikatoren gefragt wie Quote der Studienabbrecher, Drittmittel für Lehre und Forschung, Relation von Lehrenden zu Studierenden, Relation von Lehrenden zu Absolventen.
- (2) *Erarbeitung der Selbstevaluation – des Lehrberichts – in den Fachbereichen.*
- (3) *Auswahl der Gutachter* – „jeweils drei Hochschullehrer der Biologie und Germanistik“ (Reissert 1994, S. 5). Die Fächer einigten sich auf gemeinsame Vorschlagslisten, die jeweils neun Namen auswiesen. Die Behörden beriefen die Gutachter, indem sie die vorgegebene Reihenfolge einhielten.
- (4) *Vorbereitung der Vor-Ort-Begehung.* Die Gutachter stimmten untereinander auf der Basis ihrer Kenntnis der Lehrberichte die Begehung ab.
- (5) *Vor-Ort-Begehung.* Der Ablaufplan einer solchen Begehung ist in Kasten 1 wiedergegeben (aus Reissert 1994, Anlage 3).
- (6) *Anfertigung der Gutachten.*
- (7) *Stellungnahme der Fächer zum Gutachten.*
- (8) *Veröffentlichung der Abschlußberichte.*

*Kasten 1: Musterplan „Ablauf Begehung“*

<p><b>Vorbereitung der Begehung</b></p> <ul style="list-style-type: none"> <li>- Abendessen mit Hochschulleitung, Dekan, Vor-Ort-Koordinatoren und Gutachtern</li> <li>- eventuell Abstimmung kurzfristiger Korrekturen der Begehung</li> </ul> <p><b>1. Tag</b></p> <ul style="list-style-type: none"> <li>- Auftaktgespräch mit Vor-Ort-Arbeitsgruppe „Lehrbericht“ (Schwerpunkt u. a. Informations- und Verständnisfragen, Austausch über die Erfahrungen mit der Evaluation – insbesondere Selbstevaluation)</li> <li>- Gespräch mit             <ul style="list-style-type: none"> <li>• Dekan und Institutsdirektoren (Schwerpunkte u. a. Umfeld für die Evaluation/Hochschulpolitik, Lehrangebot, Ausbildungsziele, Stärken und Schwächen im Studium und Lehre, Lehrmanagement und -organisation)</li> <li>• Studienkommission „Studium und Lehre“ (Schwerpunkt u. a. Studienprogramm, Studierbarkeit, Studienplan, Studienreform)</li> <li>• Prüfungsausschuß (Schwerpunkt u. a. Prüfungsablauf und -organisation, Leistungsstand der Studierenden)</li> </ul> </li> </ul> <p><i>Mittagessen</i></p> <ul style="list-style-type: none"> <li>- Gruppengespräche mit Studierenden getrennt nach Studienphasen (Schwerpunkt u. a. Lehre, Studien- und Prüfungsverlauf, Studierbarkeit des Studiengangs, Betreuung und Beratung)</li> <li>- Gespräch mit Lehrenden – Hochschullehrer und wissenschaftl. Mittelbau – (Schwerpunkt u. a. Lehre, Lehr- und Lernformen, Lehre und Prüfungsbelastung, Studienvoraussetzungen und -anforderungen)</li> <li>- Gespräch mit Fachstudienberatung (Schwerpunkt u. a. Studierbarkeit, Studienhindernisse, Beratung und Betreuung)</li> </ul> <p><b>2. Tag</b></p> <ul style="list-style-type: none"> <li>- Führung durch das Fach (Schwerpunkt u. a. Kennenlernen der zentralen Studien- und Lehrorte des Faches)</li> <li>- Einzelfachgespräche „Gutachter-Lehrende“ (Schwerpunkt u. a. Qualität der Lehre, fachliches Lehrangebot des Institutes/des Gesprächspartners, Studien- und Lehrprobleme aus der fachlichen Perspektive eines Lehrgebietes)</li> </ul> <p><i>Mittagessen</i></p> <ul style="list-style-type: none"> <li>- internes Gespräch der Gutachter (Schwerpunkt u. a. Resümee der Begehung und Formulierung erster Eindrücke/Empfehlungen)</li> <li>- Abschlußgespräch mit Vor-Ort-Koordinatoren und -Arbeitsgruppe sowie Schlüsselakteuren des Faches (Schwerpunkt u. a. Darstellung der ersten gutachterlichen Eindrücke/Empfehlungen sowie offener Meinungs-austausch über das Evaluationsverfahren)</li> <li>- eventuell abends oder am nächsten Tag „Nachbereitung der Vor-Ort-Begehung“ (nur Gutachter und/oder Geschäftsstelle)</li> </ul>
--

Der staatlich initiierte Evaluationsprozeß wurde erweitert durch ein autonomes Verfahren im Verbund norddeutscher Universitäten. Diesem „Nordverbund“ gehören die Universitäten Rostock, Kiel, Hamburg, Bremen und Oldenburg an. Sie haben in einer ersten Runde parallel die Fächer Germanistik und Biologie evaluieren lassen und setzen nunmehr den Prozeß mit weiteren Fächern – bspw. Betriebswirtschaftslehre (jedoch nicht in

Kiel, wo dieses Fach durch das Projekt des Wissenschaftsrats bereits evaluiert wurde) – fort (Lüthje, 1995 a, b). Auf die Erfahrungen der Beteiligten mit diesem Projekt wird noch eingegangen.

## II. Das Pilotprojekt des Wissenschaftsrates

In seinen „10 Thesen zur Hochschulpolitik“ vom Januar 1993 hatte der Wissenschaftsrat u. a. konstatiert: „Die offenkundigen Funktionsmängel und die öffentliche Kritik an den Hochschulen verlangen nach internen und externen Verfahren der Evaluation.“ Zur Umsetzung dieser Forderung hat der Wissenschaftsrat im Frühjahr 1994 einen ständigen „Ausschuß Lehre“ eingerichtet, der sich insbesondere der Entwicklung und Förderung einer überregionalen Evaluation der Lehre annehmen soll. Dieser Ausschuß hat dann im Juni 1994 entschieden, in die exemplarische praktische Erprobung eines zweistufigen Evaluationsverfahrens in ausgewählten Fächern/Studiengängen an verschiedenen Hochschulen einzutreten. „Ziel ist die Prüfung der Tragfähigkeit eines interne und externe Evaluation verknüpfenden Instrumentariums. Für die angestrebte größere Transparenz der Lehrleistungen einzelner Hochschulen und Fachbereiche sollen aber auch solche Daten und Leistungsindikatoren identifiziert werden, die längerfristig an den Fachbereichen vorzuhalten wären.“ (Geschäftsstelle des Wissenschaftsrates 1994)

Der Leitfaden des Wissenschaftsrates ist ebenfalls stark am niederländischen Modell orientiert. Der Aufbau ist etwas abweichend, und die Fragen sind z. T. anders formuliert. Es werden aber im wesentlichen die gleichen Tatbestände erhoben. Es gibt einen qualitativen Teil und einen Teil, in dem die Ermittlung von Indikatoren vorgegeben ist. Auch der Seitenumfang ist in etwa der gleiche.

Für eine erste Erprobungsrunde hat der „Ausschuß Lehre“ die Fächer Physik und Betriebswirtschaftslehre ausgewählt und dafür jeweils eine fachspezifische Unterarbeitsgruppe als Evaluationskomitee eingesetzt. Der Unterarbeitsgruppe Betriebswirtschaftslehre gehören Praktiker, Universitätsprofessoren, Fachhochschulprofessoren sowie ein Vertreter eines Landesministeriums und ein Vertreter des Bundesministeriums für Bildung und Wissenschaft an. Drei betriebswirtschaftliche Fachbereiche an Universitäten und zwei Fachbereiche an Fachhochschulen wurden evaluiert. Das Procedere war weitgehend identisch mit dem der norddeutschen Universitäten. Als Abweichungen für die Betriebswirtschaftslehre sind festzuhalten: (1) Das Gespräch mit den Studierenden wurde auf der Basis der Erfahrungen der ersten Evaluation – von den Studierenden kam Kritik, mit der die Evaluatoren die für die Lehre Verantwortlichen konfrontieren konnten – vor dem Gespräch mit dem Dekan und den Institutionsdirektoren bzw. den Verantwortlichen für die Evaluationen geführt. (2) Es wurden keine getrennten Gespräche mit zwei Gruppen von Studierenden – Anfänger und Fortgeschrittene – geführt. (3) Eine Stellungnahme der evaluierten Fachbereiche zum Bericht war nicht vorgesehen.

Einer der Autoren dieses Aufsatzes war Gutachter, ein anderer Angehöriger eines betugachteten Fachbereichs. Auf die Erfahrungen mit diesem Pilotprojekt kommen wir ebenfalls später zurück.

### III. Weitere Projekte

In der „Exemplarische Fachanalysen zur Qualitätssteigerung von Lehre und Studium“ in Hessen bringen die Fachbereiche eine Selbstevaluation in eine „Fachkonferenz“ ein, die sich aus Vertretern aller Fachbereiche eines Faches des Landes und des Ministeriums (!) zusammensetzt (Hessisches Ministerium für Wissenschaft und Kunst 1993). Die Fachkonferenz übernimmt sozusagen die Rolle der externen Evaluatoren. Der zum Einsatz kommende Leitfaden weist deutliche Unterschiede zu denen des Wissenschaftsrates und der Universitäten Hamburg und Kiel auf: Der Schwerpunkt liegt eindeutig auf Indikatoren. Die Vorgaben zu den qualitativ zu erhebenden Tatbeständen beschränken sich auf Überschriften und einzelne Stichworte (z. B.: „Studien- und Prüfungsbedingungen (Strukturen, Erfahrungen, Engpässe, Verbesserungsvorschläge)“; „studienvorbereitende und -einführende Information und Beratung (Konzeption, Materialien, Veranstaltungen)“; „Grundstudium (Konzeption, Organisation, Studierbarkeit)“).

Während die bisherigen Projekte Peer-Evaluationen repräsentierten (allerdings mit einer partiellen Durchbrechung dieses Prinzips in Hessen), stellte das Projekt „Externe Prüfung der Wirtschaftlichkeit in Universitäten des Landes Baden-Württemberg“ (HAYEK Engineering AG 1993) eine Expertenbeurteilung dar. Eine Unternehmensberatungsgesellschaft untersuchte den Ressourceneinsatz in Forschung und Lehre. Diese Evaluation war zwar explizit auf die ökonomische Dimension gerichtet, was die Evaluatoren jedoch nicht davon abhielt, auch Empfehlungen zur Form der Lehre und zum Einsatz von Hilfskraftmitteln in Forschung und Lehre auszusprechen.

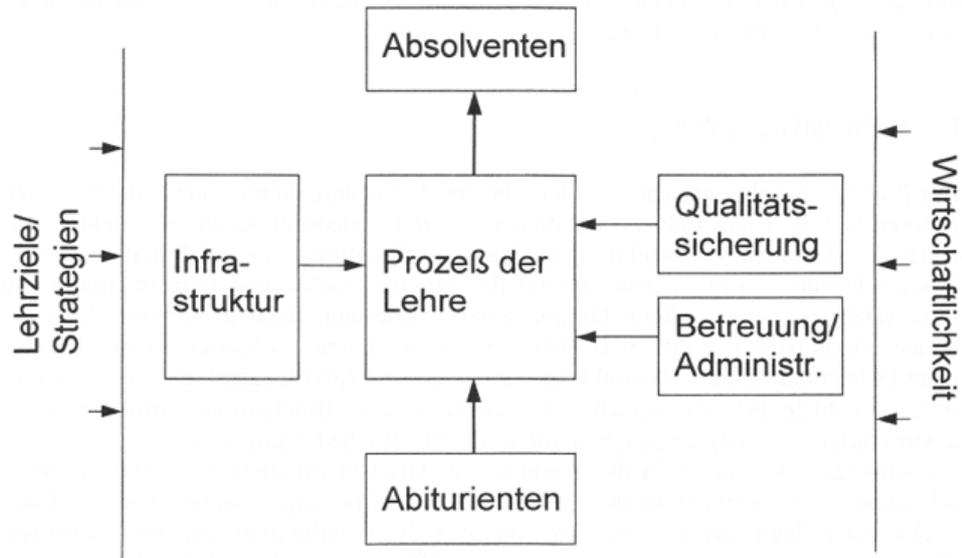
### IV. Bewertung

Die Pilotprojekte zeigen, daß das *niederländische Modell die deutschen Bestrebungen zur Einführung von Lehrevaluationen stark beeinflusst* hat. Im folgenden wollen wir uns deshalb auf dieses Modell konzentrieren. Wir fragen, welche Probleme mit seinem Einsatz verbunden sind und an welche Voraussetzungen er geknüpft ist. Zunächst wollen wir aber den Gegenstand der Evaluation – die akademische Lehre an wirtschaftswissenschaftlichen Fakultäten – etwas eingehender betrachten.

#### **D. Ein Modell der wirtschaftswissenschaftlichen Ausbildung an Universitäten**

Eine erste wichtige Aufgabe im Rahmen der Konzipierung einer Evaluation ist es, ein Modell des zu evaluierenden Systems zu erstellen, um die relevanten Dimensionen der Evaluation identifizieren zu können (De Landsheere 1990). Ein Modell der Lehre an wirtschaftswissenschaftlichen Fachbereichen könnte das in Abbildung 1 wiedergegebene Aussehen haben.

Abb. 1: Ein Modell der wirtschaftswissenschaftlichen Ausbildung an Universitäten



### I. Die Zielkomponente

Wenn man Wettbewerb befürwortet, muß man auch einen Wettbewerb um Lehrziele zulassen. Es gibt wohl keinen Zweifel, daß ein wirtschaftswissenschaftlicher Fachbereich Studierende für einen Einsatz im Wirtschaftssystem qualifizieren muß. Daneben kann es aber auch noch ein Ausbildungsziel sein, die Studierenden generell auf eine kritische Auseinandersetzung mit gesellschaftlichen Problemen und auf die Lebensbewältigung außerhalb des engeren beruflichen Feldes vorzubereiten (Persönlichkeitsbildung). Schließlich gibt es sicherlich auch Unterschiede in der Interpretation der beruflichen Qualifikation, die etwa mit den Begriffen „Berufsfähigkeit“ und „Berufsfertigkeit“ umrissen werden. Es erscheint uns angemessen, wirtschaftswissenschaftliche Fachbereiche vor allem danach zu beurteilen, inwieweit sie *ihren eigenen Zielvorstellungen* gerecht werden.

Darüber, wie solche Ausbildungsziele am besten zu erreichen sind (was ja einer impliziten Konkretisierung der Ziele gleichkommt), gibt es kaum Konsens. Es gibt keine einheitlichen Auffassungen über die gültigen Theorien zur Lösung wirtschaftlicher Probleme. Selbst über den Charakter dieser Probleme besteht häufig Dissens. Strittig kann deshalb auch sein, ob ein in die Lehre übernommenes Forschungsergebnis einen Fortschritt darstellt. Die Folge ist, daß über die weitgehend geteilte Forderung nach Integration von Forschung und Lehre hinaus über die erforderlichen Kernqualifikationen des zukünftigen Managers sowie über den Inhalt und den Umfang der Wissensvermittlung (und damit auch über die Studiendauer) unterschiedliche Vorstellungen bestehen.

Am ehesten kann man noch definieren, welche Kenntnisse für bestimmte Berufsfelder (z. B. öffentlicher Dienst) oder für bestimmte Einsatzgebiete (z. B. Rechnungswesen) er-

forderlich sind (vgl. z. B. die Empfehlungen verschiedener Fachkommissionen für Ausbildungsfragen der Schmalenbach-Gesellschaft – Deutsche Gesellschaft für Betriebswirtschaft e. V. in den 80er Jahren).

## II. Der Prozeß der Lehre

Der Prozeß der Lehre umfaßt vor allem die zur Anwendung kommende Didaktik. Dazu gehören (s. z. B. Thom 1980): Auswahl der Lehrziele, Auswahl der zu vermittelnden Inhalte und Methoden, Auswahl der Dozenten (Lehrbeauftragte aus der Praxis vs. akademische Dozenten), Strukturierung der Inhalte, Einsatz verschiedener Lernsituationen und Lernverfahren (Selbststudium, Gruppenarbeit, Vorlesung, Seminar, Übung, Tutorien, Planspiele, Praktika), Wahl von Lehrstrategien (Anordnung des Stoffes, Auswahl geeigneter Darstellungsmittel), Auswahl der verschiedenen Lehrstile (autokratisch vs. interaktiv), Auswahl der Prüfungsformen (studienbegleitende vs. Blockprüfung; strukturierte vs. unstrukturierte Prüfungsfragen, mündliche vs. schriftliche Prüfung usw.).

Fachbereiche überlassen in aller Regel die didaktische Gestaltung den einzelnen Hochschullehrern. Das enthebt sie des Problems, Konsens in Fragen herzustellen, zu denen höchst unterschiedliche Vorstellungen unter den Hochschullehrern bestehen. Fachbereiche funktionieren als „loosely coupled systems“ (Weick 1976). Lediglich in Bereichen, in denen verschiedene Hochschullehrer zusammenwirken müssen – in wirtschaftswissenschaftlichen Fachbereichen sind dies vor allem das Grundstudium, die Allgemeine Betriebswirtschaftslehre und die Allgemeine Volkswirtschaftslehre im Hauptstudium – ist eine intensivere didaktische Diskussion nicht zu vermeiden. Sie beschränkt sich aber meist auf die grobe Umgrenzung des zu vermittelnden Lehrstoffes und der Lehrformen (Vorlesungen, Übungen, Seminare, Tutorien) und überläßt die Gestaltung der Details den mit den verschiedenen Lehrveranstaltungen jeweils betrauten Dozenten.

## III. Infrastruktur der Lehre

Die Qualität der Lehre hängt von der Ausstattung ab, worunter u. a. fallen: Hörsäle (einschließlich Ausstattung mit technischen Hilfsmitteln), Gruppenarbeitsräume, Bibliotheken (Ausstattung mit Büchern und Periodika, Öffnungszeiten, Recherchemöglichkeiten usw.), EDV-Arbeitsplätze und Rechnerkapazitäten.

Die Ausstattung sollte von der Didaktik bestimmt sein (wenn sich derzeit auch die Didaktik in der Regel nach der zu kurzen Decke der knappen Mittel strecken muß). So setzen bspw. Gruppenkonzepte der Lehre Gruppenarbeitsräume und Tutorenprogramme Mittel für ungeprüfte Hilfskräfte voraus; der Einsatz von Planspielen erfordert EDV-Arbeitsplätze, theorieorientierte Diplomarbeiten eine gut ausgestattete Bibliothek usw. Es ist also wichtig, daß die *Ausstattung* nicht als solche oder nur relativ zu anderen Fachbereichen beurteilt wird, sondern *im Hinblick auf die Lehrkonzeption*.

#### IV. Maßnahmen der Qualitätssicherung

Im Rahmen von Diskussionen zur Qualität als Wettbewerbsvorteil ist häufig die These anzutreffen, daß die Erhaltung und Steigerung von Qualität die Institutionalisierung von Maßnahmen der Qualitätssicherung bedingt. In diesem Sinne wird etwa die Zertifizierung nach ISO 9000 ff. nicht an die Prüfung der Qualität des Outputs eines Unternehmens geknüpft, sondern an das Vorhandensein von Prozessen der Qualitätssicherung. Auf Fachbereiche übertragen würde das bedeuten: Es wird angenommen, daß Fachbereiche, die Verfahren der Qualitätssicherung institutionalisiert haben, eine bessere Qualität der Ausbildung realisieren als Fachbereiche ohne solche Maßnahmen. Zu Maßnahmen dieser Art gehören etwa: Absolventenbefragungen, Evaluationen von Lehrveranstaltungen, Erhebung und Verfolgung von Kennzahlen wie durchschnittliche Studiendauer, Drop-out-Raten, Vorhandensein von Didaktikkursen und Anreizkonzepten für Lehrende usw.

Das Vorhandensein solcher Maßnahmen sagt jedoch nichts darüber aus, mit welchen Intensitäten sie genutzt werden und wie ihre Qualität zu beurteilen ist (z. B.: ein Didaktikkurs kann gut oder schlecht sein, intensiv oder kaum besucht werden).

#### V. Betreuung von Studierenden/Administration des Studienprozesses

In der Massenuniversität kann man nicht voraussetzen, daß Lehre und Betreuung von Studierenden Hand in Hand gehen. Dazu sind die Lehrveranstaltungen zu groß. Der einzelne Studierende wird in der Massenuniversität leicht isoliert. Man muß verhindern, daß Studenten scheitern, weil sie schlecht informiert sind. Maßnahmen zur Betreuung von Studierenden umfassen etwa: Informationsveranstaltungen oder Broschüren für Schüler zur Vorbereitung auf das Studium, für Studienanfänger und zur Auswahl von Speziellen Betriebswirtschaftslehren und Wahlfächern sowie Studienberatung durch spezialisierte Stellen, aber auch durch die Lehrenden selbst, was u. a. über deren Ansprechbarkeit erfaßt werden kann, spezifische Beratungsangebote für Problemgruppen wie Langzeitstudierende, Studenten mit Prüfungsphobien usw.

Zu den Aktivitäten zur Administration des Studienprozesses gehören weiterhin: zügige Korrektur von Prüfungsarbeiten, Sicherstellung der rechtzeitigen und umfassenden Aushängung von Seminar- und Diplomarbeitsthemen an Kandidaten, Sicherstellung eines ausreichenden Angebots an Lehrveranstaltungen, das einen reibungslosen Studienverlauf ermöglicht, Sicherstellung des Zugangs zu prüfungsrelevanten Seminaren und Übungen, ein ausreichend großes Angebot an Plätzen in Veranstaltungsräumen, möglichst frühe Wiederholungsmöglichkeiten für Studierende, die bestimmte Prüfungen nicht bestanden haben, usw.

#### VI. Wirtschaftlichkeit

Ein Beurteilungskriterium der Effizienz des Mitteleinsatzes erfordert zunächst die Ermittlung der verfügbaren Mittel – eine Anforderung, die schon angesichts der engen Verzahnung von Forschung, Lehre und Verwaltung nicht immer leicht zu erfüllen ist. Neben

der Frage, ob verfügbare Mittel überhaupt für die Lehre eingesetzt werden, ist die Wirksamkeit des Einsatzes – die Effektivität und Effizienz – zu klären. Effektivität liegt vor, wenn ein Mitteleinsatz in Übereinstimmung mit didaktischen Prinzipien erfolgt. Bei der Effizienz ist bei einer als gegeben unterstellten Qualität das Verhältnis zwischen Output und Input beim Wissenstransfer zu analysieren (zur Evaluation der Wirtschaftlichkeit s. Gerlich 1993, S. 47 ff.).

## E. Probleme und Entwicklungsbedarf von Evaluationen nach dem niederländischen Modell

### I. Zielkonflikte und unvollständiger Konsens über Voraussetzungen guter Lehre

Evaluationen der Lehre an Fachbereichen nach Maßgabe des holländischen Modells haben vor allem mit drei Problemen zu kämpfen: (1) Es gibt *keinen Konsens darüber, was gute Lehre ausmacht*. (2) Die *Ziele* der Evaluation stehen *teilweise im Konflikt* zueinander. (3) Die Methoden sind weder valide noch zuverlässig.

Daß ein Konsens weder in der Gruppe der Evaluatoren noch in der Gruppe der Evaluierten und erst recht nicht zwischen diesen beiden Gruppen zu erwarten ist, hat unsere Entwicklung eines Modells der universitären wirtschaftswissenschaftlichen Ausbildung gezeigt. Die Evaluationspraxis an deutschen Hochschulen umschifft dieses Problem – und auch das Problem, daß eine detaillierte Erfassung der Lehrprozesse in einer Evaluation gar nicht zu schaffen ist –, indem sie sich *auf die indirekten Funktionen der Lehre – auf Aktivitäten zur Qualitätssicherung, zur Betreuung von Studierenden und zur Administration von Lehrprozessen – konzentriert*. Diesem Vorgehen liegt die *Annahme zugrunde, daß intensive indirekte Aktivitäten auf gute Lehre schließen lassen*. Diese Annahme ist nicht unproblematisch. Es ist durchaus möglich, solche indirekten Aktivitäten als Fassaden aufzubauen (im Sinne des institutionalistischen Ansatzes, s. Meyer/Rowan 1977). U. U. sind sie auch ein Indikator für Probleme im Kernprozeß der Lehre: Eine spezifische Beratung für Langzeitstudierende wird z. B. erforderlich, weil das System viele Langzeitstudierende produziert. Ein Angebot in didaktischer Unterweisung für Lehrende ist u. U. ein Respons auf gravierende Probleme in diesem Bereich.

Auch auf Konflikte zwischen den Zielen der Evaluation haben wir bereits hingewiesen. Wenn Ministerien aus Evaluationen Daten für die Zuweisung von Finanzmitteln und für andere Maßnahmen erhalten wollen, leidet zwangsläufig die Offenheit der internen Evaluation und damit der Druck im Hinblick auf Verbesserungen. *Taktisches Verhalten* wird unweigerlich an den Tag gelegt. In den Niederlanden wird das Aufkommen von taktischem Verhalten bereits beklagt (Vroeijenstijn, A. I. 1991, bes. S. 373 f.).

Evaluationen für folgenreiche externe Entscheidungen (durch Ministerien) setzen eine hohe Validität und Zuverlässigkeit voraus. Die bisher in den Pilotprojekten eingesetzten Verfahren können diese nicht in Anspruch nehmen. Panels aus in der Lehreevaluation nicht geübten Peers kommen ausgehend von Selbstevaluationen, die mit Hilfe schwach strukturierter Leitfäden von mehr oder weniger motivierten Fakultätsangehörigen erstellt wurden, und auf der Basis von relativ kurzen Hearings mit verschiedenen, mehr oder minder zufällig zusammengesetzten Gruppen zu Bewertungen und Empfehlungen. Wir erinnern

uns: „interpanel reliability is poor“ (Scriven 1993, S. 50). Folgeschwere Entscheidungen, Zuteilung von Finanzmitteln etwa, lassen sich auf solche Verfahren nicht gründen.

*Es wird deshalb darauf ankommen, die Evaluation der Lehre von Fachbereichen im Rahmen einer Meta-Evaluation einer laufenden Überprüfung zu unterziehen.* Die Vorstellung, daß zu einem gewissen Zeitpunkt ein „überprüftes und mängelfreies“ Verfahren vorliegt, ist dem Gegenstand nicht angemessen.

*Weil das Instrumentarium nicht präzise ist, kommen auch der Legitimation und der Akzeptanz der Evaluatoren und damit dem Verfahren ihrer Auswahl besondere Bedeutung zu.* Kompetenz, Erfahrung und Integrität der Evaluatoren können die unausweichlichen Mängel des Verfahrens zum Teil kompensieren.

Sehr wohl angemessen sind Evaluationen vom Typ Holland jedoch, um Reformdruck in den Fachbereichen zu erzeugen. Hier können die Anforderungen an Validität und Zuverlässigkeit reduziert werden. Denn: Stellen die Fachbereiche in dem Reformprozeß fest, daß sich in die Evaluation Fehler eingeschlichen haben oder empfohlene Maßnahmen nicht greifen bzw. unintendierte Nebenwirkungen haben, sind in einem internen Diskussionsprozeß eher Korrekturen möglich als bei Eingriffen des Ministeriums. Diese Dialoge innerhalb von Fachbereichen werden denn auch als wesentliches Ergebnis der von Hochschulen selbst initiierten Evaluationen bezeichnet (Lüthje 1995 a).

## II. Indikatoren

Einige Pilotprojekte – bspw. die Projekte der Hamburger und Kieler Universität, des Wissenschaftsrats und des Landes Hessen – übernahmen vom niederländischen Modell eine Reihe von Indikatoren wie etwa Zahl der Absolventen, Verhältnis von eingesetzten Sach- und Personalmitteln zum Output an Absolventen, Durchfallerquoten und Studienabbrecher. In den Niederlanden werden *Indikatoren dieser Art zur Allokation von Finanzmitteln herangezogen*: Ca. 23% der Mittel werden den Hochschulen als „lump sum“ gewährt; „die restlichen 77% der im gradualen und postgradualen Bereich benötigten Gelder werden in Abhängigkeit von Studentenzahl und Absolventenquoten, also z. T. von input- bzw. outputorientierten Daten nach dem sog. Plaatsen-Geld-Model zugewiesen. Danach erhält die Hochschule die Mittel für das Studienfach entsprechend der Anzahl der eingeschriebenen Studenten. Für das dreijährige doctoraal-studium werden für jeden erfolgreichen Absolventen die Aufwendungen für 4,5 Studienjahre und für jeden dropout die Aufwendungen für nur 1,3 erstattet.“ (Richter 1991 b, S. 89). Es ist zu befürchten, daß bei Evaluationen ermittelte Daten auch in Deutschland zu Zwecken der Mittelzuweisung genutzt werden, selbst wenn dies von allen Beteiligten immer wieder zurückgewiesen wird.

Zu verlockend ist in Zeiten knapper Kassen das Argument, daß einzelne Fachbereiche in Abhängigkeit von ihren Leistungen ihre Finanzmittel sogar noch erhöhen können. Das Gesamtbudget über alle Universitäten eines Landes ließe sich dann immer noch senken. Bereits jetzt werden Sachmittel und Mittel für studentische und wissenschaftliche Hilfskräfte in Rheinland-Pfalz und in Nordrhein-Westfalen nach Kennzahlen, basierend auf Studenten- und Absolventenzahlen, Promotionen und Drittmitteln, verteilt. In Schleswig-Holstein wird ein derartiges Modell gemeinsam zwischen Hochschulen, Landesregierung und dem Centrum für Hochschulentwicklung erarbeitet. Im Prinzip können Mittelzuweisungen

auf der Basis von Indikatoren auch ohne Evaluationen erfolgen. Werden die auf der Basis von Indikatoren erzeugten Daten jedoch in Evaluationen eingebracht, erhalten sie dadurch zusätzliche Legitimation. Natürlich können Evaluationskommissionen auch etwas zu den Ursachen von Abweichungen gegenüber dem Mittelwert aussagen. Es bleibt dann aber abzuwarten, ob solche qualitativen Relativierungen Einfluß auf die Zuweisungen haben.

Das Hauptproblem quantitativer Indikatoren liegt darin, daß universitärer Output – als öffentliches Gut – nicht eindeutig bewertet werden kann (die folgende Argumentation basiert auf (österreich.) Bundesministerium für Wirtschaft und Forschung 1993, S. 38 f.). Die Indikatoren, mit denen man sich behelfen muß, schaffen insofern Probleme, als von ihnen nicht nur erwünschte Anreizwirkungen ausgehen können. Um das Setzen von falschen Anreizen, für das es im öffentlichen Bereich viele Beispiele gibt, zu verhindern, benötigt man harte und eindeutige Kriterien, die eine echte Steuerungsfunktion erfüllen und nicht zu anfällig sind gegenüber „rent seeking“ verschiedener Beteiligter. Gleichzeitig müssen solche Kriterien auch flexibel genug sein, um *erwünschte* Anpassungs- und Ausweichreaktionen zuzulassen, ja sogar zu induzieren. „Sachlich berechnete“ Abweichungen, die der Steuerungsabsicht entsprechen, müssen zugelassen sein. Auch sollte eine Überkomplexität des Evaluationssystems vermieden und auf eine Überschaubarkeit des Kriterienkatalogs geachtet werden.

Diesen Anforderungen bei der Konstruktion von Kriterien zu genügen, ist extrem schwierig. Nehmen wir an, es stehen valide und zuverlässige Indikatoren zu (1) Output von Absolventen, (2) Studienerfolgsquoten, (3) Abbrecherquoten und (4) durchschnittlicher Studiendauer zur Verfügung. Stellt man Anpassungsverhalten in Rechnung, zeigt sich, daß diese Indikatoren kaum eine sinnvolle Bewertung der Lehrqualität gestatten. Denn ein Fachbereich kann, um im Hinblick auf diese Indikatoren gut abzuschneiden, Belastungen und Prüfungsanforderungen reduzieren. Solange Daten vom Arbeitsmarkt (z. B. Anfangsgehälter, durchschnittliche Wartezeit vom Examen bis zur ersten Beschäftigung) nicht vorliegen, gibt es keinen Indikator, der die unerwünschten Nebeneffekte aufzeigt. Ein Fachbereich kann jedoch auch mit extrem schweren frühen Prüfungen, scharfen Vorgaben von zulässigen Höchstsemestern für Zwischenprüfung und Examen sowie mit einer Reduzierung der Wiederholungsmöglichkeiten von Prüfungen auf diese Indikatoren reagieren, um nur die leistungsfähigsten Studierenden anzuziehen – zumindest solange eine starke Nachfrage nach einem Fach besteht –, wodurch dann ebenfalls unerwünschte Nebeneffekte produziert werden (vor allem Externalisierungen auf andere Fachbereiche und „gescheiterte Existenzen“) (s. a. Bormans et al. 1987, S. 189).

Im Grunde können mehr qualitativ orientierte Evaluationen durch Peers Korrektive zu den schlichten quantitativen Steuerungen generieren. Indem Qualitätssicherungsprozesse durch Evaluationen in Gang gesetzt werden, kann etwa der oben dargestellten Niveauabsenkung zur Verbesserung der Absolventen/Ausstattungs-Relation entgegengewirkt werden.

Generell erscheinen Indikatoren, die den Beschäftigungswert der Hochschulausbildung in Abhängigkeit vom Einkommen des Absolventen, seinem sozialen Status und der Länge seiner Sucharbeitslosigkeit bestimmen, vorzuziehen (Timmermann 1983). Unterstellt wird bei einem solchen Vorgehen freilich, daß der Markt die Leistungsfähigkeit der Absolventen einigermaßen zutreffend beurteilt. Solche objektivierbaren Daten darüber, wie gut Absolventen vom Arbeitsmarkt aufgenommen werden, können kaum durch An-

passungsverhalten der Beteiligten unterlaufen werden. Sie sind neutral, weil sie jede Form der Qualifikation, welche am Markt Wertschätzung erfährt, gleich behandeln. „Diese Neutralität ist verknüpft mit einer Entlastung der steuernden Instanz von einer Fülle von Informationsbedarf. Insgesamt imitieren solche ökonomischen Indikatoren weitgehend einen marktlichen Steuerungsmechanismus und weisen daher ganz ähnliche Vorteile auf wie der Markt.“ (Bundesministerium für Wissenschaft und Forschung 1993, S. 44).

Leider liegen Indikatoren dieser Art noch nicht vor. Es ist auch fraglich, in welchem Umfang der Datenschutz sie zuläßt. In diesem Zusammenhang sind auch Rankings nach amerikanischem Vorbild relevant, da auch sie Bewertungen von Märkten widerspiegeln. Es wäre bedauerlich, wenn in der derzeitigen Evaluationseuphorie diese Form der Evaluation durch den Markt aus dem Blick gerieten.

### III. Leitfaden/Lehrbericht

Es ist zu erwarten, daß *die Berichte der Fachbereiche, die diese auf der Basis eines Leitfadens erstellen, unterschiedliche Qualität aufweisen*. Erste Erfahrungen mit den Pilotprojekten, insbesondere mit dem des Wissenschaftsrats, bestätigen diese Annahme ebenso wie die Praxis in den Niederlanden: „Etwa 10% der Berichte enthalten neben einer Beschreibung auch eine klare Analyse und zeigen ein Verständnis für Selbstbeurteilung“ (Vroeijenstijn, A. I. 1991, S. 373). Die Fachbereiche verwenden nicht nur unterschiedliche Sorgfalt auf die Verfertigung des Berichts, sondern setzen auch unterschiedliche Organisationskonzepte ein: Eine Gruppe von Professoren erstellt bspw. den Bericht oder eine einzelne Person (Professor oder Dekan). Mitarbeiter der Universitäts- und Dekanatsverwaltung werden als Datenlieferanten oder Koautoren beteiligt. Manchmal werden Studenten und Assistenten beteiligt, manchmal nicht. Dies hat Auswirkungen auf das Ergebnis der Selbstevaluation.

Es ist weiter anzunehmen, daß *die Items/Fragen des Leitfadens unterschiedlich interpretiert werden und nicht in allen Fällen geeignet sind, die Besonderheiten des jeweiligen Fachbereiches hervortreten zu lassen*. Fachbereiche, die sich eng an den Leitfaden halten, produzieren mit großer Wahrscheinlichkeit ziemlich unzusammenhängende Darstellungen mit vielen Redundanzen. Häufig geht die Intention der Fragestellung verloren. Die Lösung dieses Problems liegt u. E. nicht in einer stärkeren Detaillierung und Strukturierung des Leitfadens, sondern in einer Aufforderung an den Fachbereich, zu den verschiedenen Problemkomplexen in einer freien Form Stellung zu nehmen. Z. B.: Statt vieler Fragen zur Beratung von Studierenden etwa folgende Frage: Welches Konzept der Information von Studenten über das Studium verfolgen Sie, und wie haben Sie dieses Konzept umgesetzt? Solche Fragen würden es dem Fachbereich erlauben, seine spezifische Lösung darzustellen und zu erläutern. Der Beantworter wäre nicht gezwungen, sein Konzept nach einem Fragenkatalog darzustellen (und zwangsläufig schlecht aussehen zu lassen), der implizit von einem ganz anderen Konzept ausgeht.

Gerade die Breite und Strukturierung des Leitfadens erschweren tendenziell eine Selbstevaluation. Die Mitglieder des Fachbereichs sind vollauf damit beschäftigt, alle Fragen zu beantworten und alle Daten heranzuschaffen und kommen gar nicht zum Reflektieren. In den Niederlanden wurde übrigens ein ähnlicher Eindruck gewonnen: „Die Mehrzahl

der Selbstuntersuchungen hinterläßt den Eindruck, lediglich für die Untersuchungskommission geschrieben zu sein, statt für eine kritische Selbstreflexion. Im allgemeinen benutzen die Fakultäten die Checkliste der VSNU ... Es hat den Anschein, daß die Mehrzahl der Fakultäten noch nicht die Grundsatzfragen für eine gute Selbstbeurteilung formuliert hat.“ (Vroeijenstijn, A. I. 1991, S. 373).

Ein Problem des Leitfadens ist auch, daß nach Forschung kaum gefragt wird. *Im Selbstverständnis der Hochschule setzt gute Lehre gute Forschung voraus.* Fragen nach der Forschung würden folglich die Akzeptanz der Evaluation erhöhen. Allerdings: Die Evaluation der Forschung ist ein schwieriges Problem, zu dessen Lösung sich die Niederländer einige Jahre Zeit genommen haben, und nicht ohne Grund evaluieren sie die Forschung *getrennt* von der Lehre – in einem separaten Verfahren und mit anderen Evaluatoren.

Ein weiteres Problem liegt darin, daß, weil der Leitfaden ja fächerneutral sein soll, in den vorliegenden Versionen *nach der Lehre in den Speziellen Betriebswirtschaftslehren nicht gefragt* wird. Eine solche Analyse wäre einerseits sehr aufwendig, andererseits ist aber nicht zu vernachlässigen, daß die Ausbildung in den Speziellen Betriebswirtschaftslehren das Bild der Praxis über die Qualität der Ausbildung weit stärker bestimmt als die Ausbildung im Grundstudium oder in der Allgemeinen Betriebswirtschaftslehre.

Die Konzentration auf Aktivitäten der Betreuung und der Prozeßadministration hat zur Folge, daß Fachbereiche, die „sich intensiv um ihre Studenten kümmern“, gut abschneiden. Die Annahme, daß Studenten, die sich und ihr Studium gut selbst organisieren können, die praxistauglicheren Absolventen werden, ist jedoch nicht ganz von der Hand zu weisen. Es gibt vielleicht so etwas wie angemessene Informations- und „Betüterungs“grade, die man nicht überschreiten sollte.

Auch verschafft u. E. der Leitfaden *kleinen Fachbereichen Vorteile*: Diese sind für die Kommission transparenter, ihre Komplexität erscheint eher handhabbar.

#### IV. Der Prozeß der Evaluation

Es ist anzunehmen, daß in den Fragen der Evaluatoren im Rahmen der Hearings *unterschiedliche implizite „Modelle der guten Lehre“* zum Ausdruck kommen. Jedes Kommissionsmitglied wird sich über die verschiedenen Befragungen hinweg einen eigenen Fragenkatalog aufbauen. Fragen, die sich bewährt hatten, werden mit großer Wahrscheinlichkeit wiederholt. Es ist daher anzunehmen, daß das Frageverhalten der Kommission in einem hohen Maße von ihrer Zusammensetzung abhängt.

Im Projekt des Wissenschaftsrats konzentrierte sich die Diskussion in den verschiedenen Fachbereichen auf je spezifische Problemfelder (in München: Studien- und Prüfungsreform, in Siegen: Probleme von integrierten Studiengängen in Gesamthochschulen, in Köln: Betreuung und Information von Studierenden), wodurch andere Problembereiche zu kurz kamen. Mehr Zeit für die Begehung und/oder eine Diskussionsführung, die auf ein Vorgehen streng nach dem Leitfaden achtet, könnten diesem Problem entgegenwirken, bewirken aber andererseits einen gewissen Schematismus.

Es ist weiter zu befürchten, daß die *Evaluierungskommission frühzeitig bestimmte Deutungsmuster entwickeln wird*, um die Komplexität eines Fachbereichs „in den Griff zu be-

kommen“ (was sich dann im nachhinein vielleicht als Würgegriff herausstellt). Wenn man als Professor die Fakultät wechselt, dauert es eine ganze Weile, bis sich einem die Regelungen und die Kultur der neuen Umgebung erschließen. Es ist für eine Kommission wohl extrem schwierig, diesen Prozeß abzukürzen.

Weiter ist anzunehmen – und die Erfahrungen mit den ersten Pilotprojekten bestätigen diese Annahme –, daß der *Verlauf der Hearings stark abhängig ist von der Kritikfähigkeit der Studenten und Assistenten*. Mängel werden von der Kommission eher entdeckt, wenn diese Gruppen auf solche aufmerksam machen. Sollte die Aufdeckung von Problemen ernsthafte Konsequenzen nach sich ziehen, ist mit taktischem Verhalten der verschiedenen Gruppen zu rechnen (Überbetonung bestimmter Probleme, „Deals machen“, kritische Leute fernhalten usw.).

Zeitigen die Hearings mit den verschiedenen Gruppen *widersprüchliche Informationen, so sollte eine Klärung unter Einschluß aller Beteiligten* angestrebt werden. In den Evaluationen des Verbunds norddeutscher Universitäten wurde dieses Problem gelöst, indem der Bericht der Evaluationskommission den Fachbereichen zur Korrektur vorgelegt wurde (Christian-Albrechts-Universität zu Kiel/Universität Hamburg 1995). Taktisches Verhalten ist bei dieser Lösung allerdings wahrscheinlicher als bei einer Klärung noch während der Begehung in Form einer Konfrontation der sich widersprechenden Gruppen unter Moderation der Evaluatoren, was allerdings wiederum viel Zeit kostet.

Das *mündliche Feedback am Ende der Begehung* muß bei der Kürze der zur Analyse zur Verfügung stehenden Zeit notwendigerweise unbefriedigend für alle Beteiligten ausfallen. Vor allem ist ein sorgfältiger Vergleich des betreffenden Fachbereichs mit anderen Fachbereichen nicht möglich. Aus diesem Grund sind im norddeutschen Verbund zweitägige Auswertungskonferenzen zwischen Gutachtern, Fachvertretern, Studierenden und Hochschulleitungen im Anschluß an die Evaluationen durchgeführt worden (Lüthje 1995 a).

Es sollte möglich sein, daß die *Kommission auch Angehörige der Universität befragt, die nicht von den Initiatoren des Fachbereichs ausgewählt wurden*. In den Pilotprojekten war dies u. W. nicht vorgesehen. Generell sollte allen Beteiligten klar sein, daß die Kommission Herrin des Verfahrens ist.

## V. Nutzung von Evaluationsergebnissen

Die hochschulinterne Nutzung externer Evaluationsergebnisse erschließt sich in ihrer Bedeutung, wenn man sie als ergänzendes Element in einem vorrangig nach dem Prinzip der akademischen Selbstkontrolle und -organisation gestalteten System betrachtet. Dabei sind zwei Merkmale der Funktionsweise der Universitäten zu berücksichtigen: (1) Die Notwendigkeit, Personen mit hohem Professionalisierungsgrad beachtliche Handlungsautonomie einzuräumen und auf berufsständisch (berufsgruppenorientiert) abgesicherte Selbstkontrolle zu setzen, führt zu *hierarchiearmen Strukturen*. (2) Die Universität handelt primär auf der Ebene der Fachbereiche; einzelne Wissenschaftler und Gruppen von Wissenschaftlern sind die wichtigsten Träger ihrer Aktivitäten. Die Anpassung an sich ändernde Anforderungen im wissenschaftlichen und gesellschaftlichen Umfeld, die Generierung und Umsetzung von Innovationen, vollzieht sich zum größten Teil in den Fakultäten. Die Orientierung ist dabei in hohem Maße fächerbezogen und hochschulübergreifend.

Vor diesem Hintergrund bedeutet die externe Evaluation die Einbringung von Handlungsdruck in ein in hohem Maße durch Autonomie gekennzeichnetes System. Externe Evaluationen verschärfen (insbesondere bei Veröffentlichung vergleichender Evaluationen) den Wettbewerb zwischen den Hochschulen um Reputation und schränken die in hierarchiarmen, auf Selbstorganisation beruhenden Strukturen bestehende Tendenz zur *organizational inertia* ein.

Erste Erfahrungen im Verbund norddeutscher Universitäten zeigen, daß neben den Fachbereichen selbst die Hochschulleitungen eine besondere Verantwortung haben, die Ergebnisse der Evaluationen nicht „versanden“ zu lassen. Zu den Aufgaben der Hochschulleitung gehört in diesem Zusammenhang die Einforderung der korporativen Verantwortung des Fachbereichs sowie die Initiierung und Aufrechterhaltung des Kommunikationsprozesses im Fachbereich. Instrumente hierzu sind runde Tische, Auswertungskonferenzen sowie protokollierte Vereinbarungen zwischen dem Fachbereich und der Hochschulleitung mit Maßnahmen- und Zeitplanungen für notwendige oder angeregte Änderungen. Hierbei können auch Mittelzuwendungen zum Abbau von Schwachstellen oder gegebenenfalls Mittelkürzungen in Aussicht gestellt werden, sofern bestimmte Maßnahmen nicht realisiert werden.

Das Interesse staatlicher Institutionen an Evaluationsergebnissen und die Förderung von Evaluationsvorhaben sind legitim und notwendig. Sie sind legitim, weil der Staat (neben der Universität selbst) die Verantwortung für die Funktionsfähigkeit der Hochschule hat. Sie sind notwendig, weil es nicht von der Hand zu weisen ist, daß die gegenwärtigen Rahmenbedingungen die Fähigkeit zur Selbstkontrolle in den Universitäten beeinträchtigen. In einer Zeit des gesellschaftlichen und wirtschaftlichen Umbruchs deckt der verschärfte internationale Wettbewerb selbst in bisher weltweit führenden Unternehmungen Defizite in der Leistungsfähigkeit auf. Die Vorstellung, die Universität würde hier eine Ausnahme machen, ist unrealistisch. *Angesichts der Autonomie der Hochschulen haben staatliche Stellen allerdings keine unmittelbare Handlungskompetenz. Ihre Einwirkungsmöglichkeit über die Förderung externer Evaluationen sind indirekter Natur. Sie können den Wettbewerb um Reputation zwischen den Hochschulen verstärken und auf die Beseitigung aufgedeckter Schwachstellen hinwirken.*

## F. Empfehlungen

- (1) *Ziele der Evaluation:* Evaluationen von Lehre und Forschung sind eine permanente Aufgabe der Universität. Das dominierende Ziel von Evaluationen sollte die Erzeugung von Handlungsdruck in den Fachbereichen sein. Schon angesichts der geringen methodischen Zuverlässigkeit ist die Verwendung von Evaluationsergebnissen zur Unterstützung ministerieller Entscheidungen, insbesondere Finanzentscheidungen, und zur Information der Öffentlichkeit höchst problematisch (diese Forderung erhebt auch Vroeijensteijn, T. (1992) auf der Basis der niederländischen Erfahrungen).
- (2) *Ziele in der Evaluation:* Die Fachbereiche sollten an ihren eigenen Zielen und Konzepten gemessen werden und nicht an expliziten oder impliziten Modellen der für die Erstellung der Checklisten Verantwortlichen oder der Evaluatoren. Die Leitfäden, die

gegenwärtig Verwendung finden, sind in diesem Sinne verbesserungsbedürftig. Empfehlenswert ist insbesondere ein geringerer Detaillierungsgrad der Leitfäden. Die Fachbereiche sollten aufgefordert werden, die Verbindung zwischen ihrer Forschung und Lehre aufzuzeigen. Nur so lassen sich die Besonderheiten universitärer Studiengänge in angemessener Weise erfassen und beurteilen.

- (3) *Evaluationsmethode*: Externe Evaluationen sind ein besonders wirksames Instrument, durch Einbringung externer Fachkompetenz die Selbstkontrolle in den Hochschulen zu stärken. Allerdings wird schon der vergleichsweise hohe Aufwand eines externen Begutachtungsverfahrens dieser Form der Evaluation den Charakter eher flankierender, unregelmäßigen Maßnahmen zuweisen. In großem Umfang sollten externe Evaluationen erst durchgeführt werden, wenn die Ergebnisse der bisherigen Pilotprojekte sorgfältig ausgewählt wurden und wenn die institutionellen Voraussetzungen für eine seriöse Beurteilung gegeben sind.
- (4) *Träger der Evaluation*: Der Aufbau einer externen Lehrevaluation erfordert eine neutrale Trägerorganisation, die mit Fachpersonal und ausreichenden Ressourcen auszustatten ist. An der Evaluation sollten Vertreter der zuständigen Ministerien nicht unmittelbar beteiligt sein. Der Auswahl von Evaluatoren kommt besondere Bedeutung zu. Berufsständische Vereinigungen von Hochschullehrern erscheinen als geeignete Institutionen, den Auswahlprozeß zu unterstützen.
- (5) *Evaluationsprozeß*: Die Qualität der Evaluation wird entscheidend von den verfolgten Prinzipien bei der Erhebung und Überprüfung der Daten bestimmt. Die Evaluationskommission sollte Gelegenheit haben, auch Personen und Gruppen zu befragen, die nicht offiziell als Gesprächspartner benannt wurden. Widersprüchliche Aussagen sollten nach Möglichkeit noch während der Begehung unter den Betroffenen durch Austausch der verschiedenen Auffassungen geklärt werden. Um Fehler zu vermeiden und Fehleinschätzungen vorzubeugen, ist dem Fachbereich Gelegenheit zu geben, vor der entgeltigen Fassung des Berichts Stellung zu beziehen. Der Evaluationsbericht wendet sich vor allem an den beurteilten Fachbereich. Für eine Berichterstattung an Dritte kommt nur ein Teil des Berichts in Frage.

## Literatur

- Alkin, M. C. (1972): Die Aufwands-Effektivitäts-Evaluation von Unterrichtsprogrammen, in: Wulf, C. (Hrsg.): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen, München, S. 146–165.
- Banta, T. W. (1992): Aktuelle Verfahren der Effektivitätsbewertung im amerikanischen Hochschulwesen, in: Holtkamp, R.; Schnitzer, K. (Hrsg.): Evaluation des Lehrens und Lernens. Ansätze, Methoden, Instrumente, Hannover, S. 65–92.
- Beeby, C. E. (1977): The meaning of evaluation, in: Department of Education (Hrsg.): Current Issues in Education. No. 4: Evaluation, Wellington, S. 68–78.
- Bormans, M. J.; Brouwer, R.; In't Veld, R. J.; Merten, F. J. (1987): The role of performance indicators in improving the dialogue between government and universities, in: International Journal of Institutional Management in Higher Education, 11. Jg., S. 181–194.
- Bundesministerium für Wissenschaft und Forschung (1993): Evaluation im Hochschulwesen, Wien.

- Christian-Albrechts-Universität zu Kiel; Universität Hamburg (Hrsg.) (1995): *Gemeinsamer Abschlußbericht der Universitäten Hamburg und Kiel zum Pilotprojekt Evaluation in den Studienfächern Biologie und Germanistik*, Hamburg/Kiel.
- Cook, T. D.; Reichardt, C. S. (Hrsg.) (1979): *Qualitative and Quantitative Methods in Evaluation Research*, Beverly Hills, CA.
- Cronbach, L. J.; Ambron, S. R.; Dornbusch, S. M.; Hess, S. M.; Hornik, R. C. et al. (1980): *Towards Reform of Program Evaluation*, San Francisco.
- Daniel, H. (1994): Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung, in: Winkler, H. (Hrsg.): *Kriterien, Prozesse und Ergebnisse guter Hochschulausbildung. Dokumentation eines Kolloquiums an der Universität Gesamthochschule, Kassel*, S. 31–49.
- De Landsheere, V. (1990): Taxonomies of educational objectives, in: Walberg, H. J.; Haertel, G. D. (Hrsg.): *The International Encyclopedia of Educational Evaluation*, Oxford/New York et al. S. 179–188.
- Ewell, P. T. (1992): Lehrevaluation in the USA – ein Wegweiser durch die Vielfalt neuer Assessment-Ansätze, in: Holtkamp, R.; Schnitzer, K. (Hrsg.): *Evaluation des Lehrens und Lernens. Ansätze, Methoden, Instrumente*, Hannover, S. 51–64.
- Frederiks, M. M. H.; Westerheijden, D. F.; Weusthof, P. J. M. (1994): Effects of quality assessment in Dutch higher education, in: *European Journal of Education*, 29. Jg., S. 181–199.
- Gerlich, P. (1993): *Hochschule und Effizienz. Anstöße zur universitären Selbstreflexion*, Wien.
- Geschäftsstelle des Wissenschaftsrates (1994): *Leitfaden zur Evaluation von Studium und Lehre*, Köln.
- Guba, E. G.; Lincoln, Y. S. (1989): *Fourth Generation Evaluation*, Newbury Park, CA.
- HAYEK Engineering AG (1993): *Externe Prüfung der Wirtschaftlichkeit in Universitäten des Landes Baden-Württemberg. Abschlußbericht*, Zürich.
- Hessisches Ministerium für Wissenschaft und Kunst (1993): *Der hessische Weg. Materialien zur Studienreform*, Wiesbaden.
- Holtkamp, R.; Schnitzer, K. (Hrsg.) (1992): *Evaluation des Lehrens und Lernens – Ansätze, Methoden, Instrumente*, Hannover.
- HRK – Hochschulrektorenkonferenz (1995): *Entschließung des 176. Plenums vom 3. 7. 1995: Zur Evaluation im Hochschulbereich unter besonderer Berücksichtigung der Lehre*, Bonn.
- Hummel, T. R. (Hrsg.) (1988): *Leistungsorientierte Kennzahlen im Spannungsfeld von Hochschulen und Staat*, Frankfurt/M.
- Kellermann, P. (1992): Ranking- und Review-Verfahren in den Vereinigten Staaten, in: Altrichter, H.; Schratz, M. (Hrsg.): *Qualität von Universitäten*, Innsbruck, S. 212–226.
- Lenn, M. P. (1992): The US accreditation system, in: Craft, A. (Hrsg.): *Quality Assurance in Higher Education. Proceedings of an International Conference Hong Kong, 1991*, London/Washington, S. 161–168.
- Lüthje, J. (1995 a): Interview zur Evaluation im Verbund norddeutscher Universitäten. *Wissenschaftsmanagement*, 1. Jg., Heft 4 (i. Dr.).
- Lüthje, J. (1995 b): Evaluation im Verbund, in: Müller-Böling, D. (Hrsg.): *Qualitätssicherung in Hochschulen*, Gütersloh (i. Dr.).
- Madaus, G. F.; Scriven, M. S.; Stufflebeam, D. L. (Hrsg.) (1983): *Evaluation Models*, Boston.
- Meyer, J. W., Rowan, B. (1977): Institutional organizations: Formal structure as myth and ceremony, in: *American Journal of Sociology*, 83. Jg., S. 340–363.
- Reissert, R. (1994): Evaluation der Lehre – interne Selbstevaluation und externe Begutachtung durch Peers, in: *HIS Kurzinformation A8/94*.
- Richter, R. (1991 a): Verfahren der Qualitätsevaluation in der Lehre an den niederländischen Hochschulen und erste Erfahrungen, in: *HIS – Beiträge zur Hochschulforschung*, 2. Jg., S. 85–111.
- Richter, R. (1991 b): Qualitätsevaluation von Lehre und Forschung an den Universitäten der Niederlande. Eine Bilanz der letzten 10 Jahre, in: Webler, W.; Otto, H. (Hrsg.): *Der Ort der Lehre in der Hochschule*, Weinheim, S. 337–362.
- Rossi, P. H.; Freeman, H. E.; Hofmann, G. (1988): *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*, Stuttgart.

- Sanders, J. R. (1994): *The Program Evaluation Standards*, 2. Aufl., Thousand Oaks, CA.
- Schwandt, T. A.; Halpern, E. S. (1988): *Linking Auditing and Metaevaluation*, Newbury Park, CA.
- Scriven, M. (1980): *The Logic of Evaluation*, Inverness, CA.
- Scriven, M. S. (1983): Evaluation ideologies, in: Madaus, G. F.; Scriven, M. S.; Stufflebeam, D. L. (Hrsg.): *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, Boston, S. 229–260.
- Scriven, M. (1993): *Hard-won Lessons in Program Evaluation*, San Francisco, CA.
- Shadish, W. R. J.; Cook, T. D.; Leviton, L. C. (1991): *Foundations of Program Evaluation. Theories of Practice*, Newbury Park, CA.
- Sluis, A. v. (1992): Qualitätsbeurteilung der Lehre an Hochschulen in den Niederlanden: die Praxis: Bericht aus der Untersuchungskommission Mathematik und Informatik, in: Holtkamp, R.; Schnitzer, K. (Hrsg.): *Evaluation des Lehrens und Lernens. Ansätze, Methoden, Instrumente*, Hannover, S. 19–32.
- Straw, R. B.; Cook, T. D. (1990): Meta-evaluation, in: Walberg, H. J.; Haertel, G. D. (Hrsg.): *The International Encyclopedia of Educational Evaluation*, Oxford/New York et al. S. 58–60.
- Stufflebeam, D. L. (1983): The CIPP model for program evaluation, in: Madaus, G. F.; Scriven, M. S.; Stufflebeam, D. L. (Hrsg.): *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, Boston, S. 117–142.
- Thom, N. (1980): Organisationslehre, Didaktik und Methodik der, in: Grochla, E. (Hrsg.): *Handwörterbuch der Organisation*, Stuttgart, Sp. 1655–1672.
- Timmermann, D. (1983): Ein qualitatives Cost-Benefit-Modell zur kennzifferorientierten Steuerung und Erhöhung der Effizienz der Hochschulen, in: Widmaier, H. P. (Hrsg.): *Das Arbeitskräfteangebot zwischen Markt und Plan*, Berlin, S. 149–164.
- Vroeijenstijn, A. I. (1991): Qualitätssicherung basiert auf Selbstbeurteilung und Expertenkritik: Erfahrungen an niederländischen Universitäten mit der externen Beurteilung, in: Webler, W. D.; Otto, H.-U. (Hrsg.): *Der Ort der Lehre in der Hochschule*, Weinheim, S. 363–383.
- Vroeijenstijn, T. (1992): External quality assessment: Servant of two masters? The Netherlands university perspective, in: Craft, A. (Hrsg.): *Quality Assurance in Higher Education. Proceedings of an International Conference Hong Kong, 1991*, London/Washington, S. 109–132.
- Weick, K. E. (1976): Educational organizations as loosely coupled systems, in: *Administrative Science Quarterly*, 21. Jg., S. 1–19.
- Wolf, R. L. (1990): Judicial evaluation, in: Walberg, H. J.; Haertel, G. D. (Hrsg.): *The International Encyclopedia of Educational Evaluation*, Oxford/New York et al. S. 79–81.
- Wolf, R. M. (1990). The nature of educational evaluation, in: Walberg, H. J.; Haertel, G. D. (Hrsg.): *The International Encyclopedia of Educational Evaluation*, Oxford/New York et al. S. 8–14.
- Wottawa, H.; Thierau, H. (1990): *Evaluation*, Bern.

## Zusammenfassung

Nach einer Einführung in grundlegende Positionen der Evaluationsforschung in den USA und einem Überblick über laufende Projekte zur Evaluation von wirtschaftswissenschaftlichen Studiengängen in Deutschland werden die grundsätzlichen Schwierigkeiten solcher Evaluationen analysiert. Vor allem drei Probleme werden identifiziert: (1) Es gibt keinen Konsens darüber, was unter guter Lehre zu verstehen ist. (2) Evaluationen haben verschiedene Ziele, die untereinander in Konflikt stehen. (3) Die Methoden der Evaluation sind weder valide noch zuverlässig. Aus diesen Schwierigkeiten werden einige Forderungen an zukünftige Evaluationen abgeleitet.

## Summary

After an introduction into different approaches to program evaluation in the U.S. and an overview over current projects on evaluations of university programs in business administration and economics in Germany, the basic difficulties of such evaluations are analysed. Three main problems are identified: (1) There is no consensus on contents and methods of teaching business administration. (2) Evaluations serve different goals that are conflicting. (3) Evaluation methods are neither valid nor reliable. These difficulties lead to some requirements for future evaluations.