



Universität Potsdam

Philipp Pohlenz

## Datenqualität als Schlüsselfrage der Qualitätssicherung von Lehre und Studium an Hochschulen



# Datenqualität als Schlüsselfrage der Qualitätssicherung von Lehre und Studium an Hochschulen



Philipp Pohlenz

**Datenqualität als Schlüsselfrage  
der Qualitätssicherung von  
Lehre und Studium an Hochschulen**

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

**Universitätsverlag Potsdam 2008**

<http://info.ub.uni-potsdam.de/verlag.htm>

Universitätsverlag Potsdam, Am Neuen Palais 10, 14469 Potsdam

Tel.: +49 (0)331 977 4623 / Fax: 4625

E-Mail: [ubpub@uni-potsdam.de](mailto:ubpub@uni-potsdam.de)

Das Manuskript ist urheberrechtlich geschützt.

Elektronisch veröffentlicht auf dem Publikationsserver der Universität Potsdam

URL <http://pub.ub.uni-potsdam.de/volltexte/2008/1975/>

URN [urn:nbn:de:kobv:517-opus-19755](http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-19755)

[<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-19755>]

**ISSN 1862-8664**

Zugleich gedruckt erschienen im Universitätsverlag Potsdam:

ISBN 978-3-940793-48-5

**Datenqualität als Schlüsselfrage der  
Qualitätssicherung von Lehre und Studium an  
Hochschulen**

Dissertation zur Erlangung des akademischen Grades  
Doktor der Wirtschafts- und Sozialwissenschaften

(Dr. rer. pol.)

eingereicht an der Wirtschafts- und  
Sozialwissenschaftlichen Fakultät der Universität Potsdam

von

Dipl.-Soz. Philipp Pohlenz

Potsdam, August 2007

Erstgutachter Prof. Dr. Erhard Stölting

Zweitgutachter Prof. Dr. Helmut Kromrey





# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
<b>2</b>	<b>Hochschulpolitische Rahmenbedingungen</b>	<b>15</b>
<b>3</b>	<b>Evaluation - Forschungsprogramm und Instrument der politischen Praxis</b>	<b>25</b>
3.1	Das Forschungsprogramm der Evaluation . . . . .	26
3.1.1	Evaluationsforschung und Grundlagenforschung . . . . .	27
3.1.2	Wirkungsmessungen und Prozessanalysen . . . . .	28
3.2	Paradigmen der Evaluation . . . . .	33
3.2.1	Rationalistische vs. Naturalistische Evaluation . . . . .	34
3.2.2	Perspektiven für die Versöhnung . . . . .	38
3.2.2.1	Wirklichkeitswissenschaft . . . . .	39
3.2.2.2	Theory-based Evaluation . . . . .	42
<b>4</b>	<b>Ansätze der Qualitätssicherung für die Hochschulentwicklung</b>	<b>45</b>
4.1	Qualitätskultur . . . . .	48
4.2	Evaluation und Ansätze der Hochschulentwicklung . . . . .	52
4.2.1	Partizipative Evaluation . . . . .	53
4.2.2	Evaluation und Lernende Organisationen . . . . .	60
4.2.3	Organisationslernen in der Praxis der Hochschulsteuerung	63
4.2.4	Evaluation als Informationsbasis für Wissensmanagement .	68
4.2.5	„Evidenzbasierte Praxis“ in der Hochschulsteuerung . . . .	71

<b>5</b>	<b>Qualitätssicherung mittels Lehrevaluation und Leistungskennzahlen</b>	<b>75</b>
5.1	Evaluation als Mittel zur Qualitätsentwicklung . . . . .	76
5.1.1	Studentische Lehrveranstaltungs-kritik . . . . .	78
5.1.2	Interne Evaluation auf der Ebene von Studienfächern . . . . .	79
5.1.3	Peer Review . . . . .	81
5.1.4	Follow-up Maßnahmen . . . . .	83
5.2	Quantifizierung von Qualität: Leistungsorientierte Mittelvergabe . . . . .	87
5.2.1	Leistungsindikatoren in der Hochschulsteuerung . . . . .	88
5.2.2	Indikatoren im Leistungsbereich Lehre . . . . .	90
5.2.3	Analyse von Studienbiographien mit Daten der Hochschulstatistik . . . . .	99
5.2.3.1	Datenschutzrechtliche Regelungen . . . . .	101
5.2.3.2	Studienverlaufsstatistiken an der Universität Potsdam . . . . .	103
<b>6</b>	<b>Kriterien für die Aussagekraft von Evaluationsergebnissen</b>	<b>115</b>
6.1	Testgütekriterien . . . . .	117
6.2	Objektivität, Reliabilität, Validität . . . . .	118
6.3	Testgütekriterien und Lehrevaluation . . . . .	123
6.3.1	Konstruktvalidität . . . . .	124
6.3.2	Zeitliche Stabilität von Qualitätsurteilen . . . . .	126
6.3.3	Sachbezogenheit von Qualitätsurteilen . . . . .	128
<b>7</b>	<b>Datenbasis für die Analysen</b>	<b>133</b>
7.1	Das Potsdamer Modell der Lehrevaluation . . . . .	133
7.1.1	Evaluation der Lehrpräsentation . . . . .	134
7.1.2	Qualitätseinschätzungen auf der Ebene von Studiengängen . . . . .	136
7.1.3	Datenerhebungen im Längsschnitt . . . . .	138
7.1.4	Realisierte Einzelstudien . . . . .	144

<b>8</b>	<b>Relevante Dimensionen zur Beurteilung der Studienqualität</b>	<b>147</b>
8.1	Konfirmatorische Faktorenanalyse . . . . .	148
8.2	Fragestellung und Untersuchungsansatz . . . . .	154
8.2.1	Hypothesen . . . . .	155
8.2.2	Ergebnisse . . . . .	156
8.2.2.1	Modellspezifikation . . . . .	156
8.2.2.2	Parameterschätzung . . . . .	157
<b>9</b>	<b>Beziehungen der Qualitätsdimensionen</b>	<b>163</b>
9.1	Lineare Strukturgleichungsmodelle . . . . .	164
9.2	Kausalanalyse zur Studienzufriedenheit . . . . .	166
9.2.1	Hypothesen . . . . .	167
9.2.2	Ergebnisse . . . . .	168
9.2.2.1	Modellspezifikation . . . . .	168
9.2.2.2	Parameterschätzung . . . . .	170
9.3	Stabilität von Qualitätseinschätzungen . . . . .	174
9.3.1	Variabilität im zeitlichen Verlauf der Studienphasen . . . . .	174
9.3.1.1	Hypothesen . . . . .	176
9.3.1.2	Modell und Parameterschätzung . . . . .	177
9.3.2	Längsschnittdaten in der Lehrevaluation . . . . .	182
9.3.2.1	Untersuchungsansatz und Ergebnisse . . . . .	182
9.3.2.2	Modellspezifikation . . . . .	184
9.3.2.3	Parameterschätzung . . . . .	185
9.3.3	Ein Ausblick: Längsschnittdaten und Qualitätsentwicklung	187
<b>10</b>	<b>Rationalität studentischer Qualitätsurteile</b>	<b>188</b>
10.1	Analysen zur Sachbezogenheit studentischer Urteile . . . . .	190
10.2	Erklärung der Determinanten der Studienzufriedenheit . . . . .	191
10.2.1	Einschätzungen auf der Ebene der Studiengänge . . . . .	192
10.2.2	Rückblickende Urteile von Studienabbrechern und Absolventen . . . . .	194

<b>11 Diskussion und Ausblick</b>	<b>199</b>
11.1 Potenzielle Beeinträchtigungen von Validität und Reliabilität . . .	199
11.1.1 Studentische Qualitätsurteile . . . . .	200
11.1.1.1 Dimensionierung und Kontextsensitivität der Ur- teile zur Studienqualität . . . . .	200
11.1.1.2 Verzerrungen durch subjektive Qualitätsurteile .	202
11.1.2 Hochschulstatistische Kennzahlen . . . . .	204
11.2 Evaluation und Hochschulentwicklung . . . . .	206
11.3 Weiterentwicklung von Qualitätssicherungsverfahren . . . . .	208
11.4 Evaluation oder Evaluitis? . . . . .	211

# Tabellenverzeichnis

3.1	Axiome der Evaluationsparadigmen . . . . .	35
5.1	Fallzahlen in den einbezogenen Studienanfängerkohorten . . . . .	106
5.2	Fachwechsel und Fachschwund: das Beispiel VWL (Studienanfängerjahrgang 1999/2000) . . . . .	110
5.3	Fachwechsel und Fachschwund: das Beispiel Volkswirtschaftslehre (alle Studienfälle) . . . . .	113
7.1	Thematische Schwerpunkte der Befragungen . . . . .	140
7.2	Bislang realisierte Befragungen . . . . .	145
8.1	Latente Variablen und Indikatoren (Konfirmatorische Faktorenanalyse; KFA) . . . . .	155
9.1	Endogene latente Variable, Indikatoren (Strukturgleichungsmodell)	168
9.2	Latente Variablen und Indikatoren (Stabilitätsanalyse) . . . . .	183
10.1	Prädiktoren des Gesamturteils Studierender . . . . .	193
10.2	Prädiktoren des Gesamturteils der Studienabbrecher . . . . .	195
10.3	Prädiktoren des Gesamturteils der Absolventen . . . . .	197

# Abbildungsverzeichnis

5.1	Schematische Darstellung sequenzieller Kreuztabellen . . . . .	108
8.1	Modell der konfirmatorischen Faktorenanalyse . . . . .	150
8.2	Pfaddiagramm der KFA; standardisierte Schätzer . . . . .	158
9.1	Modell eines linearen Strukturgleichungsmodells . . . . .	165
9.2	Kausalanalyse der Studienzufriedenheit; standardisierte Schätzer .	171
9.3	Qualitätsurteile in späteren Studienphasen; standard. Schätzer . .	178
9.4	Studentische Qualitätsurteile im zeitlichen Längsschnitt . . . . .	186

# Kapitel 1

## Einleitung

Die Qualitätsbewertung akademischer und wissenschaftlicher Leistungen gehört seit jeher zu den Kennzeichen des Handelns von Hochschulen und den sie bildenden Wissenschaftlern.<sup>1</sup> Forschungsergebnisse und -methoden (aber auch die Leistungen in der Lehre) wurden schon immer dem wissenschaftlichen Publikum zur Diskussion gestellt und haben sich der entsprechenden Kritik stellen müssen. In dieser Kritik durch (mehr oder weniger) fachnahe Kollegen spiegelt sich die Qualität der erbrachten Leistungen. „Es hat also an den Hochschulen immer schon ein Verfahren der Qualitätssicherung gegeben, nämlich: die kollegiale Kritik, die freie Auseinandersetzung unter Fachleuten über die Qualität wissenschaftlicher Leistungen“ (Kreckel 2002).

Auch der Wettbewerb zwischen Hochschulen oder Fakultäten ist nicht erst mit der Einführung neuer Steuerungssysteme zur Bewertung und Belohnung von Hochschulleistungen entstanden, die das Leitbild einer wettbewerbsorientierten und marktadäquat handelnden Hochschule unter den Vorzeichen von „New Governance“ und „New Public Management“ (z.B. Schimank 2004; Nickel 2003) prägen. Vielmehr konkurrierten Hochschulen schon immer, wobei sich die Konkurrenz allerdings eher auf fachliches Prestige und Reputation bezog. Diese blie-

---

<sup>1</sup>Hier wie an jeder anderen Stelle sind mit der Nennung der männlichen Bezeichnung Menschen beiderlei Geschlechts gemeint. Der Verzicht auf die Verwendung beider Formen dient lediglich der sprachlichen Vereinfachung.

---

ben fachintern und wurden innerhalb der Gemeinschaft der Fachkollegen bewertet (Stölting 2002).

Die Steuerung der Hochschulen in Deutschland ist traditionell durch eine starke akademische Selbstorganisation (insbesondere auf der Ebene der einzelnen Professuren) bei gleichzeitiger staatlicher Regulierung auf der Ebene der Institutionen gekennzeichnet (Schimank 2004). Beides soll durch die Einführung der genannten neuen Steuerungsmodelle abgelöst werden, mit dem Ziel, wettbewerbliche Elemente in die Hochschulen einzuführen und die Ebene der Hochschulleitungen zu stärken, womit wiederum zu einer Planungsautonomie auf der Institutionenebene beigetragen werden soll (Schimank 2004).

Eine Voraussetzung für die Einführung marktadäquater Wettbewerbselemente in die Steuerung von Hochschulen ist eine Abbildung der Leistungen, mit denen die Leistungserbringer für sich werben und ihre -möglichst vorteilhafte- Marktposition legitimieren und behaupten können. Eine rein fachinterne und auf akademischer Reputation beruhende Leistungsmessung wird vielfach als nicht angemessen dargestellt. Im Leitbild der Marktvergesellschaftung von Hochschulen werden Transparenz und Rechenschaft gegenüber der Gesellschaft besonders betont (z.B. Müller-Böling 2006). Aus diesem Grund wird versucht, Leistungen anhand von Indikatoren zu messen und sie so transparent und objektivierbar zu machen. Eine Vergleichbarkeit zwischen den Leistungserbringern zum Zweck der Ermittlung von besonders guten und besonders minderwertigen „Anbietern“ soll auf diese Weise erreicht und so die auf die Honorierung guter Leistungen bezogenen Marktkräfte entfesselt werden.

Für eine Beurteilung von Hochschulleistungen haben sich verschiedene Verfahren der Evaluation etabliert (z.B. Kromrey 2000; Kromrey 2001). Diese zielen in unterschiedlicher Weise sowohl auf die interne Qualitätsentwicklung als auch - insbesondere im Falle leistungsbezogener Mittelvergabeverfahren durch Indikatoren trifft dies zu - auf die externe Steuerung und Sanktionierung als exzellent bzw. minderwertig identifizierter Hochschulleistungen.

Herangezogen zur Leistungsmessung werden dabei direkte und indirekte Indikatoren (vgl. Stölting 2002). Direkte Indikatoren beziehen sich auf solche Hoch-



---

schulleistungen, von denen angenommen wird, dass sie einer unmittelbaren und objektiven Messung zugänglich sind. Dies trifft insbesondere auf solche Leistungsbereiche zu, die eine (mehr oder weniger) direkte *Quantifizierung* ermöglichen, wie bspw. die „Absolventenquote“ oder „die Zahl der Studierenden in der Regelstudienzeit“. Indirekte Indikatoren beziehen sich dagegen auf Tatbestände, die einer direkten Quantifizierung nicht zugänglich sind, die aber trotzdem zur Leistungsbeurteilung von Hochschulleistungen herangezogen werden. Als Beispiel seien dazu studentische Zufriedenheitsurteile über die Qualität von Lehre und Studium genannt. Diese stellen bspw. in den vom Centrum für Hochschulentwicklung (CHE) durchgeführten Hochschulrankings<sup>2</sup> eine relevante Informationsquelle zur Leistungsbeurteilung von Hochschulen dar. Sie fußen aber nicht wie die genannten Beispiele auf der Erhebung objektiv messbarer Sachverhalte, sondern auf den subjektiven Qualitätseinschätzungen Studierender. Die Vielgestaltigkeit (hinsichtlich möglicher Variationsquellen und zugrunde gelegter Urteilsdimensionen) dieser Urteile wird möglicherweise nicht angemessen berücksichtigt, wenn sie zu einem oder wenigen globalen Indikatoren zur Messung der studentischen Zufriedenheit verdichtet werden, wie dies vielfach in der öffentlichen Darstellung der entsprechenden Ergebnisse der Fall ist (Berghoff u. a. 2006).

Abgesehen davon, dass auch mit den direkten Indikatoren verschiedene methodologische und auf die entsprechenden Messungen bezogene Probleme<sup>3</sup> verbunden sind, ist zu betonen, dass insbesondere die indirekten Indikatoren - und hier eben auch die genannten studentischen Qualitätseinschätzungen - diskutabel sind, wenn sie für Zwecke des objektivierenden Leistungsvergleichs herangezogen werden.

Studentische Qualitätseinschätzungen haben mittlerweile über Evaluations-

---

<sup>2</sup>Zur Methode des CHE-Hochschulrankings vgl. Berghoff u. a. (2003); Berghoff u. a. (2006). Zu Hochschulrankings allgemein vgl. Bayer (2004) sowie Engel (2001).

<sup>3</sup>Fallstricke, die mit der Objektivierung von Hochschulleistungen durch den Einsatz einer indikatorengestützten Leistungsbeurteilung verbunden sind, werden unten ausführlich diskutiert (vgl. Kapitel 5).

---

verfahren umfangreich Einzug in die Beurteilung von Hochschulleistungen gehalten. Dabei beschränken sie sich mittlerweile nicht mehr auf studentische Beurteilungen der konkreten Lehrpräsentation in Lehrveranstaltungen, sondern werden auch zum Zweck der Beurteilung der Studienqualität insgesamt, bspw. auf der Ebene der Studienfächer und als rückblickende Kritik von Absolventen erbeten.

Insofern sie eine diesbezüglich relevante Rolle spielen, die unter Umständen sogar in der Einflussnahme auf Allokationsentscheidungen gipfelt, ist zu hinterfragen, inwieweit sie geeignet sind, den mit ihrer Erhebung verbundenen Zweck - die zuverlässige und gültige Beschreibung der Studienqualität - zu erfüllen. Die methodologische Frage, die an diesen Anspruch zu knüpfen ist, ist diejenige nach der Reliabilität und Validität von studentischen Befragungsdaten, die zum Zwecke der Lehrevaluation gesammelt werden sowie von herangezogenen Leistungsindikatoren, wie der Absolventen- und Studienabbrecherquote. Diese Frage steht im Zentrum der vorliegenden Arbeit.

Die Validität von studentischen Qualitätsurteilen auf der Ebene einzelner Lehrveranstaltungen war in der Vergangenheit bereits Gegenstand einer breiten Diskussion und vielfältiger diesbezüglicher Forschungen (vgl. z.B. Rindermann 1995a; Rindermann und Amelang 1994). Studentische Urteile, die auf höheren Aggregatebenen, wie derjenigen der Studienfächer, erhoben wurden, sind dagegen noch nicht in vergleichbarem Maße hinsichtlich dieser Kriterien untersucht worden.<sup>4</sup> In der Verfügbarkeit valider Daten zur Beurteilung der Hochschulleistungen wird eine zentrale Forderung gesehen, die verstärkt in die Diskussion um die in den letzten zwei Jahrzehnten eingeführten Bewertungs- und Evaluationsverfahren einzubringen ist: Zur Beurteilung des Systems Wissenschaft sind aus vorliegender Sicht Bewertungskriterien und -maßstäbe anzulegen, die diesem System selbst entstammen, wie nämlich die nach möglichst aussagekräftigen, weil zentralen Testgütekriterien genügenden Datenerhebungen.

---

<sup>4</sup>Zur allmählichen Etablierung von Evaluationsverfahren, die über die einzelnen Lehrveranstaltungen hinausgehen und die (zu evaluierende) Studienqualität auch auf der Ebene der allgemeinen Studiensituation ansiedeln, vgl. bspw. Knapper (2001); Cannon (2001).

---

Die angestrebte Entstaatlichung der Hochschulen zielt auf eine Steigerung ihrer betriebswirtschaftlichen Effizienz und die Beurteilung des Wissenschaftssystems gemäß entsprechender Leistungskriterien wie Nachfrage oder Nützlichkeit (Stölting 2002). Das Anlegen dieser Kriterien für die Beurteilung von Hochschulen oder des Wissenschaftssystems allgemein birgt insofern Konfliktpotenzial, als sie mit den systemimmanenten Beurteilungstraditionen, die ihrerseits auf fachlicher Reputation gründen, in Einklang gebracht werden müssen (a.a.O.). Darüber hinaus sind unerwünschte Effekte denkbar, die sachfremde, weil an Effizienzkriterien, nicht aber am wissenschaftlichen Erkenntnisfortschritt orientierte Entscheidungen bei der Planung und Durchführung wissenschaftlicher Unternehmungen provozieren.

Die Tragweite und die verschiedenen Implikationen der paradigmatischen Veränderungen in der Steuerung und Bewertung von Hochschulleistungen sollen im Rahmen dieser Arbeit nicht bewertet werden. Ziel ist es aber durchaus, darauf aufmerksam zu machen, dass mit der Verwendung von (mehr oder weniger objektivierenden) Indikatoren zur Beurteilung von Hochschulleistungen *methodologische*, auf die entsprechende Datenerhebung bezogene Probleme verbunden sind. Potenzielle Fehleinschätzungen und Fehlinterpretationen sind auf mögliche Bedrohungen der Validität und Reliabilität der Messungen zurückzuführen, die bei der Datenerhebung zu den als relevant erachteten Indikatoren durchgeführt werden.

Insofern die Anwendung interner Leistungsbeurteilungsverfahren im oben beschriebenen Sinne ohnehin zum Hochschulalltag gehörten und gehören, soll aber auch gefragt werden, ob nicht die Anwendung der vglw. jungen indikatorgestützten Qualitätssicherungsverfahren nicht auch Chancen für die Hochschulen bergen. Diese könnten - sofern sie sich auf eine valide und dementsprechend rationale Beurteilungsgrundlage stützen können - zu einer größeren Transparenz der Beurteilung und damit möglicherweise zur Steigerung der öffentlichen Akzeptanz von Hochschulleistungen beitragen. Die Notwendigkeit von Qualitätssicherung und -entwicklung sowie einer Qualitätsorientierung und an erbrachter Leistung ausgerichteten Honorierung an sich scheint unbestritten zu sein. „Es sollte mitt-

---

lerweile selbstverständlich sein, dass dauerhaft knappe staatliche Finanzmittel für die Hochschule nicht mit der Gießkanne oder nach irgendwelchen arbiträren Gesichtspunkten wie etwa ‚Erbhöfen‘ verteilt werden dürfen“ (Schimank 2006, S. 215). Die innerwissenschaftliche Akzeptanz der entsprechend eingesetzten Beurteilungsverfahren hängt aber entscheidend davon ab, ob die eingesetzten Verfahren in der Lage sind, basierend auf einer wissenschaftlichen Kriterien genügenden Leistungsbeurteilung, konsensual getragene Entwicklungsziele zu formulieren und zu deren Erreichung geeignete Strategien zu entwickeln. Von zentraler Bedeutung ist demgemäß die interne Nutzung von Maßnahmen und Verfahren der Qualitätssicherung oder Evaluation. Von dieser hängt in hohem Maße ihr Beitrag zum Gelingen von Reformanstrengungen ab.

Neben der methodologischen Frage nach der Übereinstimmung von Messungen mit einschlägigen Testgütekriterien ist hier ein weiteres wichtiges Anliegen der vorliegenden Arbeit zu sehen. Dieses besteht darin, ungeachtet etwaiger Beeinträchtigungen durch eingeschränkte Validität und Reliabilität zu diskutieren, inwieweit Daten und Auswertungsergebnisse der Lehrevaluation (sowie anderer Qualitätssicherungsverfahren) nachhaltig und an den Entwicklungszielen der Hochschule orientiert, genutzt werden können.

Diesen Anliegen entspricht der Aufbau der Arbeit. Im folgenden zweiten Kapitel werden die Rahmenbedingungen, innerhalb derer sich die skizzierten hochschulpolitischen Veränderungsprozesse abspielen, weiter diskutiert. Dies soll dazu beitragen, das Verständnis für die in den letzten Jahren gestiegene Bedeutung von Qualitätssicherungsverfahren und insbesondere der Evaluation von Lehre und Studium zu stärken.

Evaluation als Forschungsprogramm ist Gegenstand des dritten Kapitels. Grundlegende begriffliche Abgrenzungen von Evaluation werden vorgenommen. Dabei wird auf die (notwendige) Trennung von alltags- und wissenschaftssprachlicher Benutzung des Begriffs eingegangen. In ersterer wird Evaluation vielfach auch mit Alltagsbewertungen gleichgesetzt und nicht auf den Begriff der Evaluation als systematisches Forschungsprogramm zur wissenschaftlichen Begleitung und/oder Bewertung von Programmen, Studiengängen, Lehrveranstaltungen, etc.

---

rekurriert (vgl. z.B. Kromrey 2000). Evaluation hat sich als Begriff für ein Bewertungsinstrument in der politischen Praxis etabliert und wird in unterschiedlich anspruchsvoller Weise verstanden bzw. praktiziert. Es werden daher allgemeine paradigmatische Grundlagen verschiedener Forschungsrichtungen und -ansätze der Evaluation(sforschung) erläutert.

Mit der Frage der Nutzbarkeit von Evaluationsdaten für nachhaltige Reformen oder Strategien der Organisationsentwicklung an Hochschulen beschäftigt sich das vierte Kapitel. Dabei wird auch auf die doppeldeutige Funktion bzw. Nutzung von Evaluation hingewiesen: Einerseits ist es ihr Selbstverständnis, Akteuren der jeweils evaluierten Handlungsfelder (Studiengänge, Lehrveranstaltungen) Wege aufzuzeigen, wie diese in eigener Anstrengung ihr Handeln verbessern können. Sie versteht sich in diesem Sinne als Instrument für die interne Entwicklung und Optimierung, bspw. zur Untersuchung von Stärken und Schwächen eines Studienprogrammes. Gleichzeitig werden ihre Ergebnisse aber zunehmend auch dafür benutzt, Steuerungsentscheidungen, die in zuweilen äußerst wirksamer Form (bspw. wenn Allokationsentscheidungen betroffen sind) in den Alltag der beteiligten Akteure eingreifen, zu rechtfertigen. Evaluation wird dann durchaus als Kontrollinstrument und nicht als Instrument der Selbstanalyse verstanden. In diesem Spannungsfeld müssen Wege gefunden werden, die es ermöglichen, dass Evaluationsinstrumente auf der Basis geteilter Vorstellungen und Erwartungen der beteiligten Akteure eingesetzt werden. Verschiedene Konzepte und Ansätze der Qualitätssicherung werden vor diesem Hintergrund diskutiert.

Im fünften Kapitel werden die landläufig in der Qualitätssicherung und -entwicklung von Hochschulen eingesetzten Evaluationsverfahren diskutiert. Gegenstand sind zudem die zur leistungsbezogenen Mittelvergabe zunehmend eingesetzten, oben genannten Indikatoren. Anders als Evaluationsverfahren, die vielfach die Erhebung (subjektiver) studentischer Qualitätsurteile betonen, werden in der indikatorgestützten Leistungsbeurteilung hauptsächlich quantitative Daten zur Beschreibung der Leistungserbringung (Zahl der Studierenden in der Regelstudienzeit, Absolventen- und Schwundquoten, etc.) herangezogen. Ob diese mit potenziellen Validitätsbedrohungen belastet sind, ist Gegenstand der Diskussion

---

sowie die Frage, inwieweit sie tatsächlich zur Qualitätsentwicklung von Hochschulen beitragen können. Dazu werden die Ergebnisse von Sekundäranalysen hochschulstatistischen Materials zur Frage der Belastbarkeit bzw. der begrenzten Aussagefähigkeit verschiedener Leistungskennzahlen, vorgestellt.

Kapitel 6 greift die eingangs gestellte Frage nach der Qualität erhobener Daten und ihrer Übereinstimmung mit verschiedenen Aspekten zentraler Testgütekriterien wieder auf. Bevor entsprechende Untersuchungen, die anhand von Daten des Lehrevaluationsmodells einer Hochschule durchgeführt wurden vorgestellt werden, sollen die testtheoretischen Grundlagen dieser Gütekriterien kurz erläutert werden. Darüber hinaus sind Teilaspekte dieser Gütekriterien, die aus vorliegender Sicht als besonders relevant für den Kontext der Lehrevaluation erscheinen, im entsprechenden Zusammenhang Gegenstand der Diskussion. Im Anschluss wird in Kapitel 7 eine kurze Beschreibung der für die genannten Auswertungen herangezogenen Datenbasis geliefert.

In den Kapiteln 8 bis 10 werden die Auswertungsergebnisse berichtet. Dabei wird auf die Frage nach der Dimensionierung von studentischen Einschätzungen zur Qualität der Lehre (Kap. 8), die zeitliche Variabilität dieser Urteile (Kap. 9) und auf die Sachbezogenheit studentischer Qualitätsurteile (Kap. 10) eingegangen. Abschließend werden die Analyseergebnisse in Kapitel 11 zusammenfassend und vor dem Hintergrund der Frage nach einer künftigen Entwicklung von Verfahren der Lehrevaluation diskutiert.

# Kapitel 2

## Hochschulpolitische Rahmenbedingungen

Das deutsche Hochschulsystem befindet sich in einem Prozess grundlegenden Wandels. Eine steigende Bildungsnachfrage und Bildungsbeteiligung seit den 1960er und 1970er Jahren (vgl. z.B. Müller 2001; Kreckel 1992; Hradil 2001; Reissert 1991; Jaspers und Rossmann 1961) trug zu einer Veränderung des Aufgabenspektrums der Hochschulen sowie der Erwartungen, die den Hochschulen seitens der Öffentlichkeit entgegengebracht werden, bei. Diese Erwartungen bestehen beispielsweise in einer Orientierung von Studium und Lehre auf eine klar definierte Berufsqualifizierung (Schröder 2003) sowie einer Verkürzung der Studiendauer (Reissert 1991). Erreicht werden sollen diese Ziele durch die Entwicklung zunehmend verregelter Studiengänge.<sup>1</sup> Universitäten und Fachhochschulen sind

---

<sup>1</sup>Der derzeitige Höhepunkt dieser Entwicklung ist in der verbindlichen Einführung von Bachelor- und Masterstudiengängen als Regelstudienabschluss zu sehen. Diese Studienstrukturreform, die durch die Bologna-Deklaration der europäischen Bildungsminister 1999 eingeleitet wurde, soll eine europaweite Harmonisierung und Vergleichbarkeit der Bildungsabschlüsse erreichen (vgl. z.B. Kultusministerkonferenz 2000; Hochschulrektorenkonferenz 2004). Vordringlichstes Ziel ist es daneben, die Teilnehmer an höherer Bildung in zeitlich möglichst kurzen und dementsprechend stark strukturierten Studienprogrammen mit berufsqualifizierenden Kompetenzen auszustatten.

---

mittlerweile berufsqualifizierende Ausbildungsinstitution für mehr als ein Drittel der Geburtsjahrgänge (Schreier 2002)<sup>2</sup>, im Studienjahr 2003/04 betrug die Zahl der insgesamt eingeschriebenen Studierenden mehr als 2 Millionen (BMBF).<sup>3</sup> In einem Statusbericht zur Untersuchung der Entwicklung des europäischen Hochschulsystems beschreibt bspw. das European Network for Quality Assurance (ENQA) den Wandel der europäischen Hochschulen von einer Bildungsinstitution, die weitgehend den Eliten vorbehalten war, hin zu einer (Aus-)Bildungsstätte für breite Gesellschaftsschichten: „First and foremost Western European higher education has seen a change from elite systems to mass systems of higher education with a dramatic increase in student enrolment“ (ENQA 2000). Eine weitere gesellschaftliche Öffnung der Hochschule und ein weiterer Anstieg der Beteiligung an höherer Bildung ist eines der vordringlichsten Ziele auch der deutschen Hochschulentwicklungspolitik (BMBF).

Gemäß ihrem gesamtgesellschaftlichen Bildungsauftrag müssen die Hochschulen verschiedenen Qualitätserwartungen gerecht werden. So wird bspw. die Wirtschaft, als Abnehmerin der (beruflich) qualifizierten Nachwuchskräfte ein Hochschulstudium dann für qualitativ hochwertig halten, „wenn es für den jeweiligen Wirtschaftszweig besonders gut verwendbare Absolventen gibt, die ein möglichst kurzes training on the job benötigen, da die einzelne Firma in dieser Zeit noch in die junge Frau bzw. den jungen Mann investieren muss“ (Seidel 2001, S. 23). Von seiten der Wissenschaft und den Hochschulen selber wird an das Studium die Erwartung gerichtet, dass ein hervorragender wissenschaftlicher Nachwuchs das Ausbildungsprofil der jeweiligen Universität bestätigt, Eltern wünschen sich

---

<sup>2</sup>Gleichwohl liegt Deutschland mit diesem Anteil auf einem der hinteren Ränge im OECD-Vergleich. Die durchschnittliche Beteiligung an höherer Bildung lag in den OECD-Staaten im Studienjahr 2003/04 bei 51 Prozent (BMBF). Allerdings sind in diesem Vergleich auch die unterschiedlichen nationalen Bildungstraditionen und entsprechend unterschiedlichen Bildungswege (bspw. das für Deutschland typische duale System von schulischer und beruflicher Bildung) zu berücksichtigen.

<sup>3</sup>Vgl. auch die diesbezüglichen Angaben des Statistischen Bundesamtes unter <http://www.destatis.de/basis/d/biwiku/hochtab2.php> (zuletzt 14.02.2007).



---

für ihre Kinder eine Ausbildung, die ihnen gute Lebenschancen eröffnet u.s.w. Zusammengefasst stehen die Universitäten vor einem „öffentlichen Legitimationsproblem, (...) bezüglich des Umgangs der Hochschule mit Ressourcen (Personal, Raum, Sachmittel), mit der Lebenszeit der Studierenden, mit der Erzeugung gesellschaftlich notwendiger Qualifikationen und mit den (Aus-)Bildungschancen von Individuen, die von den Auswirkungen Studienzeit verlängernder Lehrorganisation auf die Ausbildungskapazitäten betroffen sind“ (Webler 1991, S. 77).

Hochschulen und insbesondere die Universitäten vollziehen dabei einen Wandel zum Dienstleister, der die steigende Nachfrage nach Bildungsangeboten möglichst effizient zu befriedigen sucht. Ein solches, vom New Public Management inspiriertes Leitbild von der Universität (s.o.; Schimank 2006) sieht in diesen kompetitive Marktakteure, die sich in der Konkurrenz um Kunden seines (Bildungs-)Angebots gegen die Wettbewerber durchsetzt, indem er Exzellenz und beste Qualität nachweist.

Dieses Leitbild kollidiert mit dem traditionellen (Selbst-)Verständnis von Hochschulen und den aus diesem Selbstverständnis erwachsenen (Selbst-)Steuerungsprinzipien der Hochschulen. In diesem sind Universitäten weniger als Organisationen, denn als Institutionen zu verstehen. Dieser Unterschied ist insofern relevant, als eine Institution als „stabiles Muster strukturierter menschlicher Beziehungsgeflechte“ (Nickel 2003, S. 224) zu sehen ist, oder -wie Schelsky es mit Blick auf die Idee von den Universitäten bzw. der Wissenschaft ausdrückt- als soziale Lebensform (Schelsky 1963, S. 65). Darin drückt sich der Daseinszweck der Hochschulen aus: Die Universität als Institution „stellt den Rahmen zur Verfügung und der Staat beaufsichtigt dezent die eigenwilligen Forscher/innen, die in Einsamkeit und Freiheit Wissen schaffen“ (a.a.O.).<sup>4</sup>

---

<sup>4</sup>Schelsky zitiert seinerseits mit dem Titel seines Buches „Einsamkeit und Freiheit“ (Schelsky 1963) Wilhelm von Humboldt zu dessen idealistischer Gründungsidee der Universität: „Der Universität ist vorbehalten, was nur der Mensch durch und in sich selbst finden kann, die Einsicht in die reine Wissenschaft. Zu diesem Selbstaktus ist nötig *Freiheit und hilfreich Einsamkeit*, und aus diesen beiden Punkten fließt zugleich die ganze äußere Organisation der Universitäten“

---

Universitäten sind dementsprechend „Korporationen mit Selbstverwaltung“ (Jaspers und Rossmann 1961, S. 2), „Sie haben ihr vom Staat freigelassenes Eigenleben aus der unvergänglichen Idee, einer Idee überstaatlichen, übernationalen, weltweiten Charakters. Daher beanspruchen sie und wird ihnen gewährt die Freiheit der Forschung und der Lehre. Das heißt, sie sollen die Wahrheit suchen und lehren unabhängig von Wünschen und Weisungen, die sie von außen oder von innen her beschränken möchten“ (a.a.O.). Diese universitäre Freiheit und Selbstverantwortlichkeit wird insbesondere in den Personen der Wissenschaftler angesiedelt, der Professor hat „die Freiheit des Lehrers“ (Jaspers und Rossmann 1961, S. 2).<sup>5</sup>

Im Gegensatz zu diesem Verständnis von Universitäten als Institutionen, in dem diese eine Entlastungsfunktion zur Absicherung der wissenschaftlichen Freiheit der einzelnen Wissenschaftler haben, beschreiben Organisationen eher die „zielgerichtete Ordnung bzw. Regelung von Aufgaben (Funktionen) und Tätigkeiten (Arbeitsvorgängen) in sozialen Gebilden in der Weise, dass alle Elemente der Organisation (Aufgaben, Tätigkeiten) und alle daraus gebildeten Organisationseinheiten (Stellen, Abteilungen, Arbeitsprozesse) in das Gefüge des Sozialgebildes eingegliedert sind“ (Nickel 2003, S. 225).

---

(Humboldt 1809, zit. nach Schelsky 1963, S. 68, Kursivsetzung im Original; vgl. auch Müller 1990, S. 274). Auch andere Autoren beschreiben die zentrale Bedeutung des genannten Zitats für die humboldt'sche Universitätskonzeption, die ihrerseits das beschriebene Selbstverständnis der (deutschen) Universität begründet: „His (W. v. Humboldts; P.P.) much-quoted formula of ‚Einsamkeit und Freiheit‘ (solitude and freedom) must be understood in this sense: the pursuit of knowledge, independent of any social or vocational consideration, bestows the confidence to render ourselves open to the secrets of the world“ (Hahn 1998, S. 5).

<sup>5</sup>Dieses Leitbild (hochschulischer) Bildung ist auf die oben bereits angesprochenen Grundsätze der Humboldt'schen Bildungslehre zurückzuführen (vgl. Fußnote 4), deren drei übergeordnete Ideale in der Individualität (im Sinne von Selbsterkenntnis und dem genannten Streben nach Wissen als Weg zur Selbstwerdung des Menschen), der Totalität (als Entfaltung aller intellektuellen, emotionalen sowie praktischen Fähigkeiten) und Universalität (als Aneignen der Welt, mittels der kontinuierlichen Vermehrung des Wissens) bestehen (Hahn 1998, S. 5).

---

Die Einbindung (ehemals) einsam und frei forschender Wissenschaftler in eine Organisation, in der sie auf gemeinsame Organisationsziele verpflichtet werden können, stärkt die Position der Leitungsebene eben dieser Organisation und schwächt die freiheitliche Position der einzelnen Wissenschaftler. Vorgegebene Organisationsziele können dabei die staatlichen bzw. ministerialbürokratischen Interessen (Effizienz, Steigerung der Absolventenquote, Senkung der Studiendauer, etc.) zu bedienen versuchen oder aber sich auf akademische Ziele (Profilierung, Herausbildung von Exzellenz in a priori festgelegten Wissenschaftsfeldern) beziehen.

Der Wandel des Verständnisses von der Hochschule als Institution hin zu einem, in dem Hochschulen eher als Organisation aufgefasst werden, schafft den Rahmen, in dem Marktvergesellschaftung, Wettbewerb, Autonomie -kurz: die Kennzeichen der derzeitigen hochschulpolitischen Entwicklungstrends überhaupt möglich sind. Ein wichtiges Element in diesem Zusammenhang besteht in dem Verhältnis zwischen Hochschulen und Staat. Die traditionell sehr enge Beziehung zwischen Staat und Universität wurde unter dem Eindruck der Expansion des Hochschulwesens und der Krise staatlicher Haushalte seit den 1990er Jahren schrittweise gelockert (z.B. Stifter 2002). Knappe öffentliche Mittel werden in die Verantwortung der einzelnen Hochschulen übertragen. Der Staat zieht sich aus der Detailsteuerung der hochschulinternen Abläufe zurück und beschränkt sich auf die Funktion der Rechtsaufsicht (z.B. Fischer-Bluhm 2000; Voßkuhle 2004). Mit dieser Übertragung und Gewährung von lokaler Autonomie wird die Vermutung und auch die Hoffnung verbunden, dass die den Hochschulen zur Verfügung gestellten Mittel vor Ort am effektivsten, weil problemorientiert eingesetzt werden können. In der diesbezüglichen Debatte setzt sich zunehmend die Auffassung durch, „dass deregulierte, differenzierte Hochschulsysteme mit Institutionen, die über hohe Autonomie und entsprechende institutionelle Profile verfügen, eher in der Lage sind, mit Wandel umzugehen und vielfältige Bedürfnisse, wie sie an ein Massenhochschulsystem herangetragen werden, zu befriedigen“ (Pellert 2002). Gleichzeitig soll die Gewährung von Autonomie einen Wettbewerb der Hochschulen untereinander auf dem zunehmend heiß umkämpften Bildungsmarkt indu-

---

zieren (z.B. Lange 2002; Kreckel 2002). Angesichts knapper Kassen<sup>6</sup> wird dabei erwartet, dass die Hochschulen selbstständig Profile bilden und exzellente Bereiche fördern, während sie weniger gute Bereiche „absterben“ lassen. Nicht überall könne alles gleich gut angeboten werden. Hinlänglich bekannte Haushaltszwänge werden dabei umgedeutet in die Sorge um die Qualität der Hochschulleistungen. Die in einem Wettbewerb stehenden und autonom, als Wettbewerber agierenden Hochschulen sind in diesem Sinne aufgefordert, die Qualität ihrer Leistungen in Lehre und Forschung selbst zu sichern und zu entwickeln, um auf den als Profilschwerpunkten definierten Gebieten Exzellenz zu erreichen. Im Gegenzug für die gewährte Entscheidungs- und Planungsautonomie wird von den Hochschulen Rechenschaftslegung bezüglich der erbrachten Leistungen gegenüber dem Staat, der unverändert der primäre Mittelgeber der Hochschulen ist, erwartet.

Neben diesem Legitimationsdruck gegenüber dem Staat, sehen sich die Hochschulen mit einer zunehmend kritischen Öffentlichkeit konfrontiert. Diese beurteilt die Hochschulen als ineffizient, träge, zu wenig leistungsorientiert und zu wenig leistungsbereit (Schröder 2003, S. 19). Im Sinne der genannten veränderten Aufgaben von Hochschulen (hinsichtlich ihrer Ausbildungsfunktion) wird in der öffentlichen Diskussion insbesondere die *Qualität der Lehre* in den Mittelpunkt gestellt und den Hochschulen vielfach die Eignung abgesprochen, dass vielfältige gesellschaftliche Bedürfnis nach Bildungsangeboten zu befriedigen. Dabei werden Schlagworte wie das von den „faulen Professoren“ (Enders und Schimank 2001) in einem Atemzug mit den Zuständen der „Massenuniversität“ genannt und die Hochschulen für die entsprechende Situation haftbar gemacht, obwohl zumindest der letztgenannte Problemkreis weitgehend außerhalb des unmittelbaren Einflusspielraumes der Hochschulen liegt (Hocke 2003).

Durch Hochschulrankings soll die interessierte Öffentlichkeit über die Leistungs(un)fähigkeit der einzelnen Hochschulstandorte informiert werden (vgl. z.B.

---

<sup>6</sup>Bereits Mitte der 1990er Jahre warnte die Hochschulrektorenkonferenz vor den Folgen der Unterausstattung der Hochschulen für die Qualität von Lehre und Forschung und mahnte eine entsprechend veränderte Prioritätensetzung der verantwortlichen Politik an (HRK 1996).

---

Hornbostel 2001; Hornbostel 2001a; Müller-Böling u. a. 2001). In den entsprechenden Verfahren ermittelte und in Kennzahlen ausgedrückte Qualitätsunterschiede werden in Rangplätze überführt, die darüber Auskunft geben sollen, welcher Universitätsstandort hinsichtlich der „gerankten“ Aspekte der Beste sei. Losgelöst von der methodologischen Diskussion über die (methodische) Qualität solcher Rankings, die sich bspw. auf die Validität der Informationen, die ihrerseits die *Qualitätsunterschiede der Hochschulen quantifizieren* sollen bezieht (Engel und Krekeler 2001), haben diese in der öffentlichen Wahrnehmung und Beurteilung der Hochschulen inzwischen einen festen Platz gefunden. Studieninteressierte greifen auf die präsentierten Informationen zurück, um zu einer Entscheidung über den zu wählenden Studienstandort zu kommen, Wissenschaftler orientieren ihre Bewerbungsstrategie an den Aussagen der Rankings, etc. Eine gute Position in einem einschlägigen Ranking wird für die Hochschulen demgemäß immer mehr zu einem relevanten Wettbewerbsfaktor.

Aus den geschilderten Entwicklungen ergibt sich die Bedeutung von Qualitätssicherung und -entwicklung für die Hochschulen. Diese wird zunehmend zu einer strategischen und nahezu vitalen Aufgabe ersten Ranges, da die Fortsetzung der Mittelzuweisung in immer größerem Maße an Erfolge, bzw. berichtete Erfolge geknüpft wird. „Die systematische Sicherung und Verbesserung von Qualität sind unausweichliche Konsequenzen von mehr Eigenverantwortung der Hochschulen und weniger staatlicher Vorschriften im Rahmen eines europäischen und zunehmend globalen Wettbewerbs um die besten Köpfe“ (Hochschulrektorenkonferenz 2002). Der Wettbewerb, in dem sich Hochschulen untereinander befinden, bezieht sich auf sämtliche Leistungsbereiche. Reputable Leistungen in Forschung und Lehre stellen das Kapital dar, mit dem Hochschulen Studieninteressierte für ihre Studienprogramme begeistern, exzellente Forscher an sich binden, die ihrerseits durch die erfolgreiche Einwerbung von Drittmitteln die Reputation der Hochschule mehren und so fort. Überspitzt formuliert, wird die Abhängigkeit der Hochschulen von der staatlichen Steuerung ersetzt durch die Konjunkturverläufe des Bildungsmarktes. Für die entsprechenden Bemühungen zur Qualitätssicherung und -entwicklung werden die Hochschulen selber verantwortlich gemacht,

---

wobei in die Wahl der einschlägigen Indikatoren zur Bestimmung dieser Qualität durchaus von außen eingegriffen wird. Dieser Eingriff besteht insbesondere in der Festlegung von Qualitätskriterien, die zu einem beträchtlichen Teil von bildungspolitischer bzw. ministerialer Seite vorgegeben werden. Dementsprechend stehen sie in der Gefahr, nicht ausschließlich wissenschaftsimmanente Maßstäbe für die Qualitätsbeurteilung heranzuziehen und so zu - bezogen auf die Beurteilung des Wissenschaftssystems - potenziell sachfremden Beurteilungen zu kommen. Bestes Beispiel hierfür sind die rein quantitativ orientierten Indikatoren, die sich an den auf den zahlenmäßigen Output der Hochschulen bezogenen Interessen der Bildungspolitik und Ministerialbürokratie beziehen und in die leistungsorientierte Mittelvergabe Eingang gefunden haben.

Neben der Überprüfung von Leistungen, die sich im zahlenmäßigen Output der Hochschulen (Absolventenzahlen, Studierende in der Regelstudienzeit, etc.) ausdrücken, hat sich als Beurteilungsinstrument die Evaluation von Lehre und Studium etabliert (vgl. z.B. Spiel 2001b). Deren Hauptanliegen besteht darin, Informationen für qualitätsbezogene Aussagen zur Studiensituation auf verschiedenen Ebenen (Lehrveranstaltungen, Fächer/Fachbereiche, Hochschulen) zu sammeln und den beteiligten Akteuren zur Verfügung zu stellen. Diese Informationen sollen die Selbstreflexion über etwaige Verbesserungsmöglichkeiten anschieben. Nichtsdestotrotz werden auch Evaluationsergebnisse in zunehmendem Maße hinsichtlich ihrer Verwendbarkeit für Leistungsbeurteilungen und -vergleiche, auf die sich Steuerungsentscheidungen stützen, diskutiert. So empfiehlt bspw. die Kultusministerkonferenz in einem grundlegenden Konzept zur Qualitätssicherung in der Lehre Kernelemente, unverzichtbare Bestandteile und Standards von Qualitätssicherung und Evaluation von Lehre und Studium.<sup>7</sup> Unter diesen Kernelementen wird bspw. eine als Anreizsystem bezeichnete Verschränkung von Evaluationsergebnissen mit der leistungsorientierten Mittelvergabe sowie die Einführung von

---

<sup>7</sup>Das Konzept zur Ausgestaltung verbindlicher Standards der Qualitätssicherung von Lehre und Studium an allen Hochschulen wurde auf der 311. Plenarsitzung der KMK am 13./14.10.2005 in Potsdam verabschiedet (Kultusministerkonferenz).

---

Sanktionen als Konsequenz der Evaluation (Kultusministerkonferenz, S. 7) subsumiert. Evaluation entwickelt sich so von einem internen Instrument der Diskussion und Selbstreflexion hin zu einem Kontrollinstrument der Hochschulsteuerung. Für die Hochschulen und hier insbesondere die dem „Prozess des Absterbens“ (s.o.) preisgegebenen Bereiche stellen sich in diesem Zusammenhang verschiedene relevante, teils geradezu vitale hochschulpolitische Fragen: Welche Bereiche werden im Prozess der Formulierung von Hochschulprofilen verzichtbar? Wie und durch wen wird darüber entschieden? Welche Maßstäbe sind für die Entscheidung anzulegen? Wie lassen sich sachbezogene Kriterien zur Beurteilung guter Qualität an der Hochschule finden und sich gleichzeitig das Anlegen von sachfremden, bspw. rein fiskalischen Kriterien vermeiden? Welche Daten zur Beurteilung der Leistungsfähigkeit von Hochschulfächern sind heranzuziehen und welche werden tatsächlich herangezogen? Wie steht es also um das Verhältnis von Anspruch und Wirklichkeit in der aktuellen Diskussion und Praxis der Leistungsbeurteilung von Hochschulen?

Über diese unmittelbar hochschulpolitischen Fragen hinaus, ist durchaus auch nach dem Selbstverständnis bzw. dem Anspruch der Evaluationsforschung - und hier der Evaluation von Lehre und Studium - als vergleichsweise junger wissenschaftlicher Teildisziplin zu fragen. Insbesondere ist zu diskutieren, wie die Reichweite und die Aussagefähigkeit der Evaluationsergebnisse zur Beurteilung von Lehre und Studium einzuschätzen ist. Es steht zur Debatte, inwieweit die Praxis der Evaluation von Lehre und Studium und deren Ergebnisse geeignet sind, Steuerungsentscheidungen in den Hochschulen zu rechtfertigen. Ist die vorhandene Datenbasis hinreichend belastbar, sind bspw. studentische Befragungsdaten eine aussagefähige Informationsbasis für die angestrebten und nachgefragten Leistungsbeurteilungen von Lehre und Studium auf den verschiedenen Ebenen, von der konkreten Lehrveranstaltung bis zur Organisation ganzer Studiengänge? Oder gibt es vielmehr Limitationen und Bedrohungen der Validität der Evaluationsergebnisse, die zu sachfremden und ungerechten Entscheidungen führen? Nachfolgend sollen zunächst das Forschungsprogramm der Evaluationsforschung im Allgemeinen und die Praxis der Qualitätssicherung und Evaluation von Lehre

---

und Studium im Besonderen vorgestellt werden. Dies dient dazu, die methodologischen Besonderheiten kenntlich zu machen, die auch in dem Evaluationsmodell berücksichtigt wurden, welches die Grundlage für die unten beschriebenen Analysen darstellt.



# Kapitel 3

## Evaluation -

## Forschungsprogramm und

## Instrument der politischen Praxis

Evaluation hat in den letzten Jahren verstärkt an Aufmerksamkeit gewonnen und wird in vielen Politikfeldern für die Überprüfung der Qualität von Leistungen und der Effizienz der Verwendung von (zumeist öffentlichen) Ressourcen gefordert (z.B. Stockmann 2000). Mittels Evaluation sei in erster Linie zu überprüfen, ob und inwieweit sich die gewünschten Effekte eines Programms oder einer Aktivität eingestellt haben (z.B. Ganz 1984, S. 630; Wottawa und Thierau 1998). Darüber hinaus dient sie den an diesen Programmen oder Aktivitäten beteiligten Akteuren als Instrument zur Selbstreflexion, zur Analyse von Stärken und Schwächen des Programms und seiner Durchführung sowie entsprechender Entwicklungspotenziale und -bedarfe. Insofern jedoch der Term Evaluation im Sinne einer Überprüfung von Leistungen oder deren Effizienz genutzt wird, tritt dieser Zweck notwendiger Weise in den Hintergrund.

Der Begriff der Evaluation hat sich zum vielfach eingesetzten Modewort entwickelt (Stockmann 2002, S. 219; Kromrey 2000, S. 233) und Eingang in die Alltagssprache gefunden. Damit geht jedoch auch ein teils unspezifischer Gebrauch des Begriffs einher. Vielfach wird jedwede Form der Bewertung oder der „Festsetzung des Wertes einer Sache“ (Scriven 1980) mit dem Begriff Evaluation

gleichgesetzt (Kromrey 2000; Wottawa und Thierau 1998). Darüber hinaus werden vielfach die Begriffe Evaluation und Umfrageforschung synonym gebraucht, mit der Begründung, dass auch durch letztere „evaluative“ Aussagen von den jeweils befragten Personen erhoben würden, so dass die derart betriebene Forschung als Evaluation zu bezeichnen sei (Kromrey 2000, S. 233, Kromrey 2001b, S. 105). Daraus ergibt sich das Problem der inflationären Nutzung des Begriffs für Untersuchungen, die ohne grundlegende theoretische Anforderungen, die an Evaluationsforschung<sup>1</sup> im engeren Sinne zu stellen sind, durchgeführt werden (Kromrey 1995). Für diese hat sich in den letzten Jahrzehnten ein eingrenzendes Begriffsverständnis etabliert. Aus diesem ergeben sich Anforderungen, die aus wissenschaftstheoretischer Sicht an die Konzeption und Durchführung von Evaluationen zu stellen sind. Insbesondere vor dem Hintergrund der Nutzung von Evaluationsdaten für - wie im Fall der Lehrevaluation angestrebt oder praktiziert - (hochschul-)politische Steuerungsentscheidungen sind die entsprechenden Gütekriterien für den Forschungsprozess von großer Bedeutung. Nachfolgend werden deshalb die Implikationen dieses Begriffsverständnisses zunächst erläutert.

## 3.1 Das Forschungsprogramm der Evaluation

In einem strengen Verständnis bezeichnet „Evaluation in der Methodologie der empirischen Sozialforschung (...) das Design für einen spezifischen Forschungstyp, durchgeführt im idealtypischen Fall als Feldexperiment mit Kontrollgruppen“ (Kromrey 2000). Feldexperimente werden dabei als eine Kombination verschiedener Untersuchungsvarianten (experimentell vs. quasi-experimentell und Feld vs. Labor) zu einem spezifischen Forschungsdesign verstanden: beispielsweise werden randomisierte Gruppen (experimentelle Untersuchungskomponente) in ihrer

---

<sup>1</sup>Im Gegensatz zu Evaluation bezeichnet in Anlehnung an Mittag und Hager (2000) Evaluationsforschung eine systematische Anwendung wissenschaftlicher Methoden zur Informationsgewinnung hinsichtlich der Bewertung verschiedener Aspekte sozialer Programme (Mittag und Hager 2000; vgl. auch Rossi und Freeman 1993).

natürlichen Umgebung, außerhalb einer Laborsituation (Felduntersuchung) untersucht (Bortz und Döring 2002). Grundlegendes Ziel von Evaluationen ist es, Informationen über die Wirkung eines Programms oder einer Interventionsmaßnahme auf die dieser Maßnahme ausgesetzten Personen zu sammeln und mit diesen Informationen entsprechende Aussagen und Bewertungen begründen zu können. Unter Programmen wird dabei im allgemeinen ein komplexes Handlungsmodell verstanden, dass unter Einsatz angemessener Handlungsstrategien und Ressourcen die Erreichung bestimmter Ziele anstrebt, die - wie bspw. im Fall von Kampagnen (z.B. Joußen 1995; Kunczik 1994) in der Verhaltensänderung von Personen bestehen können (z.B. Lange 1983; Mittag und Hager 2000; Hellstern und Wollmann 1984). Bezogen auf den Anwendungsfall der Evaluation von Lehre und Studium ließe sich diese Definition insofern anwenden, als Studiengänge die akademische und berufliche Qualifizierung der Studierenden zwischen der Aufnahme und dem Abschluss ihres Studiums zur Aufgabe haben und dieses Ziel unter Einsatz verschiedener personeller und sächlicher Ressourcen erreichen sollen.

#### 3.1.1 Evaluationsforschung und Grundlagenforschung

Der empirische Nachweis von Wirkungen eines Programms ist ein wichtiger Anspruch von Evaluationen. Aus diesem Anspruch ergibt sich ein entscheidender Unterschied der Evaluationsforschung zur grundlagenorientierten sozialwissenschaftlichen Forschung (Vedung 2000). Diese wird dem begrifflichen Verständnis nach zweckfrei und um ihrer selbst Willen betrieben, um nach allgemein gültiger Erkenntnis zu streben (Stockmann 2000). Evaluationsforschung ist dagegen stets an einem unmittelbaren praktischen Nutzen orientiert, meist verfolgt sie dabei spezifische Absichten und Informationsbedürfnisse eines Auftraggebers (Chelimsky 1985). Evaluation soll „dazu beitragen, Prozesse transparent zu machen, Wirkungen zu dokumentieren und Zusammenhänge aufzuzeigen, letztlich um Entscheidungen treffen zu können. Z.B. mit dem Ziel, Ablaufprozesse effektiver zu gestalten, den Input effektiver einzusetzen, den Output zu erhöhen, den Wirkungsgrad zu verbessern, die Nachhaltigkeit zu sichern, etc. D.h., Evaluationen können - wie Qualitätsmanagementsysteme - dazu beitragen, die Qualität eines

Programms, einer Maßnahme oder einer Dienstleistung zu verbessern“ (Stockmann 2000, S. 219; vgl. auch Stockmann 2006).

Anders als Grundlagenforschung kommt Evaluationsforschung nicht zu wertneutralen Aussagen. Vielmehr ist die Bewertung des Evaluationsgegenstandes gerade ihr Zweck und wird entsprechend als Ergebnis von ihr verlangt. Evaluationen sollen im Sinne der angestrebten Erkenntnisziele handlungsleitende und handlungsrelevante Bewertungen bereitstellen. „Nicht Erkenntnis im Sinne einer Falsifikation von Theorien und Hypothesen ist primäres Ziel der Evaluation, sondern die Angabe von Handlungsalternativen zur Lösung auftretender Probleme (...)“ (Lange 1983).

Über die Abschätzung und Untersuchung von Wirkungen hinaus umfasst das Aufgabenspektrum der Evaluation auch die Beratung der evaluierten Programme bzw. der für ihre Durchführung Verantwortlichen und die Moderation von Reorganisationsanstrengungen, die durch die Evaluation angeregt werden. „Evaluation besteht in der Analyse und Bewertung von Einrichtungen, Programmen, Innovationen unter dem Aspekt ihrer Wirksamkeit, Akzeptanz und Qualität mit dem Ziel der Situations- und Problemdiagnose sowie der Mobilisierung von Aktivität“ (Hennen 2002, S. 100).

#### **3.1.2 Wirkungsmessungen und Prozessanalysen**

Wirkungen eines Programms sind aufzufassen als kausale Effekte dieses Programms bspw. auf Personen, die den Programminhalten ausgesetzt werden und im Anschluss an die Programmteilnahme ihr Verhalten ändern, neues Wissen erworben haben, etwas gelernt haben, etc. Um derartige kausale Wirkzusammenhänge zu beschreiben oder anders ausgedrückt, um nachzuweisen, dass sich Verhaltens-, Einstellungs- oder Zustandsänderungen tatsächlich auf das Wirken des entsprechenden Programms zurückführen lassen, bedarf es der (statistischen) Kontrolle von etwaigen Störfaktoren. Diese müssen quantifiziert und hinsichtlich des jeweils spezifischen Einflusses beschrieben werden, den sie an Stelle des gewünschten Programmeffekts hatten (Hager 2000).

Bezogen auf das Beispiel der Evaluation von Lehre und Studium könnte ein

Störfaktor in einer unterschiedlichen Wochenarbeitszeit liegen, die Studierende neben dem Studium aufbringen müssen, um ihren Lebensunterhalt zu verdienen. Sofern sich ein unterschiedlicher Studienerfolg auch auf die in verschiedenem Maße zur Verfügung stehende (Studien-)Zeit zurückführen lässt, muss die Annahme der kausalen Verknüpfung des Studienerfolgs mit der Wirkung des Studienprogramms als durch den Einfluss des unterschiedlichen Zeitbudgets verzerrt gelten.

Notwendig wäre entsprechend ein Forschungsdesign, in dem derartige Störvariablen kontrolliert werden können. Im oben beschriebenen Sinne einer Forschung in Experimentallogik wäre als Voraussetzung für die Durchführung von Wirkungsanalysen das Vorhandensein von Experimental- und Kontrollgruppen zu fordern. Diese ermöglichen durch Randomisierung, also eine zufällige Zuteilung der Untersuchungsteilnehmer auf die verschiedenen Gruppen eine Kontrolle von Störvariablen (Bortz und Döring 2002).

Weiterhin wäre es notwendig, eindeutig zu definieren, welche Wirkung durch welche Programmaktivität erzielt werden soll. Es müsste also a priori formulierte Theorien und Hypothesen zu den Wirkungsweisen und -zusammenhängen der verschiedenen Programmbestandteile und -inhalte geben. Kromrey (1995) hebt verschiedene für die Anwendung des methodologischen Forschungsprogramms empirischer Kausalanalysen<sup>2</sup> unerlässliche Bedingungen hervor, die jedoch in der Realität der Programmforschung im Bereich der Evaluation von Lehre und Studium zumeist nicht gegeben sind, so dass durch sie eher der Ideal- als der Realtyp eines Forschungsdesigns formuliert wird:

- „Vor der Entwicklung des Forschungsdesigns muss Klarheit über die Untersuchungsziele -bezogen auf einen definierbaren und empirisch abgrenzbaren Untersuchungsgegenstand- bestehen. Für die Dauer der Datenerhebung dürfen sich weder die Untersuchungsziele noch die wesentlichen Randbedin-

---

<sup>2</sup>Die analytischen Aufgabenstellungen bzgl. der zu evaluierenden Programmwirkungen sind nach Kromrey (2001b) „aufzubrechen in jene Teile, die den jeweiligen Maßnahmen als deren Wirkung zugerechnet sind und in die verbleibenden Teile, die als Effekte exogener Einflüsse (Programmumwelt) zu gelten haben“ (Kromrey 2001b, S. 120)

gungen des Untersuchungsgegenstandes in unvorhersehbarer Weise ändern.

- Vor der Entwicklung des Forschungsdesigns müssen (...) begründete Vermutungen (Hypothesen) über die Struktur des Gegenstandes wie auch über die Zusammenhänge und Beziehungen zwischen dessen wesentlichen Elementen existieren, nach Möglichkeit in Form empirisch bewährter Theorien. Erst auf ihrer Basis kann ein Gültigkeit beanspruchendes Indikatorenmodell konstruiert, können geeignete Messinstrumente entwickelt, kann über problemangemessene Auswertungsverfahren entschieden werden.
- Der Forscher muss die Kontrolle über den Forschungsablauf haben, um die (interne und externe) Gültigkeit der Resultate weitgehend sicher zu stellen“ (Kromrey 1995, S. 323).

In der Praxis der Bewertung von Lehre und Studium sind die mit den genannten Ansprüchen wissenschaftlicher Evaluation verbundenen Merkmale einer Wirkungsanalyse nicht oder nicht in hinreichendem Maße gegeben (Kromrey 1995, S. 323). So sind die Ziele der zu evaluierenden Studiengänge (die in diesem Sinne als Programme aufgefasst werden; s.o.) nicht eindeutig definiert bzw. präzise beschreibbar. Einzelnen Programminhalten können keine eindeutigen Indikatoren zur Qualitätsmessung zugeschrieben werden, die Wirkung oder deren Ausbleiben lässt sich nicht präzise auf die Programmgestaltung und -durchführung zurückführen. Aufgrund der verschiedenen denkbaren und miteinander konkurrierenden Qualitätsansprüche und multidimensionalen Qualitätskonzeptionen (Hochschulrektorenkonferenz 2000a) lassen sich keine eindeutigen, für alle beteiligten Akteure verbindlichen und mit dem Curriculum eines Studiengangs angestrebten Ziele sowie diesen Zielen zuzuordnende Ausbildungsmaßnahmen festlegen, die ihrerseits auf ihre Wirkung hin untersucht werden könnten.

Selbst wenn das erfolgreiche Absolvieren des Studiengangs als augenfälligstes und vermutlich unstrittiges Erfolgskriterium zu Grunde gelegt wird, lässt sich diesem nicht eindeutig das Einwirken des Studienprogrammes zuordnen. Vielmehr wird dieser auch durch andere Faktoren, wie dem oben genannten der aufgewendeten Zeit für Erwerbstätigkeit bestimmt.

Im Rahmen der Evaluation von Lehre und Studium ist kaum eine Datenerhebungssituation denkbar, die es erlaubt, im Sinne experimentallogischer Untersuchungsdesigns Wirkungsgruppen und Kontrollgruppen festzulegen und voneinander zu trennen. Die Mitglieder der Kontrollgruppe wären in dieser Logik von der Teilnahme am Studium auszuschließen, um die gewünschten Kontrollen von Störvariablen vornehmen zu können. Evaluation von Lehre und Studium wird aber „unmittelbar im aktiven Feld“ (Kromrey 2000) des laufenden Studienbetriebs durchgeführt. Da alle Studierenden am Studium teilnehmen müssen, ist es praktisch unmöglich, die Evaluation von Lehre und Studium als echtes soziales Experiment zu konzipieren (Kromrey 2001b).

Das Fehlen eines handhabbaren und belastbaren Untersuchungsdesigns zur Analyse von Wirkungen führte zu einem Abrücken von dem methodologisch hohen Anspruch der (Lehr-)Evaluationsforschung, kausale Wirkungszusammenhänge analysieren zu wollen, bzw. ausschließlich solche Evaluationsstudien als nutzbringend anzuerkennen, die auf einem kontrollierten Experimentaldesign beruhen (Grob 2003). „(...) Measuring program outcomes requires a great deal of rigor; it's demanding. When program evaluators discovered how difficult it was, many abandoned the attempt and decided to focus on process, which is generally more tractable. (...) Rather than assuming the difficult tasks of improving the designs and developing better measures (...) program evaluators decided to focus on something else - process“ (Smith 1994). In prozessorientierten Evaluationen liegt der Fokus auf der Untersuchung der Planung und Durchführung der zu evaluierenden Programme. Sie werden daher auch als Implementationsforschung bezeichnet (Kromrey 2001a; Kromrey 2001b). Im Vordergrund steht hier die beratende Funktion der Evaluation für die Verantwortlichen der Programmdurchführung. Während des Programmverlaufs werden durch sie Steuerungsentscheidungen erleichtert, indem sie auf relevante Informationen gestützt werden können (Stockmann 2000, S. 13). Generell hat eine derartige Begleitforschung den Zweck, „to describe and analyze the process of implemented program activities -management, strategies, operations, costs, interactions among clients and practitioners and so forth- so as to improve them“ (Chelimsky 1985). Gemäß des begleitenden Cha-

racters dieser Form der Evaluation wird sie auch als formativ bezeichnet, da ihre Ergebnisse noch während des Programmablaufes in dessen Durchführung zurückgemeldet werden und diesen auf Informationen gestützt beraten und verbessern können. „Die formative Evaluation (*Verbesserungsevaluation*) setzt sich zum Ziel, systemintern Lernprozesse auszulösen und Ansatzpunkte zur Verbesserung aufzuzeigen“ (Widmer 2000). Die Ergebnisse von formativen Evaluationen (Begleitforschungen) sind zwar besonders schwierig im Sinne einer Erfolgs- oder Wirkungskontrolle zu interpretieren, da die „Forschung den Gegenstand der Bewertung selbst fortlaufend beeinflusst und verändert“ (Kromrey 1995, S. 318). Allerdings sind sie für die Zwecke der Qualitätssicherung und -entwicklung von Programmen in hohem Maße geeignet, da sie als dynamisches Entwicklungsmodell zu sehen sind, das auch auf die Begrenzung unerwünschter Programmeffekte zielt (a.a.O.). Auch Hager (2000) stellt fest, dass wenn das Ziel der Evaluation in der Entwicklung einer Maßnahme zur Qualitätssicherung besteht, die an die gewählte Evaluationspraxis zu stellenden Forderungen hinsichtlich methodologischer Kriterien weniger streng ausfallen können, als es für wissenschaftliche Evaluationen mit kausalanalytischem Forschungsinteresse bezüglich der Programmwirkungen und -wechselwirkungen zutrifft. Wenn also das Interesse der Evaluation in der Untersuchung der Programmdurchführung und seinen Randbedingungen besteht, sind auch Forschungsdesigns, die die strengen Bedingungen der Experimentallogik nicht erfüllen durchaus als probates Instrument anzusehen (a.a.O.).

Dies trifft durchaus für die Evaluation von Lehre und Studium zu, sofern diese als internes Instrument der Selbstreflexion und Selbstvergewisserung aufgefasst wird. Ziel von Evaluation in diesem Verständnis ist es, zur Qualitätssicherung und -entwicklung beizutragen. Evaluation von Lehre und Studium hat die Aufgabe, „die Leistungen der Fachbereiche umfassend zu dokumentieren, Aufschluss über die Mittelverwendung zu geben, Maßstäbe für die aufgabenbezogene Bewertung der Qualität und Verfahren zur Qualitätssicherung auf Fach-, Fachbereichs- und Hochschulebene zu entwickeln“ (Reil 2002). Evaluation und Qualitätssicherung von Lehre und Studium orientieren sich entsprechend an einem multidimensionalen Qualitätsverständnis, das Lehre und Studium nicht allein in der Kategorie des



durch die Ausbildung vermittelten Effekts zu beschreiben, sondern dessen Qualität relational zu ermitteln versucht (Pellert 2002, S. 24) und dementsprechend auf *Qualitätsmessungen* verzichtet, da sich die Qualität von Lehre und Studium einer direkten Messung entzieht.

Studienqualität ist nur sehr bedingt im Sinne kausaler (bspw. zu beruflicher und akademischer Qualifizierung beitragender) Effekte zu messen, weil die entsprechenden messtheoretischen Voraussetzungen nicht gegeben sind. Daher besteht die Aufgabe der Evaluation als Qualitätssicherung in der Bereitsstellung von Informationen, „die es erlauben, die Güte der zu evaluierenden Sachverhalte aus unterschiedlichen Perspektiven einzuschätzen“ (Kromrey 2000). Einem solchen Anspruch an die Evaluation von Lehre und Studium werden auch solche Datenerhebungen gerecht, die nicht in einer methodologisch strengen experimentologischen Untersuchungsanordnung erhoben wurden. Studentische Befragungsdaten, auf die sich die Evaluation von Lehre und Studium vielfach stützt, können wertvolle Informationen über die Studiensituation und die Studienqualität liefern, sofern sie valide und verzerrungsfreie Qualitätseinschätzungen widerspiegeln (Engel und Pohlentz 2001). Die Ergebnisse derartiger Untersuchungen können mit den selbst gesteckten Qualitätszielen der Verantwortlichen kontrastiert werden, um zu überprüfen, wo und inwieweit Handlungsbedarf zur Verbesserung der Situation besteht. Auf diese Weise wird die Evaluation von Lehre und Studium ihrem Anspruch gerecht, einen Beitrag zur Qualitätssicherung und -entwicklung von Lehre und Studium zu leisten.

## 3.2 Paradigmen der Evaluation

Mit den oben beschriebenen verschiedenen Ansprüchen an Evaluation sind unterschiedliche methodologische Anforderungen an die entsprechenden Forschungsdesigns und damit die Begründung des zu Grunde gelegten Evaluationsparadigmas verbunden. Aus den Sozialwissenschaften wurde eine Vielzahl von Paradigmen in die Evaluationsforschung übernommen, die sich in der Anwendung auf deren Ziele und Kontexte teils in scharfem Kontrast und entsprechendem Widerstreit

zueinander befinden. Kennzeichnend für diesen Streit in der Entwicklung der noch vglw. jungen Disziplin der Evaluationsforschung ist dabei insbesondere der Gegensatz zwischen den beiden Hauptrichtungen (Stockmann 2000; Patton 1997; Knaap 2004, S. 26 ff.) der empirisch-analytischen Herangehensweise der rationalistischen Evaluation und der konstruktivistisch oder naturalistisch orientierten Evaluation. Zwischen diesen Polen des Kontinuums verschiedener paradigmatischer Ansätze und Positionen wurde in den letzten Jahrzehnten eine zuweilen äußerst kontroverse Debatte geführt, deren Schärfe dadurch gekennzeichnet wird, dass sie von verschiedenen Autoren auch als der „kalte Krieg“ der Paradigmen in der Evaluation bezeichnet wird (Lee 2000; Stockmann 2000). „These differences have been the subject of prolonged and persistent debate, sometimes characterized as the paradigm wars“ (Lee 2000, S. 140).

Guba und Lincoln (1985) fassen die wichtigsten axiomatischen Unterschiede der genannten gegensätzlichen paradigmatischen Positionen, hinsichtlich ihrer erkenntnistheoretischen Ansprüche zusammen (vgl. Tabelle 3.1).<sup>3</sup> Nachfolgend sollen die Argumente der verschiedenen Seiten kurz einleitend dargestellt werden.

#### 3.2.1 Rationalistische vs. Naturalistische Evaluation

Im rationalistischen Paradigma der Evaluationsforschung (Beywl 1988, S. 92) wird diese als anwendungsorientierter Teilbereich empirischer Sozialforschung aufgefasst. Die in diesem Zusammenhang eingesetzten Methoden der empirischen Sozialforschung wenden eine Forschungslogik an, die die Tradition des kritischen Rationalismus aufgreift.<sup>4</sup> Dabei werden -anders als in der Grundlagenforschung- spezifische Erkenntnis- und Verwertungsinteressen an den Evaluationsergebnissen,

---

<sup>3</sup>Zit. nach Beywl (1988).

<sup>4</sup>Vgl. Axiom 1 in Tabelle 3.1; Kernelement der Theorie des kritischen Rationalismus besteht darin, dass es nur eine objektiv beschreibbare Realität gibt, an die sich die Forschung durch das Eliminieren von Irrtümern mittels der Falsifikation von Hypothesen annähern kann (z.B. Prim und Tilman 1997; Popper 1993; Popper 1994).

<i>Axioms about...</i>	<i>Positivist Paradigm</i>	<i>Naturalist Paradigm</i>
... the nature of reality	Reality is single, tangible and fragmentable.	Realities are multiple, constructed and holistic.
... the relationship of knower to the known	Knower and known are independent, a dualism.	Knower and known are interactive, inseparable.
... the possibility of generalization	Time- and context-free generalizations (nomothetic statements) are possible.	Only time- and context-bound working hypotheses (idiographic statements) are possible.
... the possibility of causal linkages	There are real causes, temporally precedent to or simultaneous with their effects.	All entities are in a state of mutual simultaneous shaping, so that it is impossible to distinguish causes from effects.
... the role of values	Inquiry is value-free.	Inquiry is value-bound.

Tabelle 3.1: Axiome der Evaluationsparadigmen

bspw. durch den Evaluationsauftraggeber berücksichtigt. Der Einsatz dieser Forschungsmethoden ist darauf ausgerichtet, die Realität des evaluierten (Studien-)Programms objektiv zu beschreiben. Ziel ist es, empirische Regelmäßigkeiten des evaluierten Gegenstandes aufzuzeigen und generalisierbare Aussagen über diese zu machen (Beywl 1988, S. 94). Das diesem Anspruch zugrunde liegende rationalistische Paradigma wird auch als empirisch-wissenschaftliches (Stockmann 2002, S. 227 ff.) bzw. empirisch-analytisches Paradigma bezeichnet. Die Grundannahme dieser bspw. von Guba und Lincoln (1985) als „positivistisches Paradigma“ bezeichneten Forschungstradition besteht darin, dass eine unabhängig und außerhalb der Forschungssubjekte bestehende Realität vorliegt, die als beobachtbar gegeben ist (Beywl 1988, S. 93). Aus dieser Beobachtung, die aus einer außen stehenden (neutralen und objektiven) Forscherperspektive vorgenommen wird, lassen sich Gesetzmäßigkeiten des evaluierten Gegenstandes (bspw. Studienprogramms) ableiten. Ziel der Forschung bzw. der Evaluation ist es, zu generalisieren, verallgemeinernde Vorhersagen über Wirkungen und Effekte der evaluierten Programme vorzunehmen und diese Effekte hinsichtlich vorgegebener Bewertungsstandards zu beurteilen.

Das zentrale Kriterium ist dabei zumeist die Effizienz oder Effektivität im weitesten Sinne der Bedeutung dieser Begriffe. Bezogen auf Bildungsprogramme

ließe sich die Effizienz in dem Maß der Qualifizierung aufgrund der Teilnahme an dem Bildungsprogramm und gemäß der mit diesem Programm verbundenen Bildungsziele darstellen. Grundlage für die Bewertung von Programmen und ihren Effekten stellen in der rationalistischen Forschungslogik Kausalbeziehungen dar, die bspw. die Wahrscheinlichkeit des Eintretens eines bestimmten Effekts durch den Einsatz einer bestimmten Ressource beschreiben. „Evaluative research (...) begins with the statement of a causal relationship hypothesized between some program or activity (the independent variable) and some desired effect (the dependent variable). The verification of this evaluative hypothesis requires the design of a research project, which would show that the desired effect was more likely to occur in the presence of the program being evaluated than in its absence“ (Suchman 1967). Evaluationsdesigns dieses Zuschnitts wurden vielfach in den 1960er und 1970er Jahren durchgeführt, wie auch das genannte Zitat verdeutlichen soll, das einem diesbezüglichen Standardwerk dieser Zeit entnommen ist. Die oben beschriebene Erkenntnis, dass -insbesondere im Fall von Bildungsprogrammen- kausale Wirkungen von Programmen auf die teilnehmenden Personen nur sehr schwer im Rahmen von Evaluationen zu analysieren sind, ist dementsprechend in deren Konzeption noch nicht berücksichtigt worden. Gleichwohl hat der Anspruch, dass das Ergebnis einer Evaluation in einer generalisierenden Beschreibung des beurteilten Programms besteht, nicht an Aktualität verloren. So ist bspw. der Diskussion der Evaluation von Hochschulleistungen vielfach zu entnehmen, dass Effizienzkriterien, die sich als das Ergebnis von Kausalbeziehungen (im Verhältnis von Input zu Output) beschreiben lassen, eine überaus zentrale Stellung einnehmen und dass die Erwartung an Evaluation gerichtet wird, diese in die entsprechenden Konzeptionen einzubeziehen.

Gemäß dem Anspruch, durch Evaluation verallgemeinernde Beschreibungen der Realität zu ermöglichen, auf die Bewertungen gestützt werden können, werden in den entsprechenden Forschungen, die dem rationalistischen Paradigma verpflichtet sind, zu einem großen Teil standardisierte empirische Forschungsmethoden eingesetzt (bspw. standardisierte Befragungen von Programmbeteiligten, etc.). Von diesen wird die Erhebung repräsentativer (also verallgemeinerungsfähiger)

ger) Daten erwartet. „The scientific, largely quantitative methods used in evaluation (...) resulted in considerable emphasis in the profession on skill in the use of statistical analysis“ (Lee 2000, S. 140). Gleichzeitig leitet sich aus dem Anliegen einer systematischen Anwendung (quantitativer) sozialwissenschaftlicher Forschungsmethoden auch die zentrale Bedeutung von Testgütekriterien für die Qualität der entsprechenden Forschung ab (Beywl 1988, S.50): Validität und Reliabilität von Messungen werden als wichtige Kriterien zur Beurteilung des Forschungsprozesses bzw. der in ihm erhobenen Daten eingeführt (z.B. Rossi und Freeman 1993; Beywl 1988, a.a.O.). Anhand dieser Kriterien für die Beurteilung der Datenerhebung kann überprüft werden, inwieweit, diese zuverlässige und gültige Aussagen der befragten Personen erhoben haben und eine entsprechend rationale Bewertungsgrundlage zur Verfügung stellen.

Im Widerspruch zum empirisch-analytischen oder rationalistischen Forschungsparadigma befindet sich auf der anderen Seite des Spektrums das naturalistische Paradigma der Evaluation (Beywl 1988, S.92). Zentrales Kennzeichen der entsprechenden Forschungsrichtung ist die Ablehnung des Vorhandenseins einer einzigen objektivierbaren Realität des evaluierten Gegenstandes. Vielmehr wird davon ausgegangen, dass die Realität des Programms aus den verschiedenen individuellen Perspektiven der Beteiligten und vor dem Hintergrund ihrer unterschiedlichen Interessen sozial konstruiert wird. „According to constructivists there can be no universally applicable claims to truth or reality“ (Knaap 2004, S. 27). Das Erkenntnisinteresse von Evaluation in diesem Sinne besteht in der Analyse der konfliktiven Sichtweisen, Handlungen und Präferenzen der Beteiligten sowie der sich in diesem Spannungsfeld ergebenden Machtasymmetrien. In diesem Zusammenhang ist „Evaluation (...) deshalb nicht eine möglichst realitätsnahe Darstellung und Bewertung der vorgefundenen Verhältnisse, sondern eine Veränderung dieser Verhältnisse zu Gunsten der Benachteiligten. Evaluation wird zu einem transformatorischen Akt“ (Stockmann 2002, S. 227 ff.). In diesem Zusammenhang wird auch von einem emanzipatorischen oder transformatorischen Paradigma der Evaluation gesprochen (z.B. Lee 2000; Mertens 1998; Mertens 2000). Die Bezeichnung als naturalistisches Paradigma leitet sich von

der zentralen Annahme entsprechender Forschungsansätze ab: „(...) Inquiry must be carried out in a ‘natural setting‘ because phenomena of study, (...) take their meaning as much from their context as they do from themselves. (...) Reality constructions cannot be separated from the world in which they are experienced and (...) any observations (...) are inevitably time- and context-dependent“ (Guba und Lincoln 1981, S. 189). Analog zur deutlichen Betonung der Wichtigkeit der Kontexte von Evaluationsgegenständen und beteiligten Personen sowie der Interpretationen und sozialen Konstruktionen von Realitäten in diesen Kontexten, wird von Anhängern dieser Forschungsrichtung der Einsatz eher qualitativ orientierter Forschungsmethoden präferiert (Stockmann 2000, S. 18). So kommt als ein gängiges Verfahren in diesem Zusammenhang bspw. die Technik des Naturalistic Interviewing zum Einsatz (vgl. Greene 1989; Greene 1989a). Dieses „is designed as a small-group discussion activity. The activity is oriented around key issues that arise when conducting open-ended, unstructured qualitative interviews within a naturalistic evaluation framework“ (Greene 1989, S. 123 ff.).

#### 3.2.2 Perspektiven für die Versöhnung

Zwar ist der Krieg der Paradigmen inzwischen ausgetragen. Mittlerweile wird bei der Wahl des Untersuchungsdesigns eine Kombination der ehemals unversöhnlichen Paradigmen und mithin der verschiedenen methodischen Ansätze angestrebt. „One can believe in both objective and subjective reality, and still choose a variety of methods“ (Lee 2000, S. 135). Nichtsdestotrotz ist die beschriebene Schulbildung auch für zeitgenössische Evaluationsansätze oder Konzeptionen von Qualitätssicherungsstrategien bedeutsam geblieben, was auch an der Begründung der eingesetzten Forschungsmethoden sichtbar wird. So orientieren sich bspw. Ansätze partizipativer Evaluation, deren zentrales Merkmal ein Einbezug der Perspektiven beteiligter Personen (oder „Evaluationssubjekte“) in die Konzeption und Durchführung der Evaluation ist, stärker an qualitativen, auf Kommunikation und Diskurs ausgerichteten Forschungsmethoden als Evaluationsansätze, die auf eine Beschreibung der Effektivität von Programmen hinsicht-

lich fester Qualitätskriterien abzielen.<sup>5</sup> Nicht zuletzt anhand der unterschiedlichen Präferenzen bei der Wahl von Forschungsdesigns für die Umsetzung evaluationsmethodologischer bzw. wissenschaftstheoretischer Grundlegungen und Fragestellungen lässt sich vermuten, dass die Übernahme der unterschiedlichen Paradigmen nach wie vor in weitgehend getrennten Sphären praktischen Evaluationshandelns mündet.

Um die Paradigmen auch theoretisch begründet und nicht lediglich aufgrund forschungspragmatischer Erwägungen in Einklang zu bringen und entsprechend für eine integrierte Forschungsperspektive nutzbar zu machen, können verschiedene Konzepte herangezogen werden. So schlägt Beywl (1988) schon früh die von Max Weber entworfene Wirklichkeitswissenschaft als Weg zur Versöhnung rationalistisch und naturalistisch orientierter Evaluationsansätze vor. Ein anderer diesbezüglicher Ansatz kann in dem Konzept der Theory-based Evaluation gesehen werden. Beide werden nachfolgend kurz beschrieben.

#### 3.2.2.1 Wirklichkeitswissenschaft

Grundlegend für das Konzept der Wirklichkeitswissenschaft, welches von Max Weber in den „Gesammelten Aufsätzen zur Wissenschaftslehre“ (Weber 1988) formuliert und hier insbesondere im „Objektivitätsaufsatz“<sup>6</sup> dargelegt wurde (Tenbrück 1986, S.15), ist die Unterteilung in zwei unterschiedliche Wissenschaften, die sich aufgrund ihres verschiedenen Erkenntnisinteresses ergibt. Bei diesen handelt es sich um die Gesetzeswissenschaft und die Wirklichkeitswissenschaft. Das Erkenntnisinteresse der Gesetzeswissenschaft bezieht sich auf die Ermittlung der allgemeinen Gesetzmäßigkeiten ihres Gegenstandes, während die Wirklichkeits-

---

<sup>5</sup>Verschiedene Evaluationsansätze, die unter Berücksichtigung der beschriebenen Paradigmen für die Qualitätssicherung von Lehre und Studium eingesetzt werden, werden unten ausführlicher diskutiert (vgl. Kapitel 4). Dabei wird auch auf die Frage nach ihrer Nützlichkeit für das genannte Ziel der Qualitätssicherung eingegangen.

<sup>6</sup>Vollständiger Titel: „Die Objektivität sozialwissenschaftlicher und sozialpolitischer Erkenntnis“ (Weber 1988, S. 146 ff.).

wissenschaft nach dessen spezifischen (stets historischen) Eigenarten und Besonderheiten forscht.

Beide Konzepte sind wissenschaftslogisch widerspruchsfrei möglich, so dass die Wahl für die eine oder andere nicht wissenschaftstheoretisch begründet, sondern gemäß der mit der Erkenntnis verbundenen Ziele und Interessen getroffen wird. Die entsprechende Entscheidung ist von großer Tragweite, weil sie Sinn, Ziel und Wesen der Wissenschaft mitbestimmt und die Entscheidung mithin die vom Forscher an die Forschung geknüpften Werte widerspiegelt.

Weber hat gegen die Gesetzeswissenschaft Partei genommen (Tenbruck 1986, S. 15).<sup>7</sup> Deren Grundauffassung, dass alle Wissenschaft es sich zum Ziel setze, die Gesetzmäßigkeiten ihres Gegenstandes in einer systematischen Theorie zu erfassen, die es erlaube, Vorhersagen und technische Beherrschung zu erreichen, stellte Weber sein Konzept der Wirklichkeitswissenschaft entgegen. Den Versuch, der Entwicklung eines definitiven Systems von Begriffen zwecks der Entwicklung einer definitiven Theorie, aus der sich die Wirklichkeit vorhersagen lasse, bezeichnete Weber als Irrweg (Tenbruck 1986, S. 18). Die Formulierung einer endgültigen Gesellschaftstheorie könne schon deshalb nicht das Endziel der Sozial- und Kulturwissenschaften sein, weil sich die Kulturprobleme, die die Menschen bewegen, immer wieder ändern. Wissenschaft wird aber betrieben, um praktische Probleme zu lösen. Sie muss daher hinsichtlich der Begriffs- und Theoriebildung immer wieder neue, an die jeweiligen Probleme angepasste Wege gehen. Dementsprechend begrenzt sich die Objektivität der wissenschaftlichen Arbeit jeweils auf einen begrenzten Ausschnitt der Wirklichkeit.

Der Entwicklung von (gesetzeswissenschaftlichen) Theorien zur endgültigen Abbildung der Wirklichkeit werden als methodisches Instrument in diesem Zusammenhang Idealtypen gegenübergestellt. Erstere haben das Ziel, die Wirklichkeit umfassend über Gesetzmäßigkeiten zu erklären, mit dem Ergebnis, dass sie,

---

<sup>7</sup>Diese Parteinahme ist vor dem Hintergrund des Methodenstreits zu beurteilen, „in dem grob gesagt Positivismus und Historismus über die Grundfragen der Erkenntnis auf dem Gebiet der Kultur- und Sozialwissenschaften rangen“ (Tenbruck 1986, S. 15).



sofern ihnen dieses nicht gelingt, abgeändert oder weiterentwickelt werden, bis sie in der Lage sind, Wirklichkeit in gesetzeswissenschaftlich begründeten Abläufen abzubilden. Idealtypen dagegen stellen eine spezifische Art der Begriffsbildung dar, die eher als die Verstehbarkeit sozialer Phänomene anstrebende Konstrukte bezeichnet werden können (Hauck 1993, S. 88). Ziel ist es nicht, die Wirklichkeit in Begriffen und Gesetzmäßigkeiten zu fixieren. Vielmehr besteht in der Untersuchung der Abweichungen realer Handlungen von den mit diesen Handlungen verbundenen idealtypischen, bei Vorliegen von Handlungsrationalität erwarteten Ausprägungen eine Möglichkeit, historisch-gesellschaftliche Phänomene zu beschreiben und zu verstehen. Idealtypen dienen der Untersuchung der Besonderheiten von sozialen Phänomenen. Anhand des Abweichens dieser Phänomene vom Idealtyp können sie in ihrer historisch-gesellschaftlichen Einzigartigkeit beschrieben werden.

Wendet man diese Überlegungen auf den Paradigmenstreit zwischen rationalistischen und naturalistischen Evaluationsansätzen an, ließe sich mit Weber argumentieren, dass Evaluationen nur in dem jeweils spezifischen Kontext des Evaluationsgegenstandes angesiedelt, anschlussfähige Erkenntnisse über dessen Eigenarten hervorbringen, die ihrerseits für eine Bewertung des Gegenstandes genutzt werden können. Diese Position spricht der Auffassung naturalistischer Evaluation das Wort. Gleichzeitig wird aber auch argumentiert, dass die in ihrem spezifischen Kontext evaluierten Gegenstände durchaus eine beschreibbare, mithin objektivierbare Realität besitzen und dass sie entsprechend auch einer empirischen Untersuchung zum Zweck der Beschreibung, Standardisierung und Schaffung von Vergleichbarkeit zugänglich sind. Beywls Vorschlag eines Rückgriffs auf Webers Konzept der Wirklichkeitswissenschaft ist zwar bereits vor nahezu 20 Jahre vorgestellt worden, hat aber hinsichtlich einer vielfach beklagten fehlenden Theoriebindung von Evaluation (Schmidt 2005) nicht an Aktualität eingebüßt. Der Kern der Diskussion besteht in einer Konzeption und Durchführung von Evaluation vor dem Hintergrund einer (theoretischen) Reflexion über die Art des Erkenntnisinteresses und die Eigenarten (bspw. bezüglich des Sinns und der Werte- oder anders formuliert: die Qualitätsdimensionen) des evaluierten Gegenstandes.

### 3.2.2.2 Theory-based Evaluation

Diese fehlende theoretische Reflexion über den Evaluationsgegenstand wird im Konzept der theoriebasierten Evaluation angestrebt, welches seit einigen Jahren erneut verstärkt in der Evaluationsliteratur aufgegriffen wird (Knaap 2004, S. 16).

In der theoriebasierten Evaluation stellen die mit dem evaluierten Programm<sup>8</sup> verbundenen (impliziten und expliziten) Annahmen und die auf diesen Annahmen basierenden Programmaktivitäten sowie schließlich die normativen Ziele des Programms den Ausgangspunkt der Evaluation dar. Unter den Annahmen eines Programms werden die unterstellten Effekte verstanden, bspw. in dem Sinne, dass an die Durchführung einer hochschuldidaktischen Weiterbildungsmaßnahme für Lehrende die Erwartung verknüpft ist, dass diese Maßnahme zu einer Verbesserung der Lehrqualität beitragen wird. Eine Verbesserung der Lehrqualität wäre entsprechend das normative Ziel der durchgeführten Programmaktivität, während sich aus dem Verständnis darüber, was unter Lehrqualität zu verstehen ist, die Inhalte der durchgeführten Qualifizierungsmaßnahme ableiten ließen.

Ausgehend von diesen Elementen basiert die Konzeption des Evaluationsdesigns auf einem theoretischen Modell des Programms (vgl. z.B. Christie und Alkin 2003; Green und McAllister 1998; Garacelli 2000). „An evaluation that is designed to develop a program theory which in turn is used to guide the program evaluation, is known in the evaluation theory literature as a Theory-Driven Evaluation (TDE)“ (Christie und Alkin 2003, S. 374).

Ziel ist es, Hypothesen zu den (kausalen) Wirkungszusammenhängen zwischen Programmaktivitäten und zu erwartenden (intendierten sowie nicht intendierten) Effekten zu entwickeln und diese empirisch zu testen. „Evaluation seeks to discover whether programmes work; programmes are theories. Therefore it fol-

---

<sup>8</sup>Auch hier wird durch die Verwendung des Begriffs „Programm“ auf den konkreten Fall von Studiengängen rekuriert.

lows that: evaluation is theory-testing“ (Pawson 2003, S. 472). Die Gleichsetzung von Programmen und Theorien hebt dabei auf die Tatsache ab, dass der Planung und Durchführung eines jeden Programmes eine (theoretische) Annahme über die Art der Auswirkung dieser Programmdurchführung vorangeht, die ihrerseits letztlich Gegenstand der entsprechenden Evaluation ist.

Die Stärke des Ansatzes der theoriebasierten Evaluation wird darin gesehen, dass sie dazu beiträgt, der Komplexität des kausalen Wirkungsgeflechts von Programmaktivitäten und den Einflüssen der Programmumwelt gerecht zu werden. „One of the involuntary virtues of the theory-driven approach to evaluation is that it forces us to contemplate programmes in their true and awesome complexity. By starting with the programme theory, one understands immediately just how many and varied the processes are that may lead to an intervention’s success or failure“ (Pawson 2003, S. 483).

Im Zusammenhang mit der Frage nach einer Synthese zwischen den oben beschriebenen widerstreitenden Evaluationsparadigmen lässt sich die entsprechende Bedeutung des Ansatzes der „theory-based evaluation“ damit erklären, dass ihr Gegenstand die Frage danach ist, welche Programmaktivitäten in welchen Kontexten für welche Personengruppen funktional oder (positiv) wirksam sind. Diese Frage wird durch empirische Analysen, die sich auf kausale Beziehungen zwischen Aktivitäten und zugehörigen Wirkungen beziehen, untersucht (Hansen 2005). Während (Programm-)Evaluationen, die dem oben beschriebenen rationalistischen Paradigma folgen, auf ein bestimmtes, festgelegtes Untersuchungsziel ausgerichtet sind (bspw. auf die Klärung der Frage nach dem Ausmaß der Vermittlung berufsqualifizierender Kompetenzen durch ein Studienprogramm), erlauben theoriebasierte Evaluationen es, einen holistischen Blick auf die Programmwirklichkeit zu werfen und die Komplexität der (Wechsel-)Wirkungen der in dem Programm durchgeführten Aktivitäten zu analysieren. „Where the classical results models focus only on concrete goals or effects related to a specific intervention or organization, and aim to evaluate the results of these, the programme theory model opens up the underlying black box of the programme theory, uncovers mechanisms and raises the focus to a cluster of interventions or to an organizational

field. The aim of the programme theory model is to revise and further develop programme theory and thus learn what works for whom in which contexts“ (Hansen 2005, S. 450).

Im Ansatz der theoriebasierten Evaluation werden eine kontextsensible Untersuchung des Evaluationsgegenstandes mit dem Anspruch einer kausalanalytischen, auf Wirkungen bezogenen empirischen Untersuchung von Programmeffekten vereint. Kausalanalytische Untersuchungen zeichnen sich dadurch aus, dass sie geeignet sind, empirisch feststellbare Effekte und Zusammenhänge nicht nur als solche zu beschreiben, sondern in ihren ursächlichen Wirkungszusammenhängen und Wirkungsrichtungen zu analysieren. Die unten beschriebenen Datenanalysen greifen diesen Anspruch insofern auf, als sie versuchen, die kausalen Verknüpfungen zwischen verschiedenen Elementen bzw. Beurteilungskategorien von Studienprogrammen zu beschreiben. Dabei wird nicht der Versuch gemacht, eine vollständige implizite Programmtheorie (z.B. im Sinne eines Modells zur Erklärung der Auswirkungen des Studiums auf die „Berufstauglichkeit“ der Absolventen) zu testen. Vielmehr besteht die Absicht der berichteten Analysen insgesamt darin, die Datenqualität der für Evaluationszwecke herangezogenen Erhebungen zu untersuchen. Jedoch dienen die berichteten kausalanalytischen Modelle durchaus dazu, über die rein deskriptive Untersuchung von Zusammenhängen verschiedener Qualitätskriterien des Studiums hinaus, zu einer hypothesengeleiteten Erklärung ausgewählter Beurteilungsaspekte der Studienqualität beizutragen. So stellt bspw. die Feststellung, dass Studierende das Studium oder eine Lehrveranstaltung als zu stark oder zu wenig stark strukturiert empfinden, an sich nur sehr bedingt eine anschlussfähige Information dar, aufgrund derer das Studienangebot weiterentwickelt werden kann. Hierfür wäre vielmehr nötig zu ermitteln, welche latenten Beurteilungskategorien sich hinter dem studentischen Urteil verbergen, bzw. was sie unter einem „gut strukturierten“ Studium verstehen. Insofern finden die Prinzipien der theoriebasierten Evaluation über ihren theoretischen Beitrag hinsichtlich der Versöhnung verschiedener Evaluationsparadigmen hinaus, auch Anwendung bei der Untersuchung der im Kontext der vorliegenden Arbeit relevanten Fragestellungen.

# Kapitel 4

## Ansätze der Qualitätssicherung für die Hochschulentwicklung

Der Hauptanspruch, der an Evaluation von Lehre und Studium im allgemeinen gestellt wird, besteht darin, einen relevanten Beitrag zu Qualitätssicherung und -entwicklung in den Hochschulen zu leisten. Evaluation steckt jedoch in einer diesbezüglichen „paradigmatischen Zwickmühle“ (Ehlert und Drees 2004): einerseits soll sie die Selbstreflexion der beteiligten Hochschulmitglieder anregen und diese dazu ermutigen, in eigenverantwortlicher Weise die Qualitätsbeurteilung und Qualitätsverbesserung zu befördern, mit dem Ziel und damit verbundenen Anreiz einer Steigerung der eigenen Reputation und Leistungsstärke. Andererseits hat sie durchaus einen von außen in die Hochschule eingreifenden Kontrollcharakter (a.a.O.), ihre Durchführung ist in den Hochschulgesetzen der Bundesländer als gesetzliche Forderung verankert<sup>1</sup> und ihre Nutzung für die leistungsorientierte Mittelvergabe wird zunehmend empfohlen (vgl. z.B. Kultusministerkonferenz). Neben der Frage, ob und wie diese gegenläufigen Aspekte von Evaluation in-

---

<sup>1</sup>Informationen zu den Hochschulgesetzen der Bundesländer und der dort fixierten Rolle der (Lehr-)Evaluation finden sich zusammengefasst im Internetangebot der Hochschulrektorenkonferenz unter [http://www.hr.k.de/de/adressen\\_und\\_links/158\\_164.php](http://www.hr.k.de/de/adressen_und_links/158_164.php) (Stand 30.4.2007).

---

tegiert werden können<sup>2</sup>, ist die Frage zu diskutieren, wie bzw. durch welche Konzepte der Qualitätssicherung substantielle Beiträge zu einer wirklichen Qualitätsverbesserung von Lehre und Studium erreicht werden können. Diese auf die Umsetzung oder Anwendung von Evaluationsergebnissen bezogene Frage hebt sowohl auf Verbesserungsmöglichkeiten bei der didaktischen und pädagogischen Qualität einzelner Lehrveranstaltungen ab, als auch - und insbesondere - auf die Organisationsqualität des Studiums. Diese spiegelt sich einerseits in den Ergebnissen des Organisationshandelns (Angebot studierbarer Studiengänge, Aufbau eines effizienten und transparenten Prüfungswesens, etc.), aber andererseits auch in den Prozessen, die zur Erbringung dieser Leistungen geführt haben, wider (Führungsstil, Innovationsfähigkeit, etc.).

Gemein ist beiden genannten Dimensionen, dass in dem Vorliegen einer relevanten Informationsbasis der Ausgangspunkt bzw. die Voraussetzung für die Entwicklung von Handlungsstrategien mit dem Ziel zukünftiger Verbesserungen der Qualität zu sehen ist. Darauf aufbauend sind Verfahrensweisen zur organisationalen Nutzung dieser Informationen zu entwickeln, die im Dienste einer angestrebten Qualitätsentwicklung Erfolg versprechend sind. Innerhalb einer solchen Strategie muss die Frage nach den methodologischen Anforderungen hinsichtlich der Qualität der für Entscheidungsprozesse benutzten Informationen einen festen Platz haben, ebenso wie die entsprechenden Überprüfungen dieser Informationen und Daten. Als Ziel der Qualitätssicherung und in diesem Sinne als das wünschenswerte Ergebnis einer nach innen gerichteten und auf die Selbstreflexion der Hochschulmitglieder zielenden Evaluation könnte die Etablierung einer Qualitätskultur bezeichnet werden. Diese kann als eine Form der Hochschulsteuerung verstanden werden, die sowohl auf den Einbezug der Perspektiven der Beteiligten und deren geteiltes Qualitätsverständnis, als auch auf eine (datengestützte) Leistungsbeschreibung für Zwecke des Qualitätsmanagements setzt.

---

<sup>2</sup>Von manchen Autoren wird das Verhältnis von Rechenschaft und institutionellem Lernen als Ergebnis der Durchführung von Evaluationen allerdings weniger als gegensätzlich, sondern eher als komplementär aufgefasst (vgl. z.B. Feinstein 2002.)

---

Das Konzept der Qualitätskultur wird nachfolgend dargestellt. Im Anschluss werden verschiedene Konzepte der Hochschulentwicklung daraufhin diskutiert, inwieweit sie dem Ziel des Aufbaus einer Qualitätskultur zuträglich sein können und in welchem Verhältnis sie zu den formulierten methodischen Ansprüchen stehen, die an die Erhebung von Evaluationsdaten zu stellen sind.

Dabei wird eine (normative) „Prozesskette“ im Sinne eines Programms der qualitätsbezogenen Hochschulsteuerung entwickelt, in der unter Einsatz partizipativer Elemente in der Evaluation (Kap. 4.2.1) Prozesse des Organisationslernens (4.2.2 und 4.2.3) angeschoben und durch Wissensmanagement (4.2.4) an Hochschulen unterstützt werden können. Diese Lernprozesse münden ihrerseits in einer auf valide und reliable Informationen gestützten, „evidenzbasierten“ Praxis der Hochschulsteuerung (4.2.5).

Insbesondere der letztgenannte Aspekt berührt unmittelbar die Frage der Reichweite von Aussagen der Lehrevaluation, die in Kapitel 6 diskutiert wird: Entwicklungsstrategien zur Qualitätssicherung von Lehre und Studium sind insbesondere dann Erfolg versprechend und nachhaltig, wenn die in ihnen getroffenen Entscheidungen sich auf eine im methodologischen Sinne „robuste“ Datenbasis stützen können. Für eine solche stellen die Gütekriterien der Validität und Reliabilität eine wichtige Voraussetzung dar. Insofern ist die Datenqualität der für die strategische Qualitätsentwicklung (von Studiengangsplanung, Curriculumentwicklung, Prüfungsorganisation, etc.) herangezogenen Evaluationen das verbindende Glied der genannten Prozesskette.

Die Entwicklung einer Qualitätskultur in diesem Sinne wird nicht durch einen Beschluss von Hochschulleitungen oder Ministerien erreicht, sondern ist das Ergebnis der Kommunikation und des Engagements der beteiligten Akteure. Die kombinierte Anwendung der verschiedenen nachfolgend diskutierten Strategien und Instrumente für Qualitätssicherung und Hochschulsteuerung kann diese Entwicklung aber möglicherweise begünstigen.

## 4.1 Qualitätskultur

Die begriffliche Abgrenzung dessen, was unter Qualitätskultur zu verstehen sei, entspringt nach Kohler (2004) zwei Denkschulen. Bei diesen handelt es sich um

- die *Kultursichtweise*, in der Qualitätskultur hauptsächlich in Abgrenzung zu ihrem Gegenbegriff, dem Qualitätsmanagement aufgefasst wird. In diesem Verständnis bezieht sich das Qualitätsmanagement auf die technokratische Definition von Qualitätskriterien, von Techniken der Qualitätsmessung, etc. Berührungspunkte von Qualitätskultur und Qualitätsmanagement bestehen jedoch schon darin, dass Hochschulen bei der Steigerung der Qualität nicht ohne Techniken und Instrumente des Qualitätsmanagements auskommen, die sich ihrerseits u.a. auf die Daten- und Informationssammlung zur Beschreibung der zu beurteilenden und möglicherweise zu verbessernden Situation beziehen. Gleichzeitig hat Qualitätskultur aber einen vom Qualitätsmanagement verschiedenen Eigengehalt. Dieser besteht in der Bedeutung des „subjektiv-persönlichen und qualitativ-institutionellen Wissens und Könnens um bzw. von Qualität“ (Kohler 2004). Qualitätskultur liegt ein bottom-up Gedanke zu Grunde, in dem Sinne, dass Qualität einen geteilten Wert darstellt, basierend auf einem gemeinsamen Verständnis und einer gemeinsamen Verpflichtung der Hochschulmitglieder. Damit unterscheidet sie sich vom Qualitätsmanagement, welches eher als top-down „verordnete“ Form der Qualitätssicherung wahrgenommen wird (a.a.O.; vgl. auch Stockmann 2006, S. 83 ff.). Qualitätskultur bezieht sich auf „Werte, Haltungen und Verhalten einer akademischen Gemeinschaft, die aus der Mission der Institution ein geteiltes Qualitätsverständnis ableitet“ (Kohler 2004).<sup>3</sup>

---

<sup>3</sup>Gleichzeitig muss angemerkt werden, dass sich diese Sichtweise des Qualitätsmanagements als *Gegensatz* zum Konzept der Qualitätskultur nicht gänzlich mit dem landläufigen (Selbst-) Verständnis von Qualitätsmanagement an Hochschulen deckt. So werden bspw. Qualitätsmanagementansätze wie im EFQM-Modell bereits Erfolg versprechend auf den Kontext von Hochschulen angewendet (Hahne 2001). Müller-Böling (2001) beschreibt als Qualitätsmanagement



- Durch die *unitarische Sichtweise* wird dagegen deutlicher die begriffliche Verbindung der beiden Konzepte der Qualitätskultur und des Qualitätsmanagements angestrebt, die entsprechend in diesem Konzept nicht als Antithesen gegeneinander gesetzt werden (Kohler 2004). Insbesondere wird dies durch die Spezifizierung des Begriffs des Qualitätsmanagements erreicht. Als institutionell adaptiertes Qualitätsmanagement bezieht dieses den spezifischen Anwendungsbereich der Hochschule ein, Verfahren werden entsprechend vor dem Hintergrund des Hochschulkontextes geplant. Durch diese Modifikation wird bspw. die „mechanistische Übernahme von Qualitätsmanagementsystemen aus der Wirtschaft“ (Kohler 2004) vermieden und vielmehr die spezifische Kultur der Institution, die sich aus einer langen Tradition ergibt, berücksichtigt. Ein „hochschulspezifisches Qualitätsmanagement ist adaptiertes Qualitätsmanagement im Sinne der Integration anerkannter Verfahren der Qualitätsfeststellung, -sicherung und -erhöhung in die Eigenart von Hochschulen (...)“ (a.a.O.).<sup>4</sup>

Qualitätsmanagement ist in dem beschriebenen Verständnis Qualitätsmessung und -beschreibung (vgl. z.B. Hener 2004; Müller-Böling 2001), während Qualitätskultur eher Qualitätswissen und -wollen oder allgemeiner: Qualitätsbewusstsein

---

den kombinierten Einsatz verschiedener Instrumente der Qualitätssicherung (Evaluation, Akkreditierung, etc.), mit der Absicht, die auf der Ebene der jeweiligen Institution gemeinsam formulierten Qualitätsziele und -standards zu erreichen und den Grad der Zielerreichung zu überprüfen. Allerdings wird an der gleichen Stelle insbesondere auch die Bedeutung der Hochschulleitung für das Qualitätsmanagement betont, indem dieser die Rolle zugewiesen wird, Qualität zu planen, Qualitätssicherung zu organisieren, Mitarbeiter zu qualitätsbewusstem Handeln zu motivieren etc. (Müller-Böling 2001, S. 390). Das entsprechende Verständnis von Qualitätssicherung als einer Leitungsaufgabe vermag wahrscheinlich durchaus zu der Interpretation des Qualitätsmanagements als „top-down“-Steuerung beizutragen.

<sup>4</sup>Zu Versuchen der Anpassung von Total Quality Management Systemen (TQM) wie bspw. dem Qualitätsmanagementsystem der European Foundation for Quality Management (EFQM; vgl. Fußnote 3) an Zusammenhänge außerhalb der Wirtschaft, namentlich im Bildungsbereich vgl. z.B. Pasternack (2004); Pasternack (2005); Arnold und Faber (2000).

(Tegethoff 2005) von der individuellen bis zur institutionellen Ebene beschreibt und das Verhalten der beteiligten Personen auf der Basis gemeinsamer Ziele, internalisierter Verhaltensstandards und eines starken Solidaritätsgefühls lenkt (Engels 2004).<sup>5</sup> Beziehungen zwischen Qualitätskultur und Qualitätsmanagement bestehen in Fragen der Kommunikation und Partizipation, des (freiwilligen) Engagements, der Einheit von Wissen und Tun, der Bereitschaft, Veränderungen zu akzeptieren. Beide Konzepte begleiten das Studienangebot und den Prozess der Studienangebotsentwicklung. Beide Elemente ergänzen sich in dem Sinne, dass das Qualitätsmanagement auf das Qualitätsbewusstsein und das Engagement der Akteure angewiesen ist und gleichzeitig dazu dient, dieses und damit die Entwicklung einer Qualitätskultur zu fördern (Kohler 2004).

Andere Verständnisse von Qualitätsmanagement beschreiben dieses als Steuerungskonzept, mit dem die Hoffnung verbunden wird, dienstleistungsbezogenes Qualitätsmanagement auf die Hochschulen und die öffentliche Verwaltung allgemein anwenden zu können (Pasternack 2004, S. 34). Dabei soll das Qualitätsmanagement die „Regelsteuerung der Verwaltung (...) so ergänzen, dass trotz sich vollziehender Komplexitätssteigerungen die Bürgerorientierung verstärkt wird“ (a.a.O.). Durch Qualitätsmanagementsysteme soll insbesondere durch eine Verbesserung der Prozessqualität die Produktqualität gesichert bzw. erhöht werden. „Die Annahme lautet: Geregelte, aufeinander konzeptionell abgestimmte und nachvollziehbare Abläufe führen dazu, Motivation und Leistung der MitarbeiterInnen zu erhöhen, Fehler zu vermeiden und damit Kosten zu reduzieren“ (Pasternack 2004, S. 35). Weiter beschreibt Pasternack (a.a.O., S. 35 ff.) „adaptable Qualitätsmanagement-Instrumente“, unter die überaus verschiedene, bspw. auf die qualitätsbezogene Diskussion, Kommunikation und Motivation der

---

<sup>5</sup>Gleichwohl muss aber angemerkt werden, dass insbesondere die bereits angesprochenen TQM-Systeme sich durch eine umfassende Qualitätsorientierung des Organisationshandelns und die Verpflichtung aller Organisationsmitglieder auf das Ziel der Qualitätsentwicklung auszeichnen. Kennzeichnend ist für sie allerdings, die bereits erwähnte, zentral gelenkte oder „top-down“ durchgesetzte Formulierung von Qualitätszielen und -kriterien (Stockmann 2006).

Beteiligten ausgerichtete Ansätze (Total Quality Management, Partizipatives Management) subsumiert werden.

Hinsichtlich der Entwicklung einer Qualitätskultur sind die Kommunikation der Beteiligten, die auch informell und über verschiedene Ebenen verläuft sowie die Sicherung der Teilhabe verschiedener Statusgruppen an Entscheidungen und den Entscheidungsfindungsprozessen, die Einrichtung von Arbeitsgruppen zur Organisation gemeinsamer Lernprozesse<sup>6</sup>, die Wahrung der Transparenz von Verfahren und Entscheidungen, eine Verteilung und Verschränkung von Zuständigkeiten sowie die Begründung von (Leistungs-)Entscheidungen auf Sachkriterien von zentraler Bedeutung (Kohler 2004).

Der Einsatz dieser Instrumente geschieht nicht regellos oder gänzlich frei von Eingriffen und Steuerungen seitens der Leitungsebene. Deren Rolle besteht vielmehr darin, Rahmenbedingungen zu schaffen, die den Einsatz dieser Instrumente ermöglichen und sicherstellen. Durch die Ermöglichung von Partizipation, Sachorientierung und die Transparenz von Planungs- und Entscheidungsprozessen wird Vertrauen in die Entwicklungsplanung geschaffen (a.a.O.).

Gleichzeitig wird eine am geteilten Qualitätsverständnis orientierte Informationsgrundlage zur Qualitätsbeurteilung aufgebaut. An dieser Stelle ist die Rolle von Evaluation im beschriebenen Konzept der Qualitätsentwicklung durch den Aufbau einer Qualitätskultur zu verorten. Dieser kommt insbesondere auch die Aufgabe zu, Akteure der Hochschulplanung und -entwicklung (auf den verschiedenen Ebenen) mit relevanten Informationen zur Qualität von Lehre und Studium zu versorgen und so ihre Entscheidungen auf eine rationale Basis zu stellen. Preskill und Torres (1999) betonen, dass Evaluation die Entwicklung einer Qualitätskultur insbesondere dann unterstützt, wenn die Beteiligten in die Durchführung der Evaluation eingebunden werden und wenn diese in die laufenden Prozesse der Institution eingebettet ist. Insbesondere wird der prozessbezogene Charakter der Evaluation dadurch betont, dass diese als iterativ, sich selbst

---

<sup>6</sup>Hier ist bspw. an die Einsetzung von Qualitätszirkeln zu denken, die ihrerseits ebenfalls den Instrumenten des Qualitätsmanagements zugerechnet werden (Pasternack 2004, S. 48).

erneuernd und auf neue Fragestellungen der Institution reagierend beschrieben wird. Die Entwicklung einer Kultur der (internen) Forschung zur Qualität der Hochschulleistungen auf der Ebene der einzelnen Institutionen wird demgemäß als zentraler Beitrag zur Entwicklung einer Qualitätskultur insgesamt verstanden (Preskill und Torres 1999).

## 4.2 Evaluation und Ansätze der Hochschulentwicklung

Im oben beschriebenen Konzept der Qualitätskultur wurden drei Aspekte angesprochen, die als zentral für deren erfolgreiche Etablierung angesehen werden können. Dabei handelt es sich um die Beteiligung aller Akteure an der Planung und Durchführung von Maßnahmen zur Qualitätssicherung, um deren Kommunikation untereinander und ihr qualitätsbewusstes und -förderndes Handeln gemäß der gemeinsam formulierten Ziele sowie schließlich um eine relevante Informationsbasis, die ein rationales Handeln und Entscheiden überhaupt ermöglicht. Gemäß dem in diesem Zusammenhang genannten Verständnis von (Lehr-)Evaluation ist deren Charakter als Forschungsprogramm zur Herstellung von Wissensbeständen über die (qualitätsbezogene) Situation von Lehre und Studium, unter systematischer Anwendung wissenschaftlicher Methoden in Erinnerung zu rufen (vgl. Kapitel 3). Evaluationen kommt die Rolle zu, rationale Entscheidungen zu ermöglichen, indem sie verlässliches Wissen zu *Stärken und Schwächen* von Institutionen oder Programmen sowie zu entsprechenden Entwicklungsbedarfen generiert (vgl. z.B. Thoenig 2000). Mit dem Aspekt der Verlässlichkeit der in die Entscheidungsprozesse einbezogenen Daten ist gleichzeitig die Frage nach den Gütekriterien der (Qualitäts-)Messungen verbunden, die zum Zwecke der Evaluation durchgeführt werden.

Über die Herstellung von in diesem Sinne relevantem Wissen hinaus, sind jedoch auch die Bedingungen zu berücksichtigen, die es begünstigen, dass die gewonnenen Kenntnisse für die Formulierung und Umsetzung von Entwicklungs-

strategien auch tatsächlich eingesetzt werden.<sup>7</sup> Es ist dementsprechend die Frage zu diskutieren, welche (theoretischen) Konzepte tragfähig sind, bei dem anspruchsvollen Versuch, die genannten Aspekte einer Qualitätskultur in Routinen praktischen Handelns der Hochschulsteuerung und Qualitätssicherung zu integrieren.

Derartige theoretische Konzepte, die in der Praxis der Qualitätssicherung an Hochschulen den Rahmen für die Formulierung und Durchführung von konkreten Maßnahmen vorgeben können, werden im Folgenden vorgestellt. Dabei wurden innerhalb des umfangreichen, in der Literatur diskutierten Bestands an verschiedenen Modellen und Instrumenten der Hochschulsteuerung einige als besonders relevant erachtet. Bei der Auswahl dieser Instrumente wurde der bereits oben entwickelte Gedanke einer Prozesskette aufgegriffen: Diese unterstellt, dass durch partizipative Elemente und Verfahren der Evaluation organisationale Lernprozesse sowie Prozesse der Nutzung von (organisationalem) Wissen im Sinne eines Wissensmanagements angestoßen werden, die ihrerseits Qualitätsorientierung und Qualitätsbewusstsein fördern. Diese können in einer evidenzbasierten Hochschulsteuerung resultieren, also politischem Handeln, das sich durch eine Praxis auszeichnet, in der Entscheidungen auf gesicherten und einschlägigen Informationen und Daten aufgesetzt werden.

### 4.2.1 Partizipative Evaluation

In ihrer Beschreibung des Konzepts der Partizipativen Evaluation versuchen Ulrich und Wenzel (2003) den aus ihrer Sicht vielfach herrschenden Widerspruch der Evaluation zu überwinden. Dieser besteht zwischen der ursprünglichen Absicht von Evaluation, Lernprozesse zum Erreichen von Verbesserungen auf der Durchführungsebene von Programmen zu initiieren und ihrer tatsächlichen Nut-

---

<sup>7</sup>Als grundsätzliches Kriterium zur Beurteilung der Qualität von Evaluationen wird u.a. auch das Maß angesehen, in dem die durch sie hergestellten Erkenntnisse Eingang in Planungs- und Entscheidungsprozesse finden (z.B. Leviton 2003; Nutley u. a. 2003).

zung für Steuerungsentscheidungen insbesondere als „Ja-Nein Entscheidung“ zur Fortführung oder Abbruch eines Programms (vgl. auch Ulrich und Wenzel 2004).

Die grundlegende Annahme des Konzepts besteht darin, dass Ergebnisse von Evaluationen, die im Kontext von Bildungsprogrammen durchgeführt werden, nur schwer im Sinne einer objektiven Erfassung der Wirksamkeit dieser Programme, losgelöst von ihrer je spezifischen Programmsituation und den damit verbundenen Problemen interpretiert werden können. In Anlehnung an oben beschriebene konstruktivistisch orientierte Evaluationsparadigmen (vgl. Abschnitt 3.2) betonen Ulrich und Wenzel (2003) daher die Wichtigkeit der Kontextgebundenheit und der demokratischen und partizipativen Anlage von Evaluationsdesigns. „Evaluation ist als interaktiver Prozess zu begreifen, der durch das Einbeziehen unterschiedlicher Perspektiven blinde Flecken aufdecken kann (...)“ (a.a.O., S. 31). Demgemäß wird die Rolle von Evaluation als Kontrollinstrument zur Überprüfung des Zielerreichungsgrades hinsichtlich bestimmter Leistungsstandards sowie ihre vordringliche Funktion der Herstellung von Informationsgrundlagen abgelehnt. „Evaluation ist ein sozialer, kein messtechnischer Prozess“ Ulrich und Wenzel (2003). Dadurch, dass Evaluation „zur Ursache von neuen Lernerfahrungen (wird)“ (a.a.O.) wird ihr zugeschrieben, zur Erweiterung von Partizipation und Pluralität in Programmen<sup>8</sup> beizutragen. Die Berücksichtigung unterschiedlicher Perspektiven im Evaluationsprozess wird deshalb als zentrales Merkmal der Partizipativen Evaluation genannt, weil bedingt durch den Umstand, dass Evaluationen üblicherweise durch Ausgangsfragestellungen strukturiert werden, ganz bestimmte Beurteilungsperspektiven auf den Evaluationsgegenstand vorbestimmt und andere gleichzeitig ausgeblendet werden. Ausgangsfragestellungen werden bspw. durch das spezifische Informationsbedürfnis des Auftraggebers der Evaluation bestimmt. Aus dem Ziel, im Verlauf der Evaluation möglichst viele Per-

---

<sup>8</sup>Ulrich und Wenzel (2003) beschreiben das Konzept der Partizipativen Evaluation vor dem Hintergrund ihrer Erfahrungen in Studienprogrammen der politischen (Erwachsenen-)Bildung. Die Merkmale des Konzepts werden aber auch auf den Kontext anderer Evaluationsgegenstände übertragen.

spektiven einbeziehen zu können, leitet sich der Anspruch ab, diese demokratisch und partizipativ zu organisieren. „Durch die demokratische, d.h. gleichberechtigte Beteiligung der von der Evaluation Betroffenen, kommen verschiedene Perspektiven zur Sprache, auch solche, die bisher vielleicht unberücksichtigt blieben oder gar unangenehm sind. Partizipative Evaluation bedeutet, konstruktiv mit diesen konflikthaftern Perspektiven umzugehen“ (Ulrich und Wenzel 2003, S. 33). Besondere Bedeutung kommt in diesem Zusammenhang dem *Prozess* der Evaluation zu, in dem sich der Austausch und die Weiterentwicklung von Perspektiven und die Diskussion der Beteiligten der mit diesen Perspektiven verknüpften Konflikten vollzieht. Weniger relevant ist in diesem Konzept daher die Überprüfung von Programmwirkungen und die Beantwortung vorab gestellter Fragestellungen. Diese ergeben sich, lediglich *basierend* auf einer Ausgangsfragestellung im Verlauf der Diskussionen der involvierten Akteure (Ulrich und Wenzel 2003, S. 34). Erhobene Daten haben im Ansatz der Partizipativen Evaluation nicht die Funktion, Realitäten objektiv zu beschreiben und eine Informationsgrundlage zur Verfügung zu stellen, auf Grund derer Schlüsse über das evaluierte Programm und seine Qualität zu ziehen wären. Daten und deren Interpretation, die sich auf die Perspektiven, Meinungen und Interpretationen der Akteure beziehen, werden vielmehr in den Diskussions- bzw. Aushandlungsprozess zurückgemeldet und geben Anlass, für weiterführende Fragestellungen und sich anschließende Diskussionsrunden.<sup>9</sup> Dem Evaluator kommt in diesem Konzept nicht die Aufgabe zu, objektive Daten zu erheben, sondern Diskussionen zu moderieren, die Aushandlungsprozesse

---

<sup>9</sup>Art und Umfang der benötigten Daten werden nicht näher spezifiziert. Sie können nach Ulrich und Wenzel (2003) sowohl aus quantitativen, wie auch aus qualitativen Daten bestehen. Aus verschiedenen Praxisberichten partizipativ orientierter Evaluationsprojekte geht sowohl der Einsatz standardisierter als auch qualitativ orientierter Datenerhebungsverfahren (Brommes u. a. 2004, S. 99), ebenso wie die Verwendung von inhaltsanalytischen Methoden und Experteninterviews (Hirsland u. a. 2004, S. 61) hervor. Zentral ist, dass die Datenerhebung selber nicht als „wissenschaftlich unabhängiger Prozess“ Ulrich und Wenzel (2003) zu verstehen ist, sondern als „das Einholen unterschiedlicher Perspektiven in Bezug auf das Evaluationsobjekt“ (a.a.O., S. 35).

der Beteiligten zu begleiten, Konflikte zu erkennen und Lösungen anzubieten. Das Ergebnis einer Evaluation in diesem Verständnis wird nicht in einem Bericht über die objektive Situation eines Programms und der Entwicklung dementsprechend gebotener Maßnahmen gesehen. Die besondere Stärke des Konzepts und ihr Vorzug gegenüber einer „herkömmlichen“, objektivierenden Evaluation besteht darin, dass Handlungsempfehlungen das Ergebnis der von allen Beteiligten geführten Diskussionen sind, dementsprechend für die empfohlenen Maßnahmen mit einer hohen Akzeptanz und für etwaige Reformprozesse mit einer hohen Teilnahmebereitschaft gerechnet werden kann. Die Praxisrelevanz der Partizipativen Evaluation beweist sich in den Konsequenzen des Evaluationsberichts (Ulrich und Wenzel 2003, S.38; Ulrich und Wenzel 2004, S. 39). Gemäß der Betonung einer durch Diskussionen hergestellten und konsensual von allen Beteiligten getragenen Bewertung des evaluierten Programms und des entsprechenden Handlungs- und Entwicklungsbedarfs, stehen im Konzept der Partizipativen Evaluation Kriterien für die Beurteilung der Evaluation selber weniger im Mittelpunkt. Bei diesen Kriterien, die im folgenden Kapitel eingehend diskutiert werden, handelt es sich um Gütekriterien wissenschaftlicher Messungen, wie die Reliabilität und Validität von Daten(erhebungen). Diese werden im Konzept der partizipativen Evaluation nicht nur nicht in den Vordergrund, sondern vielmehr hinsichtlich ihrer Bedeutung zur Disposition gestellt (Ulrich und Wenzel 2004, S. 41).<sup>10</sup> Der veränderte Qualitätsbegriff zur Durchführung von Evaluation, der im Ansatz der Partizipativen Evaluation entworfen wird und sich u.a. auf den Prozess der Herstellung von

---

<sup>10</sup>Das Konzept der partizipativen Evaluation lehnt sich an den Ansatz der Appreciative Inquiry an (Ulrich und Wenzel 2003, S. 35). Diese bezeichnet ein Forschungsdesign aus dem Bereich der Aktionsforschung (zum Programm der Action Research vgl. z.B. Cooperrider und Srivastva 1987; Bushe 1995; Bushe und Kassam 2005). Gegenstand ist die Organisationsentwicklung auf Grund diskursiver Verfahren, bei denen die grundlegende Annahme darin besteht, dass es keine stabile soziale Ordnung gibt, die mittels objektiver Verfahren analysierbar wäre, sondern dass die soziale Realität von Organisationen Gegenstand stetiger Veränderung und Rekonstruktion durch die Organisationsmitglieder sei. Kategorien wie die Validität von Situationsbeschreibungen haben daher für entsprechende Untersuchungen keine Relevanz (Bushe 1995).



beurteilungsrelevanten Informationen bezieht, impliziert vielmehr neue Standards zur Beurteilung der Güte von Evaluationen. Diese bestehen in

- der *Fairness* gegenüber verschiedenen Perspektiven,
- der *Weiterentwicklung* der eigenen Perspektiven bei allen Beteiligten,
- dem Beitrag zur *Fortbildung* im Sinne einer Sicherung des Verständnisses für Bewertungen und Perspektiven anderer Beteiligter,
- der *Relevanz* der Ergebnisse für Handlungskonsequenzen,
- der *Ermächtigung* der Beteiligten (im Sinne einer Durchführbarkeit der als nötig erkannten Handlungskonsequenzen) und
- der *Glaubwürdigkeit* als Maß für das umfassende, partizipative und von den Beteiligten akzeptierte Vorgehen der Evaluation (Ulrich und Wenzel 2003, S. 42).

Das Konzept der partizipativen Evaluation, welches vielfach zur Evaluation von Programmen der politischen Bildung eingesetzt wird (vgl. z.B. Uhl u. a. 2004; Ulrich und Wenzel 2004), liefert aus vorliegender Sicht relevante Impulse auch für die Konzeption der Evaluation von Lehre und Studium. Diese bestehen insbesondere in der Betonung einer prozessorientierten Begutachtung des Einwirkens von Bildungsangeboten auf die dem entsprechenden Programm ausgesetzten Personen. Der Erfolg eines Bildungsprogramms, bspw. im Sinne einer bestimmten Qualifizierung der jeweiligen Teilnehmer, ist wie oben ausgeführt wurde, nur unter Beachtung methodischer Restriktionen als kausale Relation von Ursache und Wirkung (des Programms) zu analysieren. Unkontrollierbare Einflussgrößen können auf den Bildungserfolg oder -misserfolg wirken, die nicht ursächlich in den Programminhalten oder der Programmdurchführung begründet sind (vgl. Kapitel 3). Die Beobachtung von Prozessen mittels der Orientierung von Evaluationen auf die Analyse von (bspw. studienbiographischen) Entwicklungen im zeitlichen Längsschnitt erlaubt es, sich der Frage der Zuordenbarkeit von Programmwirkungen oder -ergebnissen zu seinen Eigenschaften bzw. zumindest den

dem Programm *zugeschriebenen* Eigenschaften zu nähern. In dem in Kapitel 7 beschriebenen Evaluationsdesign zur Beurteilung der Qualität von Lehre und Studium durch Studierende ist dementsprechend der Gedanke der Betrachtung von Studienverläufen im zeitlichen Längsschnitt, mit dem Ziel, Studienbiographien und mithin den Prozess von Lehre und Studium zu analysieren, verankert. Eine Prozessorientierung der Datenerhebung ist in diesem Design entsprechend realisiert.

Gleichzeitig wird im Ansatz der Partizipativen Evaluation die Frage nach der „Evaluierbarkeit“ von (Bildungs-)Programmen im Sinne der Überprüfung ihrer positiven oder negativen, intendierten oder nichtintendierten Wirkung auf die an dem Programm teilnehmenden Personen erneut aufgeworfen. Die bereits angeführten methodologischen Probleme bei dem Versuch Evaluationen in dieser Weise einzusetzen, werden um inhaltliche Argumente hinsichtlich des Zweifels an der Sinnhaftigkeit von Evaluation als Kontrollinstrument und Legitimationsgrundlage für Entscheidungen über die Einstellung oder Fortführung von Bildungsprogrammen erweitert.

Nichtsdestotrotz scheint auch Kritik angebracht zu sein, bezüglich der praktischen Anwendbarkeit des Ansatzes in Systemen der Qualitätssicherung von Lehre und Studium an Hochschulen. So setzt die Methode der Partizipativen Evaluation voraus, dass es überhaupt die Gelegenheit für die Akteure gibt, über die Qualität von Lehre und Studium aus verschiedenen und in der entsprechenden Diskussion gleichberechtigten Perspektiven zu diskutieren und zu streiten.

Darüber hinaus ist vorgesehen, dass Maßnahmen zur Qualitätsverbesserung aufbauend auf konsensualen Qualitätsbewertungen der Teilnehmenden geplant und realisiert werden. Insbesondere ist es dabei vermutlich schwer möglich, ein System der Repräsentanz aller relevanten Akteursgruppen eines großen Studienfachs oder Studiengangs zu schaffen, in dem die individuellen Meinungen zur Qualität hinlänglich demokratisch berücksichtigt werden und nicht dem Kräftespiel um Einflussnahme durch unterschiedlich starke Gruppen und Gruppeninteressen ausgesetzt sind. Der Annahme, dass die Partizipation als notwendige und hinreichende Voraussetzung für das von einem breiten Konsens getragene Durchsetzen

von Qualitätsverbesserungsmaßnahmen ausreicht, steht die Erfahrung entgegen, dass es durchaus auch andere externe und interne Bedingungen gibt, die den entsprechenden Bemühungen zuwiderlaufen. So sind den objektiven Möglichkeiten, am Alltag von Lehre und Studium überhaupt etwas verändern zu können, oft enge (bspw. fiskalische) Grenzen gesetzt. Die meisten als notwendig erachteten Maßnahmen zur Qualitätsverbesserung erfordern auch den Einsatz oder die Verschiebung von Ressourcen, die ihrerseits knapp sind und derentwegen - sofern es um ihre Verlagerung zu Lasten des einen Bereichs und zu Gunsten eines anderen geht - der mühsam erarbeitete Konsens möglicherweise schnell wieder aufgebrochen wird. Weiterhin steht eine Ungleichverteilung der Einflussmöglichkeiten verschiedener Akteursgruppen sowie das Vorhandensein von Beharrungskräften im Sinne der Wahrung von Vorteilen, einer konsensual von allen Beteiligten getragenen Strategieentwicklung vermutlich entgegen.

Weiterhin scheint die Bedeutung, die der Qualität erhobener Daten zugemessen wird, überaus diskutabel zu sein. Sofern eine Datenerhebung den spezifischen Zweck verfolgt, die auf Qualitätsverbesserungen zielenden Diskussionen mit Informationen zu versorgen, scheint es aus vorliegender Sicht notwendig zu sein, dass die Zustände, die mittels der erhobenen Daten beschrieben werden sollen, durch die Daten auch zuverlässig abgebildet werden. Anderenfalls besteht zumindest die Gefahr der Fehlinformation der Diskussionsteilnehmer und mithin eine Fehlinterpretation und Fehlbeurteilung der einzuschätzenden Situation.

Offensichtlich lässt sich das Konzept der Partizipativen Evaluation eher in kleineren Kontexten verwirklichen, an denen kleinere Gruppen von Akteuren beteiligt sind, als es üblicherweise für Studienfächer von Hochschulen zutrifft. Eines der wichtigen konstituierenden Merkmale der Partizipativen Evaluation, bestehend in der Kommunikation der Beteiligten, wird im Bereich der Lehrevaluation an Hochschulen (wenn überhaupt) eher auf der Ebene der Bewertung einzelner Lehrveranstaltungen verwirklicht.

In einem anderen Merkmal der Partizipativen Evaluation, nämlich der Beteiligung der Akteure an der Planung und Durchführung von Maßnahmen zur Qualitätssicherung besteht dagegen eine häufige Forderung auch für die Evalua-

tion von Lehre und Studium auf höheren Aggregatebenen, wie Studienfächern und Fachbereichen. Beteiligung ist in verschiedenen Formen zu realisieren und unabhängig von den genannten Problemen, die einer Institutionalisierung der Partizipativen Evaluation als Methode der Qualitätssicherung von Lehre und Studium möglicherweise entgegenstehen. So stellt bspw. für König (2004) die konzeptionelle Verankerung von Selbstorganisation und Selbststeuerung in Qualitätssicherungssysteme eine wichtige Voraussetzung für die Initiierung und praktische Umsetzung von Lernprozessen in Hochschulen dar: Individuelles Lernen wird dann besonders effektiv, „wenn die Akteure vor Ort bei der Planung und Vorbereitung und im Prozess selbst maßgeblich beteiligt sind. Die Verantwortlichkeiten für Qualitätsentwicklung werden auf diese Weise gestärkt“ (König 2004, S. 8). Lernprozesse, die als Schlüssel zum Erfolg von Qualitätsentwicklungsmaßnahmen angesehen werden, sind ihrerseits das Ergebnis von Beteiligung. „Offenheit, Klarheit und Transparenz im Verlauf ermöglichen echte Beteiligung. Daraus erst entstehen nachhaltige Lernprozesse, die wiederum die wichtigsten Voraussetzungen für das Gelingen von Qualitätsentwicklung sind (a.a.O., S. 9).

### 4.2.2 Evaluation und Lernende Organisationen

Im Konzept des Organisationslernens wird der individuelle Lernbegriff von der Anpassung an sich ändernde Umweltbedingungen, auf das Lernen von Organisationen übertragen (Pasternack 2004). „Lernend ist eine Organisation dann, wenn sie durch Wissensaufnahme, Wissensproduktion und Verfügbarkeitssicherung des neuen Wissens ihre Abläufe und die Handlungsorientierungen ihrer Mitglieder ständig an neue Gegebenheiten und Erfordernisse anpasst“ (a.a.O., S. 60). Gleichzeitig besteht das Lernen von Organisationen aus mehr als dem Lernen der einzelnen Organisationsmitglieder. „Organisationales oder institutionelles Wissen steckt in den personenunabhängigen, anonymisierten Regelsystemen, welche die Operationsweise eines Sozialsystems definieren“ (Willke 1998, S. 166, zit. nach Pasternack 2004). Organisationales Lernen ist demnach als „Prozess der Durchsetzung neuen Wissens innerhalb der Organisation definiert, das in neuen Routinen und Strukturen - also unabhängig von Individuen gespeichert wird“ (Wilkes-

mann 2001, S. 333). Preskill und Torres (1999) beschreiben, dass Organisationen heute sich stetig wandelnden Anforderungen ausgesetzt sind. Dieser Wandel hat zur Folge, dass Stabilität und Sicherheit, im Sinne von wiederkehrenden Routinen verloren gehen. Die Überlebenschancen von Organisationen steigen in dem Maße, in dem sie in der Lage sind, auf neue Anforderungen mit der Entwicklung neuer Fähigkeiten und Eigenschaften zu reagieren (a.a.O.). Isolierte, individuell verantwortete Arbeitsabläufe werden zunehmend zugunsten multi-disziplinärer Aufgabenstellungen verdrängt. Entsprechend prognostizieren Preskill und Torres (1999), „that tomorrow’s organizations will (1) accomplish their work through multi-disciplinary teams, (2) have permeable boundaries, (3) be focused on mental tasks, (4) be participative, diverse, and innovative, (5) support a professional culture of commitment and results, and (6) value peer-to-peer relationships. (...) Such organizations will develop a culture of continuous learning that influences the way the organization approaches its goals (...)“ (Preskill und Torres 1999, S. 43). Auch Carstensen (2004) betont, dass Organisationen durch stetigen Wandel gekennzeichnet sind und dass sie „aus vergangenen Erfolgen und Misserfolgen Lehren ziehen, die Irrtümer der Vergangenheit aufspüren und korrigieren, bevorstehende Bedrohungen erahnen und darauf reagieren, experimentieren, ständig innovativ sein und Bilder einer erstrebenswerten Zukunft aufzeigen und realisieren müssen“ (Argyris und Schön 1999, zit. nach Carstensen 2004, S. 50). Die Aktivierung der Organisationsmitglieder und ihrer Lernkompetenz ist dabei eine wichtige Voraussetzung für erfolgreiches organisationales Handeln und Lernen (Mittag 2006, S. 23 ff.).

In engem Zusammenhang mit dieser auf Dialogfähigkeit und Partizipation ausgerichteten Konzeption organisationalen Lernens<sup>11</sup> steht der Ansatz der „evaluative inquiry“ (Preskill und Torres 1999). In der Durchführung entsprechender Evaluationsprozesse, die hinsichtlich ihres Ansatzes und ihrer Durchführung

---

<sup>11</sup>Die Eigenschaften der Dialogfähigkeit und der Partizipation werden in der Literatur zu Anwendungen des Organisationslernens vielfach als relevante Kennzeichen bzw. Voraussetzungen für entsprechende Lernprozesse angeführt (vgl. z.B. Silins u. a. 2002).

mit den Prämissen der Partizipativen Evaluation vergleichbar sind (Betonung der Durchführung des Evaluationsprozesses durch die Organisationsmitglieder, Bevorzugung von Verfahren, die auf Dialog und Selbstreflexion orientiert sind), wird eine relevante Voraussetzung für das Organisationslernen überhaupt gesehen (Preskill und Torres 1999). Dieses wird durch eine stetig erneuerte Informationsbasis ermöglicht, die aus den iterativ angewendeten Evaluationsverfahren hergestellt wird und zur Beurteilung der Organisation und ihrer Arbeitsabläufe dient. Als Voraussetzung für erfolgreiches Organisationslernen werden neben der Dialogfähigkeit der Beteiligten auch der Austausch individuellen Wissens und dessen Transformation in organisationales Wissen als das Ergebnis sozialer Konstruktionen angesehen (Preskill und Torres 1999).<sup>12</sup> Kennzeichnend für den Ansatz der „evaluative inquiry“ als Instrument des Organisationslernens ist, dass er über die Bewertung der Ergebnisse des Organisationshandelns (in Anspielung auf die vorherrschende Outputorientierung traditioneller Evaluationsansätze) hinausgeht. Vielmehr wird die Debatte und Infragestellung der Werte, die dem Organisationshandeln zugrunde liegen in die entsprechende Untersuchung einbezogen.<sup>13</sup> Diese „evaluative Debatte“ wird nicht nur im Angesicht konkreter Probleme geführt, sondern im Sinne einer steten „Wertediskussion“ kontinuierlich und iterativ aufrecht erhalten. Preskill und Torres (1999) betonen, dass dies die Entwicklung einer Forschungskultur der Organisation zur Folge hat, in der die Organisationsmitglieder stetig darum bemüht sind, einschlägige qualitätsbezogene Informationen zu sammeln und in die Bemühungen zur Qualitätsentwicklung einzubringen. Teil

---

<sup>12</sup>Diese Austausch- und Transformationsprozesse finden auf der Ebene kleiner, an entsprechenden Evaluationen beteiligten Teams statt. Dabei wird davon ausgegangen, dass das Lernen von Teams -oder Gruppen von Organisationsmitgliedern- das Lernen auf der Ebene der Gesamtorganisation befördert.

<sup>13</sup>Zur Bedeutung der Werte einer Organisation und ihrer Fähigkeit, diese Werte vor dem Hintergrund neuer Situationen und Rahmenbedingungen weiterzuentwickeln bzw. völlig neu zu definieren, vgl. insbesondere das Konzept des organisationalen single-loop und double-loop Lernens von Argyris und Schön (Argyris 1992; Argyris und Schön 1999; vgl. auch Carstensen 2004).

dieser Forschungskultur sind Evaluationsverfahren, die ein relevantes Instrument zur Unterstützung von organisationalen Lernprozessen sein können (Preskill 2004, zit. nach Mittag 2006, S. 23). Hierfür bedarf es verschiedener Voraussetzungen, wie beispielsweise, dass Evaluation als permanenter, reflexiver Prozess verstanden wird, der in die organisationale Praxis eingebettet ist, dass Lernen als wichtiges Ziel der Evaluation anerkannt wird und dass sich die Durchführung von Evaluationen auf die Organisation (ihre Organisationskultur, Rahmenbedingungen, u.ä.) bezieht (Mittag 2006, S. 23).

Zu ergänzen sind diese Voraussetzungen aus vorliegender Sicht um den Aspekt der Datenqualität im Sinne der Aussagekraft erhobener Informationen, die die Grundlage für die organisationalen (und individuellen) Lernprozesse bilden sollen. Hier ist insbesondere nach der Validität und Reliabilität gewonnener Daten zu fragen. Bezogen auf den Kontext von Lehre und Studium scheint das Konzept des Organisationslernens insbesondere für die Ebene von Studiengängen vielversprechend zu sein, weil diese in der gemeinsamen Verantwortung der Lehrenden und Studierenden eines Fachbereichs sowie der nichtwissenschaftlichen Beteiligten (bspw. aus Verwaltungsbereichen, die die Prüfungsorganisation verantworten) getragen werden. Dementsprechend können sie Gegenstand der Qualitätsentwicklung sein, die verschiedene Informationsquellen, aber auch verschiedene diskursive Verfahren der Leistungsbeurteilung einbezieht.

### 4.2.3 Organisationslernen in der Praxis der Hochschulsteuerung

Lüthje (2003) beschreibt anhand eines an der Universität Hamburg durchgeführten Projekts zur Organisationsentwicklung anschaulich, wie organisationale Lernprozesse in der Praxis von Hochschulen und ihren Fachbereichen angeschoben und moderiert werden können und welche Voraussetzungen dafür zu schaffen sind. Darüber hinaus wird in seiner Beschreibung des Projektes „ProUni“ (Lüthje 2003) deutlich, welche Bedeutung Evaluationsergebnissen in diesem Zusammenhang zukommt. Unter dem „Ansatz der systemischen Universitätsentwicklung“

(vgl. Lüthje 2003, S. 46) wurde im Rahmen des genannten Projekts ein Reformprozess in Gang gesetzt, dessen Ziel in der „Analyse und Lösung von Problemen der Leitungs-, Verwaltungs- und Organisationsstruktur“ (a.a.O., S. 48) bestand. Vier Teilprojekte zur Erreichung dieses Ziels wurden in einer ersten Projektphase angeschoben:

1. Formulierung von Entwicklungszielen und eines Leitbildes der Universität auf Grund einer Stärken-/Schwächenanalyse.
2. Stärkung der Eigeninitiative und Eigenverantwortung der Fachbereiche (bspw. durch Delegation der Stellen- und Mittelbewirtschaftung von der Zentralebene auf diejenige der Fachbereiche).
3. Reorganisation der Zentralverwaltung (strategische Neuorientierung in Richtung einer Steuerung durch Zielvereinbarungen an Stelle von Vorgaben).
4. Entwicklung eines universitären Berichts- und Controllingsystems (Lüthje 2003, S. 49).

Darüber hinaus wird für das beschriebene Projekt der systemischen Universitätsentwicklung als bedeutsam herausgestellt, dass es an bereits vor Projektbeginn entwickelte Reformbausteine anknüpfen konnte. Als einer dieser Bausteine des Reformprozesses wird insbesondere auch die Qualitätssicherung in Studium und Lehre im Rahmen des Evaluationsverbundes Norddeutscher Universitäten genannt (Lüthje 2003, S. 46).<sup>14</sup> Das übergeordnete Ziel des Reformprozesses, welches darin besteht, „die Universität zu einer lernenden Organisation zu machen“ (a.a.O.), soll dementsprechend auch durch die Konsultation von Evaluationsergebnissen zur Qualität von Lehre und Studium erreicht werden.

Kennzeichnend für die Projektdurchführung ist die Einbeziehung der Entscheidungs- und Arbeitsstrukturen auf zentraler wie dezentraler Ebene.

---

<sup>14</sup>Zur Organisation und zu Zielen und Strukturen des „Nordverbundes“ vgl. auch Fischer-Bluhm (1995); Fischer-Bluhm (1998); Fischer-Bluhm u. a. (2003); Lüthje (1997).



Diese wird als Voraussetzung für die Berücksichtigung der „Informations-, Kommunikations-, Kooperations- und Entscheidungserfordernisse“ (Lüthje 2003, S. 51) beschrieben, durch sie wird eine breite Beteiligung der universitären Akteure verschiedener Ebenen ermöglicht und erreicht. Die Veränderung der Leitungs- und Verwaltungsstruktur impliziert einen Umstellungsprozess, der nicht durch eine einmalige Leitungsentscheidung herbeigeführt werden kann, sondern auf der Beteiligung der Organe, Gremien und Mitglieder der Universität beruht.

Für die methodische Umsetzung wurde ein systemisches Organisationsentwicklungskonzept entworfen. Dieses geht davon aus, dass „die erforderliche Optimierung des komplexen Gesamtsystems der Universität nicht allein durch Expertenanalyse und Leitungsentscheidungen bewirkt werden kann. Vielmehr ist ein Prozess der institutionellen Selbstorganisation unter Aktivierung möglichst aller Entscheidungsebenen und Handlungsbeteiligten und unter Einbeziehung aller den Leistungsprozess bestimmenden Faktoren einzuleiten sowie ziel- und ergebnisorientiert zu organisieren. (...) Leitidee der prozessorientierten Organisationsentwicklung ist nicht die Erarbeitung und Entscheidung umfassender Gesamtmodelle, sondern die konsensorientierte Erarbeitung möglichst breit getragener, jeweils begrenzter Veränderungsschritte im Rahmen der eigenen Handlungsmöglichkeiten. Diese Teilschritte einer vorrangig als (Selbst-)Lernprozesse organisierten Entwicklung sollen - unter Aktivierung des Sachverstandes und der Motivation der am Leistungsprozess Beteiligten - koordiniert und aufeinander abgestimmt werden“ (Lüthje 2003, S. 46).

Als drei zentrale Bestandteile der Qualitätsentwicklung in Hochschulen werden in diesem längeren Zitat (i) Expertenanalyse, (ii) Leitungsentscheidungen und (iii) institutionelle Selbstorganisation, die durch Erfahrungswissen (organisationale Selbstlernprozesse) moderiert wird, genannt. Unter Expertenanalysen wird dabei (externer sowie interner) Sachverstand zur Beurteilung der Qualität von Hochschulleistungen, kurz: Evaluation verstanden. Leitungsentscheidungen stehen für „top-down“ durchgesetzte Reformvorhaben, während die institutionelle Selbstorganisation durch Lernprozesse und „bottom-up“ angestoßene Veränderungen charakterisiert wird. Diese drei Komponenten der Hochschulsteuerung werden nicht

als notwendigerweise widersprüchlich zueinander angesehen, vielmehr beziehen sie sich aufeinander und ergänzen sich (Lüthje 2003, S. 64): Evaluation (bzw. Expertenanalyse) bildet die Informationsgrundlage für Leitungsentscheidungen und für die Nutzung, Anwendung und Erweiterung von Wissensbeständen in (organisationalen und individuellen) Lernprozessen auf verschiedenen Ebenen. Gleichzeitig ist ihre Gewinn bringende Nutzung für Veränderungsprozesse nur unter der Voraussetzung möglich, dass sie vor dem Hintergrund des spezifischen Gültigkeitsbereichs des jeweiligen institutionellen Kontextes interpretiert und diskutiert werden.<sup>15</sup> Leitungsentscheidungen sind auch in beteiligungsorientierten Reformprozessen unverzichtbar, wenn bspw. Veränderungswille bei den Beteiligten fehlt (a.a.O., S. 65). Gleichzeitig können Leitungsentscheidungen gerade das Ergebnis von Selbstreflexion oder Lernprozessen sein, bspw. darüber, wie das auf Grund von Evaluationsergebnissen formulierte Leitbild der Hochschule erreicht werden soll.

Bezogen auf die Frage nach der organisationalen Lernfähigkeit von Hochschulen wird das beschriebene Projekt positiv resümiert (vgl. Lüthje 2003, S. 66). Die Beteiligten haben „wichtige Prozess Erfahrungen gemacht, Methoden gelernt und neue Verhaltensweisen entwickelt“ (a.a.O.), die zu einer Institutionalisierung von (Selbst-)Reflexion beigetragen haben, die ihrerseits wichtiges Kennzeichen von (organisationalen) Lernprozessen ist. Den Absichten und Zielen der beteiligungsorientierten Organisationsentwicklung entsprechend, bestehen diese Verhaltensweisen in der Wahrnehmung von Möglichkeiten der Partizipation in Entscheidungsprozessen, im Formulieren und Einbringen eigener Interessen sowie in der Diskussion und Kommunikation von Gestaltungsmöglichkeiten und -rahmen, Widersprüchen und der Auseinandersetzung mit Argumenten.

---

<sup>15</sup>An dieser Stelle ist an die Ausführungen aus Kapitel 3 bezüglich verschiedener konfligierender Evaluationsparadigmen zu erinnern. Als Weg zu deren Aussöhnung oder Verbindung wurde vorgeschlagen, durchaus (objektivierende) Daten zu sammeln, diese aber vor dem Hintergrund eines jeweils spezifischen Kontextes zu interpretieren (vgl. Beywl 1988). Diesem Anspruch wird durch den hier beschriebenen Ansatz aus vorliegender Sicht durchaus gerecht.

Für den Kontext der vorliegenden Arbeit ist dieses Beispiel für ein Projekt zur Induktion organisationalen Lernprozesse aus einem ganz bestimmten Grund relevant: Teilprojekt 1 zur Erstellung eines Leitbildes und einer Profilierung der Universität knüpft unmittelbar an den vor Projektbeginn initiierten Reformbaustein Qualitätssicherung von Lehre und Studium durch Evaluation im Rahmen des Evaluationsverbundes norddeutscher Universitäten an. Basis für die Formulierung künftiger Ziele und Profile sind die Ergebnisse von Stärken- und Schwächenanalysen der Universität. Da eine Profilierung tendenziell mit einer Beschränkung auf bestimmte, als besonders leistungsstark beurteilte Bereiche einher geht, ist insbesondere die Frage danach zu stellen, ob die Daten, auf die die Ergebnisse der Stärken- und Schwächenanalysen zurückzuführen sind, die Realität bzw. die tatsächliche Qualität der beurteilten Bereiche in adäquater Weise widerspiegeln. Vitale Zukunftsentscheidungen über das Fortbestehen von Bereichen und über die Entscheidung hinsichtlich ihrer Eignung, Teil des Leitbildes und Universitätsprofils zu sein bzw. zu bleiben, sind immer auch dahingehend zu hinterfragen, ob die für die Entscheidung konsultierte Datenbasis ein sachbezogenes und valides Bild der Realität zeichnet.

An organisationale Lernprozesse wird die Erwartung gebunden, dass die Beteiligten einschlägige qualitätsbezogene Informationen über die zu verändernde Institution einbeziehen und kommunizieren und dass sie sich in ihrer Selbstreflexion von Befunden über deren Zustand und einem formulierten Entwicklungsziel leiten lassen (Preskill und Torres 1999). Es wird mithin erwartet, dass bei den Beteiligten eine „datengestützte Informiertheit“ bezüglich der zu beurteilenden Institution und etwaiger Verbesserungspotenziale vorliegt. Das Ergebnis des Lernprozesses soll in einer relevanten organisationalen Weiterentwicklung bestehen, die durch diese informierte Selbstreflexion angeschoben wird. Auf Grund dieses Merkmals der Basierung der Lernprozesse auf Informationen (bzw. zuvor oder währenddessen erhobene Daten) ist für ihr nachhaltiges Gelingen ebenfalls die Bedingung zu formulieren, dass die einbezogene Datenbasis den genannten Gütekriterien der Validität und Reliabilität entspricht. Andernfalls wäre von einer fehlenden Informiertheit und einem Lernen, das auf falschen Annahmen be-

ruht zu sprechen. Es wäre in diesem Fall davon auszugehen, dass im Rahmen der Lern- und Reflexionsprozesse potenziell die falschen Schlüsse gezogen würden, die ihrerseits zu falschen oder nicht sachgerechten Entscheidungen zu führen drohen.

#### 4.2.4 Evaluation als Informationsbasis für Wissensmanagement

Wissensmanagement wird von Pasternack (2004) als Instrument für eine erfolgreiche Anwendung des Organisationslernens angesehen: „Als Voraussetzung des organisationalen Lernens wird ein von den einzelnen unabhängiges Gedächtnis benötigt - nicht zuletzt, um relevantes Wissen einzelner Organisationsmitglieder auch dann verfügbar zu halten, wenn diese die Organisation verlassen. Insofern baut die Lernende Organisation auf einem erfolgreichen Wissensmanagement auf“ (Pasternack 2004, S. 60). Diese Funktion bezieht sich insbesondere auf die technischen Aspekte von Wissensmanagement, wie die Datenspeicherung und Verfügbarkeitssicherung von Informationen für Entscheidungsprozesse. „Es geht beim Wissensmanagement um den Kreislauf von Informationen, durch den das Wissen von Einzelpersonen für ganze Organisationen verfügbar wird“ (a.a.O., S. 60). Als konstituierend für Wissensmanagement werden vier zentrale Bestandteile gekennzeichnet. Bei diesen handelt es sich um *Wissensrepräsentation* (Dokumentation vorhandener Wissensbestände), *Wissenskommunikation* (beschreibt die Kooperation von Organisationsmitgliedern auf der Basis geteilten Wissens), *Wissensgenerierung* (verstanden als Beschaffung externen Wissens und die Schaffung von Wissensnetzwerken, bestehend aus den vernetzten Wissensbeständen individueller Organisationsmitglieder) sowie *Wissensnutzung* (als Umsetzung vorhandenen Wissens in Entscheidungen) (Pasternack 2004, S. 60 f.). Im Zusammenspiel dieser vier Elemente von Wissensmanagement ist der idealtypische Ablauf von Evaluation an verschiedenen Stellen zu verorten. Dieser besteht zunächst in der Schaffung einer relevanten Informationsbasis für die Beurteilung von Situationen und die Abschätzung von Handlungsbedarf (Generierung und Dokumentation von Informationen bzw. Wissen). Wissen, das aus entsprechen-

den „Stärken-Schwächen-Analysen“ abgeleitet wird, wird unter Einbezug seiner Interpretation durch die Verantwortlichen (Kommunikation) in konkrete Handlungsstrategien umgesetzt (Nutzung).

Die inhaltliche Verbindung von Wissensmanagement und organisationalem Lernen wird auch von Nutley u. a. (2004) beschrieben. So definieren sie dieses in Anlehnung an Scarborough u. a. (1999) als „any process or practice of creating, acquiring, capturing, sharing, and using knowledge, wherever it resides, to enhance learning and performance in organisations“ (Nutley u. a. 2004, S. 5). Allerdings wird auf einige wichtige Unterschiede der beiden Konzepte verwiesen. Diese bestehen darin, dass Organisationslernen im Gegensatz zum Wissensmanagement einen eher theoriegeleiteten und breit angelegten, auf ganze Organisationen bezogenen Fokus verfolgt, während das Wissensmanagement sich als praxisorientiert und mit einem spezifischen Fokus auf bestimmte Projekte oder Abläufe innerhalb von Organisationen bezogen versteht. Darüber hinaus betont das Organisationslernen die Reflektion der zugrunde liegenden Werte und Kulturen einer Organisation sowie das Management und die Nutzbarmachung verborgenen, individuellen Wissens für Organisationszwecke.<sup>16</sup> Das Wissensmanagement hat dagegen eher die instrumentelle Verwendung von Wissen (unabhängig von seiner Verortung innerhalb der Organisation) für konkrete projektgebundene Bedarfe zum Gegenstand. Eine praktische Konsequenz aus diesen Unterschiedlichkeiten ergibt sich daraus, dass das Ziel des Organisationslernens eher darin besteht, in die Entwicklung von Organisationsmitgliedern (hinsichtlich ihrer Kooperationsfähigkeit, etc.) und die Entwicklung der entsprechenden Organisationskultur zu investieren, während Wissensmanagement daran gelegen ist, organisationale Routinen zu verbessern, die eine bestmögliche Nutzung von vorhandenem Wissen sicherstellen. Dieses Ziel wird hauptsächlich durch (technische) Systementwicklungen, bspw. von Datenbanken und das Training von Organisationsmitgliedern in deren effektiver Nutzung erreicht (Nutley u. a. 2004).

---

<sup>16</sup>Zur Bedeutung verborgenen Wissens (tacit knowledge) für organisationale Lernprozesse vgl. z.B. auch Carstensen (2004).

Bezogen auf die Evaluation von Hochschulen wird durch das Konzept des Wissensmanagements eher die Dimension ihrer praktischen Durchführung angesprochen. Diese bezieht sich auf die Erhebung und Bereitstellung einer einschlägigen Datenbasis zum Zwecke der Erstellung umfassender Informationen als Beurteilungsgrundlage der Qualität der Leistungserbringung.<sup>17</sup> Im Gegensatz zu den genannten „technischen Aspekten“ steht die Organisationsentwicklung im Sinne der Selbstreflexion von Organisationsmitgliedern über die bereitgestellten Evaluationsergebnisse weniger im Fokus des Konzepts. Diese ist eher der Gegenstand des zuvor beschriebenen organisationalen Lernens.

In Anlehnung an Pasternack (2004) werden deshalb die beiden Konzepte des Organisationslernens und des Wissensmanagements als einander bedingend aufgefasst, in dem Sinne, dass nur durch eine solide Daten- und Informationsbasis Wissen über die Organisation für Lernprozesse eingesetzt werden kann. Organisationale Lernprozesse haben jedoch nicht den unmittelbaren oder instrumentellen Zweck, bspw. die Kosten-Nutzen Bilanz der Organisation durch modifizierte Handlungsroutinen zu verbessern, sondern zielen auf die Infragestellung und Neudefinition von Handlungsweisen, die der Leistungserbringung zugrunde liegen.<sup>18</sup> Sie sind entsprechend längerfristiger Natur und zielen auf strategische Organisationsentwicklungen. Im Zusammenhang mit den im vorliegenden Kapitel dargestellten Konzepten der Qualitätssicherung von Lehre und Studium ist das Wissensmanagement insbesondere mit Blick auf die technischen Aspekte der Etablierung und Durchführung von Evaluationsverfahren relevant. Nötig für den Aufbau einer sachgerechten Beurteilungsgrundlage zur Qualität von Lehre und Studium ist in einem leistungsfähigen System zum Management großer Datenmengen zu sehen. Diese entstehen bei studentischen Befragungen, aber auch im

---

<sup>17</sup>Feinstein (2002) bezeichnet Wissensmanagement als die Konversion von Information in Wissen, womit er auf einen Präsentationsmodus von Evaluationsergebnissen anspielt, der für ihre potenziellen Nutzer (Politiker, Aufsichtsräte, Hochschulleitungen) verständlich und dementsprechend überhaupt nutzbar ist.

<sup>18</sup>Vgl. auch Argyris (1992); Argyris und Schön (1999).

Rahmen der amtlichen Hochschulstatistik. Schon in der Frage der Datenaufbereitung bzw. in derjenigen der Verfügbarkeit von relevanten Informationen ist der Aspekt der Zuverlässigkeit und Aussagefähigkeit von Interpretationen dieser Daten angelegt. Unten (Kapitel 5) wird u.a. geschildert, inwieweit den Daten der amtlichen Statistik unterschiedliche Informationen entnommen werden können, in Abhängigkeit von der jeweiligen Betrachtungsperspektive. So können Studienstatistiken aus einzelnen Semestern, die zu Studienverlaufsdaten verknüpft werden, Informationen zu Studienbiographien einschließlich dem Studienabbruch- und Hochschulwechselfverhalten liefern. Diese lassen sich aus den semesterweise erhobenen (nicht zu Studienverläufen verknüpften) Bestandsstatistiken nicht ablesen, so dass diese für Auswertungen zu Studienabbruch und Hochschulwechsel nicht in hinreichend zuverlässigem Maße genutzt und für die Hochschulplanung eingesetzt werden können. Durch die Art des Datenmanagements ist also bereits die Art der verfügbaren Informationen mitbestimmt, so dass diesem eine hinreichende Aufmerksamkeit bei der Etablierung und Implementierung von Qualitätssicherungssystemen geschenkt werden muss.

### 4.2.5 „Evidenzbasierte Praxis“ in der Hochschulsteuerung

Als Rahmen zur gedanklichen Verknüpfung der beschriebenen Konzepte der Qualitätssicherung und -entwicklung kann das Leitbild der evidenzbasierten (politischen) Praxis („Evidence-based Policy Practice“) dienen. Dieses hat insbesondere in Großbritannien, aber zunehmend auch im restlichen Europa in den letzten Jahren in einer Reihe von Politikfeldern in denen Leistungen der öffentlichen Verwaltung angeboten werden - Gesundheit, Justiz, aber auch im Bildungswesen - an Popularität gewonnen (Nutley u. a. 2003; Nutley 2003).

Dieser Aufschwung ist vermutlich auf die Tatsache zurückzuführen, dass grundlegende Prämissen dieses Leitbildes mit denen des aufstrebenden New Public Management in Einklang zu bringen sind. So entspricht die Formulierung „what matters is what works“ (vgl. Anderson und Bennett 2002; Nutley 2003)

einer Qualitätsprüfungskultur, in der Leistungen einer instrumentellen Rationalität folgend, hauptsächlich hinsichtlich ihres unmittelbar erfassbaren Nutzens beurteilt werden.

Von dem Einbezug wissenschaftlicher Evidenz in Entscheidungsprozesse wird sich dabei besseres, weil auf fachliche Expertise und wissenschaftliche Forschungsergebnisse gestütztes (und damit rationales) politisches Handeln versprochen. Als Evidenz werden daher auch systematische wissenschaftliche Untersuchungen, mit dem Zweck, die Summe des Wissens zu vergrößern, bezeichnet (Nutley u. a. 2003). Diese Definition schließt Expertenwissen, veröffentlichte Forschungsergebnisse, vorliegende Statistiken, Experteninterviews, Evaluationen vorangegangener Maßnahmen, kurz: eine einschlägige Datenbasis zur Beschreibung und Beurteilung eines Sachverhaltes ein. Dementsprechend wird durch den Term „Evidenz“ beschrieben, was (auch bezogen auf den Kontext von Lehre und Studium an Hochschulen) von Evaluation erwartet wird: Die Unterstützung der für Steuerungsentscheidungen Verantwortlichen, mit dem Ziel, deren Entscheidungen und Entscheidungsprozesse auf einer rationalen, im Sinne von auf systematische und wissenschaftliche Methoden gestützte Informationsquelle zu basieren.<sup>19</sup> Neben diesem Aspekt, in dem sich die technische Akzentuierung des Konzepts des Wissensmanagement (s.o.) widerspiegelt, ist jedoch ein weiterer Aspekt zentral für evidenzbasierte politische Praxis. Bei diesem handelt es sich um die Kommunikation der beiden beteiligten Sphären Wissenschaft und Politik (bzw. Praxis), die in einer Weise organisiert wird, die eine tatsächlich nutzbringende Verwendung<sup>20</sup> der wissenschaftlichen Evidenz sicherstellt. Dieser Aspekt rekurriert auf die Fähigkeit von wissenschaftlichen Forschungsergebnissen oder Evaluationser-

---

<sup>19</sup>Der Begriff der Rationalität von Datenerhebungen wurde bereits mehrfach erwähnt. An dieser Stelle sei klargestellt, dass sich das vorliegende Begriffsverständnis eher eine allgemeine bzw. alltagssprachliche Definition von Rationalität zu Nutze macht.

<sup>20</sup>Als Standardwerk zur nutzenorientierten Evaluation kann die Arbeit von Patton (1997) zur „Utilization-focused Evaluation“ genannt werden. Auch in die „Standards für Evaluation“ der Deutschen Gesellschaft für Evaluation (DeGEval) ist das Nützlichkeitskriterium als zentraler Aspekt für die Beurteilung von Evaluationen aufgenommen worden (DeGEval 2002).



gebnissen, über ihre tatsächliche Nutzung für Organisationsentwicklungen zu organisationalen Lernprozessen beizutragen.

Vor dem Hintergrund der praktischen Umsetzung evidenzbasierter Steuerung sind darüber hinaus verschiedene Fragen zu klären. So ist neben dem Aspekt der zeitlichen Nähe von Evaluationen zu den durch sie unterstützten Entscheidungsprozessen (Pawson 2002) hinaus zu diskutieren, welche Anforderungen an wissenschaftliche Forschungsergebnisse zu stellen sind, damit diese als Evidenz bezeichnet werden können. Unmittelbar berührt ist damit aus vorliegender Sicht wiederum der Aspekt der Gütekriterien von Messungen. Dort wo diese bedroht oder verletzt sind, muss die Eignung der entsprechend erbrachten Ergebnisse, evidenzbasiertes Handeln und Steuern zu ermöglichen, bezweifelt werden. Ein Beispiel für die Relevanz dieser Frage hinsichtlich der Nutzbarkeit von erhobenen Daten für Steuerungsentscheidungen und Beurteilungen beschreibt Desforges (2002). Anhand der Frage ob der im Rahmen von entsprechenden Evaluationen als Qualitätsmerkmal für Schulunterricht erhobene Geräuschpegel im Klassenraum als Zeichen von Engagement oder als Indikator für die Ablenkung der Schüler gewertet werden muss, ist zu diskutieren ob „(...) evidence of busyness stand(s) as a valid evidence of comprehension or even task engagement“ (Desforges 2002). Für manche spricht ein Geräuschpegel für eine aktive Arbeitsatmosphäre im Klassenraum, andere interpretieren den gleichen Zustand als Störung und auffälliges Schülerverhalten (a.a.O.). Die Art der Operationalisierung, wie hier der Qualität von Schulunterricht und die Art der entsprechenden Datenerhebung sowie die Interpretation erhobener Daten, ist demgemäß entscheidend für spätere Nutzungsmöglichkeiten der Evaluation in (politischen) Entscheidungsprozessen. Aus diesem Grund ist die Gewährleistung sachgemäßer Analysen von relevanten und hinsichtlich ihrer Qualität überprüften Daten von Bedeutung. Gleichzeitig besteht aber auch in der gemeinsamen Planung der Evaluation bezüglich ihrer Ziele und der beurteilten Kriterien durch die am evaluierten Programm Beteiligten einerseits und die Evaluation Durchführenden andererseits eine wichtige Voraussetzung für eine auf einschlägiger wissenschaftlicher Evidenz beruhende politische Praxis.

Evaluation, die den im vorliegenden Kapitel beschriebenen Kriterien gerecht wird, kann zu einem wichtigen Instrument der Hochschulsteuerung entwickelt werden. Evaluation in diesem Sinne wird über ein Kontrollinstrument hinaus zu einem wichtigen Element bzw. zur Voraussetzung eines auf Interessenausgleich, Beteiligung der Akteure und Qualitätsentwicklung zielenden Hochschulmanagements. Ein auf diesen Prämissen beruhendes partizipatives Management von Hochschulen kann im Sinne der „Selbstorganisation komplexer Sozialsysteme“ (Willke 1998a, zit. n. Nickel 2007, S. 52) zur demokratischen Entwicklung von Hochschulen beitragen, unabhängig von den von ihnen geforderten Leistungsnachweisen und Effizienzprüfungen.

# Kapitel 5

## Qualitätssicherung mittels Lehrevaluation und Leistungskennzahlen

Das in Kapitel 2 beschriebene veränderte Leitbild von Hochschulen, in dem diese zunehmend als marktorientierte Wettbewerber gesehen werden, die auf einem (zunehmend globalisierten) Bildungsmarkt um die Abnehmer ihrer Leistungen in Lehre und Forschung konkurrieren, hat in den vergangenen 15 Jahren in Deutschland zu einem rapiden Anstieg der Anwendung von verschiedenen Verfahren zur Qualitätsbeurteilung von Hochschulleistungen geführt. Auch im Leistungsbereich Lehre fand dabei in den letzten Jahren ein „Evaluationsboom“ statt (vgl. Kap. 3), mit dem die Entwicklung und Professionalisierung bzw. Institutionalisierung von Evaluationsverfahren einher ging und diese vermehrt für die Beurteilung von Lehre und Studium nutzbar gemacht wurden (Teichler 2005).

Gegenstand des vorliegenden Kapitels ist eine Beschreibung der gegenwärtigen Evaluationspraxis an Hochschulen. Insbesondere wird dabei auf die mit den Verfahren verbundenen Begriffe eingegangen, die in diesem Feld der Hochschulpolitik und -planung vielfach mehrdeutig bzw. synonym gebraucht werden und entsprechend spezifiziert werden müssen. So werden vielfach die Begriffe Evaluation und Qualitätssicherung im gleichen Sinne verwendet. In einem engeren Verständnis ist allerdings das Verhältnis dieser beiden Begriffe eher als Prozess

zu verstehen: Evaluationsverfahren werden mit dem Ziel bzw. zum Zweck der Qualitätssicherung angewendet. Auch auf der Ebene der dementsprechend eingesetzten Verfahren sind begriffliche Spezifikationen vorzunehmen. So wird zum Zweck der Lehrevaluation vielfach auf studentische Befragungsdaten (bspw. zu deren Beurteilung der Lehrpräsentation und/oder der allgemeinen Studienbedingungen) zurückgegriffen. Die Berechnung und Nutzung von Leistungskennzahlen - bspw. von Absolventenquoten aus Daten der Hochschulstatistik - ist nach diesem Verständnis nicht zur Lehrevaluation zu rechnen, wird aber mit dem gleichen Ziel, nämlich zur Beschreibung, Sicherung und Entwicklung der Qualität beizutragen, durchgeführt.<sup>1</sup> In diesem Sinne werden die verschiedenen Verfahren in der vorliegenden Arbeit zusammenfassend als Methoden zur Evaluation von Lehre und Studium mit dem Zweck der Qualitätssicherung aufgefasst. Beide unterliegen in gleicher Weise den in dieser Arbeit aufgeworfenen Problemen hinsichtlich der Datenqualität der erhobenen Informationen.

## 5.1 Evaluation als Mittel zur Qualitätsentwicklung

Zur Beurteilung der Qualität von Hochschulleistungen werden in der derzeitigen Praxis vielfach zwei sich gedanklich ergänzende Verfahrensstränge eingesetzt. Bei diesen handelt es sich um Evaluation und Akkreditierung. Letztere wurde insbesondere durch die im Bologna-Prozess angestrebten Veränderungen der Hochschullandschaft (u.a. Harmonisierung europäischer Bildungswege und der formalen Bildungsabschlüsse, steigende Qualitätssorge auf der institutionellen Ebene der einzelnen Hochschulen)<sup>2</sup> nötig, da eine Vielzahl neu eingeführter Studiengänge

---

<sup>1</sup>Zudem werden die genannten Quoten in ungleich höherem Maße für Zwecke der Rechenschaftslegung und die Begründung von Allokationsentscheidungen, als für die Entwicklung interner Verbesserungsstrategien eingesetzt.

<sup>2</sup>Die Implikationen des Bologna-Prozesses für die Einführung neuer Studiengänge fasst bspw. Jahn (2001) zusammen.

a priori hinsichtlich ihrer Studierbarkeit, der Arbeitsmarktgängigkeit der vergebenen Abschlüsse, einheitlicher Qualitätsstandards, etc. zu beurteilen war bzw. sein wird (z.B. Reuke 2002).<sup>3</sup> Gegenüber der Evaluation als Verfahren ist hier das wesentliche Unterscheidungsmerkmal der Akkreditierung von Studiengängen zu sehen: Diese besteht in einer ex ante Bewertung von Studiengangskonzepten (Schade 2002), während Evaluation in der Bewertung erbrachter Leistungen, also einer ex post Beurteilung des Erreichten besteht.<sup>4</sup> Weiterhin ist unter den derzeitigen Bedingungen einer getrennten Durchführung von Akkreditierung und Evaluation<sup>5</sup> festzustellen, dass in Letzterer die Erhebung studentischer Qualitätsurteile anders als in Evaluationen nur eine untergeordnete Rolle spielt (Reuke 2002). Akkreditierungen zielen mehr auf die curriculare Qualität von Studiengängen (Studieninhalte, Arbeitsmarktgängigkeit der erworbenen Abschlüsse, etc.), während die Prozesse von Lehre und Studium eher im Fokus der Evaluation stehen.

Gerade diese Prozesshaftigkeit von Lehre und Studium und die Erhebung der diesbezüglichen studentischen Urteile sowie die Qualität zu diesem Zweck erhobener Daten sind aber Gegenstand der vorliegenden Arbeit. Aus diesem Grund

---

<sup>3</sup>Eben diese unüberblickbare Vielzahl von Studienangeboten und der zu erbringende Aufwand hat bereits den Weiterentwicklungsbedarf des ursprünglichen Modells der Akkreditierung auf der Ebene einzelner Studiengänge offenbart. Diese geht in Richtung der Akkreditierung ganzer Institutionen oder „Cluster-Akkreditierungen“ (Arnold 2006) sowie in Richtung einer Akkreditierung der internen Qualitätssicherungsverfahren im Sinne einer System- oder Prozessakkreditierung die den Hochschulen das Vorliegen eines effizienten Qualitätsmanagements bescheinigt sowie die Fähigkeit selbstständig für die durch die Programmakkreditierung angestrebte Qualitätssicherung zu sorgen (Hofmann 2006; Hofmann 2006a).

<sup>4</sup>Zum Verfahren der Akkreditierung vgl. bspw. auch Reil (2002); Jahn (2002); Naderer (2001); Schenker-Wicki (2002).

<sup>5</sup>Reuke (2002) betont, dass aufgrund des erwähnten Aufwandes sowie aufgrund der inhaltlichen und methodischen Nähe der Verfahren, eine Verschmelzung erwartet werden kann. Weiterentwicklungen der Akkreditierungsverfahren, die auf eine Vereinfachung, bspw. durch eine Verlagerung der Begutachtung von Studiengängen auf diejenige hochschulinterner Prozesse zur Qualitätssicherung (Prozessakkreditierung), werden derzeit diskutiert und erprobt (vgl. z.B. Hofmann 2006; Hofmann 2006a; Arnold 2006).

wird bei der folgenden Darstellung von Methoden und Verfahren der Qualitätssicherung auf eine weitere Erläuterung der Akkreditierung verzichtet und das Augenmerk auf die Praxis der Evaluation von Lehre und Studium gerichtet.

### 5.1.1 Studentische Lehrveranstaltungskritik

Studentische Lehrveranstaltungsbewertungen gehören seit den 1990er Jahren zum „Standardrepertoire der Bemühungen um die Verbesserung der Qualität der Lehre“ (Pasternack 2004, S. 100). Inzwischen ist zur Beschreibung des Verfahrens, zu den eingesetzten Datenerhebungsverfahren und -instrumenten, zu den methodologischen Implikationen, zur hochschulpolitischen Frage nach der Nutzung der Ergebnisse für Steuerungsentscheidungen sowie zu einzelnen Projekten von Hochschulen oder Fächern ein nahezu unübersehbarer Literaturkörper, allein im deutschsprachigen Raum entstanden. Überblicksartikel und Monographien zum Thema finden sich bspw. bei Spiel (2001a), Willems u. a. (1994), Hage (1996), Rindermann (1995a), Rindermann (1997). Auch existiert inzwischen eine Vielzahl von Datenerhebungsinstrumenten, die u.a. auch als Referenz zur Durchführung eigener Evaluationsprojekte vorliegen (z.B. Rindermann und Amelang 1994).

Gegenstand dieser Befragungen sind zumeist dozentenbezogene Aspekte der Lehrveranstaltung (Präsentationsstil, Betreuungsleistungen, etc.) sowie das didaktische Modell (Struktur der Veranstaltung, Schwierigkeitsgrad, Medieneinsatz, u.ä.). Darüber hinaus werden vielfach äußere Bedingungen der Lehre und organisationale Aspekte abgefragt sowie die Einschätzung der Studierenden hinsichtlich des studentischen Beitrages zum Gelingen der Veranstaltung (Engagement, Vorbereitungsleistungen und bspw. Qualität studentischer Beiträge wie Referate).

Das zentrale Anliegen der studentischen Lehrveranstaltungskritik besteht zumeist darin, Studierenden die Gelegenheit zu geben, die Qualität der konkreten Lehrpräsentation, so wie sie sich ihnen darstellt, zu beurteilen und so den Lehrenden ein entsprechendes Feedback hinsichtlich etwaiger Verbesserungsmöglichkeiten zu geben. In diesem Sinne steht die interne Optimierungsfunktion auf der Ebene der jeweils beurteilten Lehrveranstaltung im Vordergrund. Die Ergebnisse

der Befragungen Studierender sollen die Diskussion Lehrender und Studierender über Defizite anschieben und über das Gespräch die entsprechenden Verbesserungen bewirken.

In letzter Zeit wird jedoch unter dem Vorzeichen der geforderten Verantwortlichkeit und Autonomie der Hochschulen verstärkt die Forderung erhoben, Instrumente der Lehrevaluation für darüber hinaus gehende Zwecke einzusetzen, namentlich für die Leistungsbeurteilung und leistungsbezogene Mittelvergabe (vgl. z.B. Kultusministerkonferenz; Müller-Böling 2000). In diesem Zusammenhang gewinnt die in Kapitel 6 beschriebene Debatte um die Frage nach der Validität von Daten der Lehrevaluation und der studentischen Veranstaltungskritik erneut an Bedeutung: Die Daten der Lehrevaluation für die Begründung einer leistungsorientierten Beurteilung oder Mittelvergabe in Hochschulen heranzuziehen, würde voraussetzen, ihre Aussagefähigkeit - im Sinne valider und reliabler Messungen - möglichst einwandfrei zu belegen. Anderenfalls bestünde die Gefahr, die leistungsorientierte Mittelvergabe auf Fehleinschätzungen bezüglich der Lehrqualität aufzusetzen. Nicht zuletzt ein fehlender glaubhafter Ausschluss dieser Möglichkeit, resultiert vermutlich in einer Einschränkung der Akzeptanz des Verfahrens bei den beteiligten Lehrenden.

### 5.1.2 Interne Evaluation auf der Ebene von Studienfächern

Interne Evaluation ist der erste Schritt des dreistufigen Modells der Lehrevaluation, welches sich insbesondere an niederländischen Hochschulen durchgesetzt hat (Vroeijenstijn 1994). Inzwischen erfreut sich dieses Modell auch in Deutschland einer breiten Anwendung.<sup>6</sup>

Neben der Selbstbegutachtung auf der Ebene von Fächern oder Studiengängen wird dieses Modell durch die Schritte der externen Evaluation durch

---

<sup>6</sup>Einen Überblick zu verschiedenen Evaluationsinstrumenten im deutschen und europäischen Hochschulraum geben Mittag u. a. (2003).

fachnahe Gutachter („Peer-Review“) und die Vereinbarung von Folgemaßnahmen als Ergebnis der Evaluation und zur Umsetzung der aus ihr abgeleiteten Empfehlungen konstituiert.<sup>7</sup>

Gegenstand der internen Evaluation von Fächern und Studiengängen ist eine systematische Bestandsaufnahme zu den Studien- und Lehrbedingungen des Faches. Die entsprechenden Analysen werden in der Selbstverantwortung des Faches durchgeführt (vgl. z.B. Reissert und Konnerth 2001; Reissert 1994; Carstensen und Reissert 1995; ZEvA 2003). „In einem möglichst diskursiven Prozess gilt es, Konzeption und Praxis der Ausbildung zu überprüfen und zu hinterfragen sowie zu überlegen, was in Studium und Lehre künftig besser gemacht werden kann“ (Reissert und Konnerth 2001, 183). Über entsprechende Diskussionen hinaus, werden auch Bestandsaufnahmen zur Studiensituation und den sie bestimmenden Rahmenbedingungen des Studiums (Erfassung von Ausstattungsmerkmalen, Zahlen zur Auslastung) und zur Studienorganisation (Prüfungswesen, Studierbarkeit) angefertigt. Ergänzt werden diese Daten in der Regel um solche, die sich auf die hochschulstatistisch erfassten Kennzahlen wie die Absolventenrelation, die Anzahl Studierender in der Regelstudienzeit, etc. beziehen. Um derlei objektivierende Kennzahlen mit den Qualitätseinschätzungen der an der Lehre beteiligten Gruppen zu kontrastieren, kommen bei der Datenerhebung Befragungen von Lehrenden und Studierenden sowie Absolventen zum Einsatz. Mittels standardisierter Befragungen, aber auch moderierter Gruppengespräche

---

<sup>7</sup>Neben diesen Schritten werden vielfach Lehrberichte, als Sammlung von Basisdaten zu Lehre und Studium angefertigt. Diese sind zumeist Teil des von den Hochschulen geforderten und durch Hochschulgesetze der Länder geregelten Berichtswesens gegenüber den Ministerien (vgl. z.B. Hochschulrektorenkonferenz 2000a) Die Praxis der Lehrberichterstattung ist jedoch sehr unterschiedlich, insbesondere hinsichtlich der Nutzung für nachfolgende Verbesserungsmaßnahmen. Vielfach sind die Berichte nicht (hochschul-)öffentlich zugänglich und haben entsprechend nur eine geringe Reichweite bezüglich der mit Evaluation verbundenen Absicht, zu Verbesserungen des Lehr- und Studienbetriebs beizutragen. Auch beziehen Lehrberichte zumeist hauptsächlich hochschulstatistische Daten (bspw. Kennzahlen zur Auslastung der Fächer) ein (vgl. z.B. Sandfuchs und Stewart 2006).



werden deren Hinweise zu etwaigen Verbesserungsmöglichkeiten erfasst (Reissert und Konnerth 2001; Futterer 2005). Die Gruppengespräche dienen über die reine Datenerhebung hinaus dazu, den fachinternen Diskurs der Beteiligten aufrecht zu erhalten und zu intensivieren.

Die Ergebnisse der entsprechenden Datenauswertungen werden in einem Selbstreport zusammengefasst. Nach dem Verfahren des dreistufigen Modells der Evaluation bildet dieser Bericht generell die Grundlage für die externe Evaluation.<sup>8</sup> Selbstevaluationen werden aber insbesondere auch mit dem Zweck durchgeführt, unabhängig von der Art der weiteren Verwendung für externe Begutachtungen, die interne Selbstvergewisserung zur Qualität von Lehre und Studium zu bereichern. In dieser Funktion des Instruments der internen Selbstevaluation, ist sein besonderer Beitrag für das Ziel der Qualitätssicherung in Lehre und Studium zu sehen. Der Selbstreport kann Schwächen in den Organisationsabläufen zu Tage bringen und die fachinterne Diskussion über mögliche Verbesserungsmaßnahmen bereichern. Wichtig dafür ist eine regelmäßige Datenerhebung zu den genannten Kennzahlen und Qualitätsaspekten, um die Informationsbasis auf einem aktuellen Stand zu halten. Die erhobenen Qualitätsaspekte sollten ihrerseits ein gemeinsames Qualitätsverständnis der Beteiligten widerspiegeln, um eine anerkannte inhaltliche Relevanz zu besitzen (Mittag u. a. 2003).

### 5.1.3 Peer Review

Externe Evaluationen dienen dem Zweck, die Innensicht, die bei der internen Evaluation gewonnen wird, durch die unabhängige Außensicht von Gutachtern zu bereichern, bzw. zu ergänzen. „Die Verbindung von interner und externer Evalua-

---

<sup>8</sup>Der Selbstreport des Faches wird zur Grundlage der externen Begutachtung gemacht, indem er der Gutachterkommission vorab zur Verfügung gestellt wird (s.u.). Diese Vorgehensweise ist gleichzeitig der Auslöser für Kritik an dem Verfahren: Wird der Selbstreport einer Kommission aus externen Gutachtern vorgelegt und zur Grundlage für deren Beurteilung gemacht, scheint es schwierig zu sein, etwaige Selbstkritik „schonungslos“ offenzulegen (Brennan und Shah 2000, zit. nach Mittag u. a. 2003, S. 188).

tion gilt als ein Weg, mit der Schwierigkeit von Qualitätsbewertungen umzugehen: Die Doppelperspektive, so die Annahme erhöhe die Chancen, Qualitäten identifizieren zu können“ (Pasternack 2004, S. 76). Neben Fachkollegen (Peers) werden dabei auch fachfremde Wissenschaftler oder Vertreter der Berufspraxis des evaluierten Faches in die Kommission eingeladen. Das Fach hat bei der Auswahl der peers das Vorschlagsrecht. Im Mittelpunkt der Peer Review steht der Besuch des Faches durch die Gutachterkommission. Dieser wird durch die Kenntnisnahme des Selbstreports des Faches durch die Gutachter vorbereitet. Während des Besuchs führen sie Gruppengespräche mit Vertretern der Hochschule bzw. Fakultät, des evaluierten Faches sowie mit Studierenden und Lehrenden. Darüber hinaus werden vertrauliche Einzelgespräche mit Vertretern der Statusgruppen geführt (Carstensen und Reissert 1995; Reissert und Konnerth 2001).<sup>9</sup>

Zum Ende des Besuchs berichten die Gutachter gegenüber den Fachvertretern ihre ersten Eindrücke, die sie aufgrund der Begehung und des Selbstreports gewonnen haben. Später wird dem Fach ein ausführliches Gutachten mit Stellungnahmen aller Kommissionsmitglieder zugestellt. Dieses stellt die Basis für die weiteren Umsetzungsmaßnahmen der Evaluation im Fach dar (a.a.O.).

Während die externe Evaluation von Lehre und Studium auf der Ebene ganzer Studiengänge durch peers ein noch vergleichsweise junges Instrument ist, hat die Beurteilung von Leistungen in der Forschung durch Fachkollegen eine lange Tradition. Hier stellt es im Grunde das älteste Verfahren der Wissenschaftsevaluation dar und wird bspw. bei Berufungsverfahren, der Begutachtung von Anträgen auf Drittmittelförderung oder bei der Begutachtung von Publikationen praktiziert. In dem Maße, wie auch diese Form der wissenschaftlichen Fachdiskussion in jüngster Zeit ein Bestandteil der Hochschulsteuerung wird und die Qualitätsparameter von außen festgelegt werden, anstatt durch die einzelnen Wissenschafts-

---

<sup>9</sup>Für die Unterstützung der konkreten Planung und Durchführung der Begehung (Zeit- und Terminplanung, Zahl und Zusammensetzung der Gesprächsgruppen, etc.) sind bspw. vom HIS Hochschul-Informations-System GmbH differenzierte Leitfäden entwickelt worden, die im Internet einsehbar sind (<http://evanet.his.de>; zuletzt 28.7.2007).

disziplinen (bspw. mittels Quantifizierungen der Publikationstätigkeit im Rahmen der leistungsorientierten Mittelvergabe; vgl. Fröhlich 2006), wird die Anwendung der peer review in der Forschung jedoch kontrovers diskutiert.<sup>10</sup> Ungeachtet der Kritik schließt Daniel (2006), dass „sich dieses Verfahren nach Meinung der beteiligten Gutachter und der Fachangehörigen bewährt hat“, und dass durch die externe Begutachtung „ein Beitrag zur Förderung der Selbststeuerungskompetenz der evaluierten Einheit, zur strategischen Planung und zur Rechenschaftslegung gegenüber der Öffentlichkeit geleistet“ (Daniel 2006, S. 190) wird.

### 5.1.4 Follow-up Maßnahmen

Nach dem beschriebenen dreistufigen Modell besteht in der Umsetzung der durch die interne Evaluation und die im Gutachten der Peers erarbeiteten Verbesserungsvorschläge der abschließende Schritt der Evaluation eines Studienganges oder Studienfachs. Die Befürchtung, dass die entsprechenden Schritte aufgrund der tendenziell gegebenen Unverbindlichkeit des Verfahrens nicht oder nicht nachhaltig eingeleitet werden, hat dazu geführt, dass andere Steuerungsinstrumente in die Diskussion eingeführt und erprobt wurden, die stärker auf die Rechenschaftslegung der Verantwortlichen fokussieren. Die Hochschulrektorenkonferenz (2000a) betont, dass Evaluation nicht folgenlos bleiben darf. Als Konsequenz aus den im Evaluationsverfahren festgestellten Stärken und Schwächen müssen daher Konsequenzen gezogen werden. Dabei wird von der HRK auf unmittelbare

---

<sup>10</sup>So werden von den Kritikern nicht intendierte Effekte beschrieben, die bspw. einsetzen, wenn als einziger publikationsbezogener Qualitätsindikator der (zahlenmäßige) Output herangezogen wird (z.B. Fröhlich 2006). Dies kann zu einem Anstieg der Publikationstätigkeit bei gleichzeitigem Sinken ihrer Qualität im Sinne des Einflusses der Publikationen auf den Fachdiskurs (bspw. gemessen am „Journal-Impact-Factor“) führen. Die (vermeintliche) Steigerung der Produktivität entpuppt sich als opportunistischer Anpassungseffekt an die Evaluationsverfahren und wird zu einem Artefakt dieser Verfahren (a.a.O.). Weitere Kritikpunkte beziehen sich auf eine geringe Reliabilität (geringe Übereinstimmung der Urteile verschiedener Urteiler), den Einfluss nicht sachbezogener Urteilskategorien und einer geringen prädiktiven Validität der Urteile (Daniel 2006).

wie langfristige Konsequenzen angespielt. Die langfristigen Konsequenzen müssen nach Auffassung der Hochschulrektorenkonferenz (a.a.O., S. 8) in einer Verzahnung der Evaluationsergebnisse mit einer leistungsbezogenen Mittelallokation an den Hochschulen bestehen. Die Implikationen dieses Steuerungsleitbildes werden unten ausführlicher diskutiert. Hier wird auf die unmittelbaren Folgen der Evaluation eingegangen, die sowohl von der HRK gefordert werden, wie auch durch das dreistufige Modell der Lehrevaluation vorgesehen sind.

Bei diesen handelt es sich um Folgemaßnahmen, die zumeist im Steuerungsinstrument der Zielvereinbarung bestehen.<sup>11</sup> Zielvereinbarungen werden durch die Formulierung und vertragliche Festlegung von Entwicklungszielen sowie von Kriterien für die Zielerreichung gebildet (vgl. z.B. Pasternack 2004; Wottawa 2001). Geschlossen werden die entsprechenden Verträge sowohl zwischen zuständigen Ministerien und Hochschulen (z.B. Börger 2006), aber auch auf der Ebene der Hochschulen selber, hier zwischen der Hochschulleitung und einzelnen Bereichen wie Fakultäten oder Instituten (z.B. Zechlin 2006).

Die grundlegende Idee des Steuerungsinstruments der Zielvereinbarung besteht darin, durch einen Wechsel von der Input- zur Outputsteuerung die Autonomie der Hochschulen bei der Leistungserbringung zu stärken (vgl. z.B. Lewis 2005, S. 34). Unter Inputsteuerung ist die Kontrolle (bspw. durch Ministerien) des Prozesses der Leistungserbringung mittels der zur Verfügung gestellten Mittel (Haushaltsmittel, Finanzierung von Berufungen, etc.), zu verstehen. „Zu Beginn (...) wird ein Input eingespeist (Arbeitskraft, Sachmittel), der Prozess beginnt, und zum Ende erhält man einen bestimmten Output, der wiederum eine bestimmte gesellschaftliche Wirkung erzielt bzw. einen bestimmten Zielbeitrag erbringt, was als Outcome bezeichnet wird“ (Ziegele 2006, S. 78). Kennzeichnend für die Inputsteuerung ist ferner die Kontrolle von Verfahren und Verhaltensweisen der Beteiligten: Über eine Vielzahl von Regelungen und Vorschriften soll die ordnungs- und zielgemäße Verwendung der zur Verfügung gestellten Mittel

---

<sup>11</sup>Zur Praxis zum Steuerungsinstrument der Zielvereinbarung vgl. die ausführliche Beschreibung von Nickel (2007).

gewährleistet werden (Wottawa 2001; Ziegele 2006). Mit der Outputsteuerung soll insbesondere auch der starken Verregelung des Hochschulhandelns begegnet werden (Wottawa 2001). Dieses soll vielmehr an dem Grad der Leistungserbringung gemäß der festgesetzten Ziele beurteilt werden. Durch Zielvereinbarungen soll Verbindlichkeit für die Erreichung gesetzter Ziele hergestellt werden, ohne dass Eingriffe in die Leistungserbringung seitens des Mittelgebers erfolgen. Dieser stellt die zur Zielerreichung benötigten Mittel zur Verfügung, ohne die Art des Mitteleinsatzes im Detail zu überprüfen.

Vielmehr stellt der durch das autonome Hochschulhandeln erreichte Output das Maß für die Zielerreichung und damit zur Beurteilung des Hochschulhandelns dar. Je nach der Art der vereinbarten Ziele, die sich ihrerseits aus den übergeordneten Organisationszielen der beteiligten Vertragspartner ableiten (Pasternack 2004), kann es sich dabei um so unterschiedliche Dinge wie die Steigerung der Absolventenzahl in bestimmten Studiengängen oder die Einführung mehrsprachiger Studiengänge (zur Steigerung der internationalen Reputation) handeln.

Das Modell zielt auf die Beteiligung der Hochschulen (und deren Mitglieder) an der Zielformulierung und damit auf deren Motivation, an der Zielerreichung mitzuwirken. Die Übertragung von Entscheidungskompetenzen an die Akteure in den Hochschulen ist darüber hinaus von der Annahme geleitet, dass dezentrale Entscheidungen problemorientierter, weil in engerer Verbindung zu den zu lösenden Problemen getroffen werden. „Zielvereinbarungen (bieten) einen immateriellen Anreiz, der sich in der Formel ‚Verbindlichkeit gegen Autonomie‘ zusammenfassen lässt: Die Selbstverpflichtung, kooperativ ausgehandelte Ziele auf selbstgewählten Wegen zu erreichen, bringt einen Gewinn an Entscheidungskompetenz. Voraussetzung dafür ist, dass die Wege zur Zielerreichung tatsächlich ins Belieben des Fachbereichs gestellt werden. Dies wiederum erfordert, dass Zielvereinbarungen auch wirkliche Ziele vereinbaren und nicht detaillierte Maßnahmen“ (Pasternack 2004, S. 55). Voraussetzung für die Anwendung des Instruments der Zielvereinbarung ist neben der Erbringung eines „messbaren Outputs“ (Wottawa 2001, S. 157) durch den Mittelempfänger Hochschule auch, dass die entsprechenden Messungen der Zielerreichung zum Zweck der Überprüfung der

„Vertragstreue“ vorgenommen werden. Um festzustellen, ob die vom Mittelgeber zur Verfügung gestellten Mittel ziel- bzw. vereinbarungsgemäß eingesetzt wurden, muss der entsprechende Output auch nachweisbar sein. Dieser Nachweis wird wiederum durch Evaluationsverfahren erbracht, in denen die Qualität der erbrachten Leistungen nachgewiesen, bzw. weiterer Entwicklungsbedarf dargestellt wird.

Zielvereinbarungen werden als „Kern des Neuen Steuerungsmodells“ angesehen (Ziegele 2006). Dessen zentrale Kennzeichen sind *Effizienz* (im Sinne von Wirtschaftlichkeit der Prozesse der Leistungserbringung) sowie *Wettbewerb* (um knappe öffentliche Mittel). Ergebnisse des Hochschulhandelns - also dessen Output - werden hinsichtlich des Verhältnisses von Zielvorgabe und Zielerreichung beurteilt und sanktioniert. In Zielvereinbarungen werden solche Vorhaben gefördert, von denen ex ante angenommen wird, dass sie sich durch ausgezeichnete Ergebnisse bewähren werden. „Der Mittelverteiler muss (...) Werturteile treffen und im Falle inhaltsoffen geführter Zielvereinbarungsrunden festlegen, ob er zehn Prozent mehr Ausländer in Fachbereich A oder eine Ausweitung der Industriekontakte in Fachbereich B höher honorieren will. Keine einfache Aufgabe, wie bereits vielfach in Zielvereinbarungsprozessen zu merken war“ (Ziegele 2006, S. 80 f.). Vor diesem Hintergrund zeigt sich die Brisanz, die mit dem Einsatz von Evaluationsverfahren zur Beurteilung von Hochschulleistungen verbunden ist. Diese geben über die Bestandsaufnahme zur Lehrqualität implizit Empfehlungen auch hinsichtlich der zukünftigen Förderungswürdigkeit von Bereichen ab, wenn an die Ergebnisse Entscheidungen in Zielvereinbarungsprozessen geknüpft werden. Entsprechend können sie von erheblichem Einfluss auf die Mittelverteilung und damit auf die einzelnen Hochschulmitglieder sein. Im Sinne der Wahrung gerechter Beurteilungsstandards, ist sicherzustellen, dass Evaluationsverfahren in Steuerungsverfahren sachgemäß eingesetzt und ihre Ergebnisse entsprechend interpretiert werden.

## 5.2 Quantifizierung von Qualität: Leistungsorientierte Mittelvergabe

Während Evaluationsverfahren zunächst eher beabsichtigen, intern Stärken und Schwächen zu erkennen und die Entwicklung von Verbesserungsstrategien anzuschließen, wurden in der letzten Zeit auch Verfahren zur leistungsorientierten Mittelvergabe erprobt (Leszczensky und Orr 2004; Jäger 2006). Diese stellen im Gegensatz zu internen Evaluationen ein Steuerungsinstrument dar, das von außen, in die Hochschulsteuerung eingreift und direkte Auswirkungen auf die Ressourcenausstattung der Hochschulen und Fachbereiche hat. In der Anwendung leistungsorientierter Steuerungsverfahren bei der Zuteilung von Mitteln wird eine Möglichkeit gesehen, wettbewerbliche Elemente in das Hochschulmanagement einzubauen und so die gewünschten Leistungssteigerungen und wissenschaftliche Profilbildungen von Institutionen und einzelnen Wissenschaftlern zu induzieren. Das Mittelvergabeverfahren bezieht dabei sowohl Hochschulen insgesamt (bezogen auf Ressourcen, die von einer Landesregierung an die einzelnen Hochschulen vergeben werden), als auch deren Gliederungen (Mittelverteilungen von der Hochschulleitung an die einzelnen Fachbereiche) ein.

Zu diesem Zweck werden Indikatoren eingesetzt, die im Sinne eines Vergleichs die Leistungen von Hochschulen in ihren verschiedenen Kernbereichen quantifizieren sollen. Diese Quantifizierung von Leistungen wird benutzt, um den diesbezüglichen Erfolg bzw. Misserfolg (positiv oder negativ) zu sanktionieren. Auf die Ergebnisse der derart durchgeführten Leistungsmessungen werden die Entscheidungen über die Zuweisung (bzw. Nichtzuweisung) von Mitteln aufgesetzt. Die Praxis der leistungsorientierten Mittelzuweisung ist im deutschen Hochschulsystem noch vergleichsweise jung. Dementsprechend ist sie sowohl hinsichtlich der tatsächlichen Steuerungswirkungen als auch bezüglich kritischer methodischer Fragen, bspw. zur Validität und Reliabilität der Qualitätsmessungen, die ihrerseits mittels zumeist *quantitativ* orientierter Leistungsindikatoren durchgeführt werden, wenig erforscht (Jäger 2006). Nichtsdestotrotz halten die Leistungsindikatoren mittlerweile in großem Umfang Einzug in den Hochschulall-

tag (Jäger 2006). Für den Gegenstand der vorliegenden Arbeit sind sie insofern relevant, als sie -zumindest in Teilen- die gleichen Datenquellen bemühen, die auch für intern organisierte Qualitätssicherungssysteme herangezogen werden und dementsprechend hinsichtlich der Frage nach potenziellen Bedrohungen der Validität und daraus entstehenden Fehlinterpretationen und Fehlschlüssen zu diskutieren sind. Bei diesen Datenquellen handelt es sich insbesondere um die Daten der amtlichen Hochschulstatistik. So wird bspw. die Zahl der Absolventen eines bestimmten Jahres in einem bestimmten Studienfach, die in diesem Zusammenhang als „Absolventenquote“ bezeichnet wird, typischerweise als Leistungsindikator herangezogen.<sup>12</sup> Es stellt sich also im Zusammenhang mit den Indikatoren der leistungsorientierten Mittelvergabe in Hochschulen in gleichem Maße die Frage nach der „Quantifizierbarkeit“ von Qualität, wie dies für die Evaluation und Qualitätssicherung von Lehre und Studium zutrifft. Nicht zuletzt aus Gerechtigkeitserwägungen scheint es geboten zu sein, die genannten methodologischen Probleme von Leistungsbeurteilungen, auf die Steuerungsentscheidungen aufgesetzt werden und die dementsprechend für die Hochschulen überaus einschneidende Auswirkungen haben können, anzusprechen.

### 5.2.1 Leistungsindikatoren in der Hochschulsteuerung

Den Überlegungen zu den eingesetzten Indikatoren der leistungsbezogenen Hochschulsteuerung muss vorangestellt werden, dass das deutsche Hochschulsystem im

---

<sup>12</sup>Dabei muss auf die begriffliche Unschärfe hingewiesen werden, die mit der Benutzung des Begriffes Leistungsindikator in diesem Zusammenhang verbunden ist. So werden zumeist durch die Zahl der Absolventen, Studienabbrecher, Studierenden, etc. eines Jahrgangs im wesentlichen Quantitäten ausgedrückt. Die diese beschreibenden Zahlen sind dementsprechend Kennziffern und keine Indikatoren. Zu Indikatoren werden Kennziffern, wenn sie „über ein theoretisches Konzept definiert werden“ (Pasternack 2004, S. 111). Für Indikatoren konstitutiv ist, dass sie nicht unabhängig von konzeptionellen Annahmen gelten, sondern Prozesse und Ergebnisse von Prozessen in einem numerischen Relativ abbilden (a.a.O.). Trotz dieser zu treffenden Unterscheidungen werden Kennziffern zur quantitativen Beschreibung von Inputs im gängigen Sprachgebrauch als Indikatoren bezeichnet.



wesentlichen in der Verantwortung und unter der Rechtsaufsicht des Staates liegt (s.o.) und dementsprechend die angestrebte Einführung von Marktprinzipien mit den Zielen der allgemeinen Bildungspolitik in Einklang zu bringen ist. Diese bestehen z.B. in der Erwartung an das Bildungssystem, im Sinne eines Ausgleichs regionaler oder sozialer Gefälle, flächendeckend ein hochwertiges und vergleichbares Bildungsangebot zur Verfügung zu stellen (vgl. z.B. Hocke 2003; Nagel 2003). Um etwaige, diesem Ziel widersprechende Auswirkungen einer ausschließlich nach Marktprinzipien organisierten Zuweisung von Mitteln auszugleichen, wird auch dort, wo bereits die leistungsorientierte Steuerung der Hochschulen erprobt oder angewendet wird, die Grundausrüstung gemäß der Nachfrage der Leistungen der Hochschulen (hinsichtlich Studierendenzahlen, Personalstellen, etc.) sichergestellt und die leistungsbezogene Mittelvergabe zu einem gewissen Budgetanteil auf den entsprechenden Sockel aufgesetzt (vgl. Leszczensky und Orr 2004). Dieses Vorgehen soll die Funktionsfähigkeit des in staatlicher Verantwortung stehenden Hochschulsystems sichern. Die leistungsorientierte Mittelvergabe wird mit einer Kappungsgrenze versehen, die den maximalen Rückgang der Mittel auf Grund der leistungsbezogenen Mittelvergabe festlegt. Je enger die Kappungsgrenze angesetzt wird, desto geringer ist der Leistungsbezug einzuschätzen, da der Hauptteil der Zuweisung faktisch unabhängig von der erbrachten Leistung fortgeschrieben wird (Leszczensky und Orr 2004, S. 45).

Im Gegensatz dazu steht der betont wettbewerbliche Charakter des indikatorgestützten Mittelvergabeverfahrens: Zwischen den Hochschulen soll ein Wettbewerb um die gemäß der Indikatoren vergebenen Mittel induziert werden. Auf diese Weise soll eine dezentrale Entscheidung der Zuweisungsempfänger über die Mittelverwendung und ihre Profilierung ermöglicht werden (s.o.). Gleichzeitig werden die Ziele des Mittelgebers verfolgt, indem Indikatoren herangezogen werden, die seine Prioritäten in den Leistungsbereichen der Hochschulen abbilden. Erfolgreiches Handeln (im Vergleich zu allen anderen „Wettbewerbern“) wird dabei belohnt, während Misserfolge sanktioniert werden, so dass Anreize zur Zielerreichung und zur Teilnahme am Wettbewerb entstehen (Ziegele 2001). In diesem Sinne sind die Leistungen der Hochschulen zu definieren als das Maß des Er-

folgs bei der Aufgabenerfüllung hinsichtlich der hochschulpolitischen Ziele des Mittelgebers (Leszczensky und Orr 2004, S. 45). Vorteile des Verfahrens werden insbesondere darin gesehen, dass

- vom Mittelgeber keine Ergebnisse vorgeschrieben, sondern vielmehr stabile Rahmenbedingungen für dezentrale, autonome Entscheidungen der Mittelempfänger gesetzt werden, wobei der Grad der Zielerreichung im nachhinein (ex-post) gemessen wird und
- eine outputorientierte Steuerung ermöglicht wird, die Zielorientierung und wirtschaftliches Handeln provoziert.

Gerade der erste der genannten Vorteile kann sich jedoch in sein Gegenteil verkehren, wenn auf Grund der durchgeführten ex-post Bewertung der Leistungen hauptsächlich solche Strategien verfolgt werden, von denen ex-ante bekannt ist, dass sie zum gewünschten Erfolg bzw. der entsprechenden Belohnung führen. Ein innovationshemmender Effekt ist dann möglich, wenn wissenschaftliche Unternehmungen in der Forschung, deren Erfolg zu Beginn ihrer Umsetzung nicht prognostizierbar ist, aus diesem Grund nicht angegangen werden (Schröder 2003). Die mögliche Innovationsfähigkeit von Forschungsvorhaben entzieht sich einer ex-post Bewertung, vielmehr ist hier zuweilen die Investition von „Risikokapital“ in Vorhaben nötig, die nur auf Grund einer ex-ante Entscheidung getätigt werden kann.<sup>13</sup>

### 5.2.2 Indikatoren im Leistungsbereich Lehre

Nichtsdestotrotz werden insbesondere im Bereich der Lehre Steuerungswirkungen durch den Einsatz leistungsorientierter Mittelzuweisung erhofft, weil die akademische Ausbildung an Hochschulen Gegenstand öffentlicher Diskussion und Kritik

---

<sup>13</sup>Um diesen innovationshemmenden Effekt zu verhindern, werden wiederum Zielvereinbarungen als ergänzendes Instrument zur innovationsfördernden Vorfinanzierung empfohlen (vgl. z.B. Ziegele 2001; Schröder 2003). Gleichzeitig ist eine Leistungsdefinition (die ihrerseits ja ein Kernelement der leistungsorientierten Steuerung ist) nur sehr allgemein möglich.

ist (s.o.). Wichtige hochschulpolitische Themen sind dabei die Verkürzung der Studiendauer, die Steigerung der Absolventenzahlen und ein stärkerer Berufsfeldbezug der Lehre. Dementsprechend werden in Modellen der leistungsorientierten Mittelvergabe

- die Zahl der Studierenden in der Regelstudienzeit (bzw. die Auslastungsquote der Studiengänge als relative Zahl bezüglich der studentischen Nachfrage),
- die Zahl der Absolventen in ihrer Veränderung zu einem früheren Vergleichszeitraum,
- die Absolventenzahlen in Relation zur Zahl der Studierenden im Jahrgang,
- die Zahl der Absolventen in der Regelstudienzeit in Relation zur Absolventenzahl insgesamt sowie die
- Absolventenzahlen in Relation zur Zahl der Studierenden in der Regelstudienzeit

für die Leistungsbeschreibung herangezogen (Leszczensky und Orr 2004; vgl. auch Pasternack 2004, S. 118). Die Zahl der Studierenden einer Hochschule (in der Regelstudienzeit) drückt die Ressourcenbelastung der Hochschulen aus, die über die Mittelzuweisung honoriert werden soll. Der mit diesem Indikator verbundene Leistungsanreiz besteht für die Hochschulen darin, ihr Lehrangebot attraktiver zu machen, um mehr Studierende an sich zu binden. Die diesbezügliche Qualität und Reputation soll den Studieninteressierten bspw. in einschlägigen Rankings angezeigt werden (Hornbostel 2001). Allerdings steht diesen der Einwand gegenüber, dass die Studienwahl nicht ausschließlich nach qualitäts- oder reputationsbezogenen Eigenschaften der Hochschulen getroffen wird, sondern auch vor dem Hintergrund persönlicher und sachfremder Motive, wie z.B. der Nähe der Hochschule zum Heimatort (vgl. z.B. Engel und Pohlenz 2001; Federkeil 2005). Der Informationswert von Rankings ist darüber hinaus insbesondere dadurch als eingeschränkt anzusehen, dass sich die Qualitätsurteile auch auf potenziell subjektiv verzerrte studentische Einschätzungen stützen und die Einteilung der Hoch-

schulen in Ranggruppen mit Schätzunsicherheiten und Verzerrungen verbunden sein kann (Engel und Krekeler 2001).<sup>14</sup> Zudem wird der gewünschte Wettbewerb durch Beschränkungen der Zulassung von Studierenden verzerrt und die entsprechenden Gestaltungsmöglichkeiten der Hochschulen eingeschränkt. Insbesondere Hochschulen, die viele zulassungsbeschränkte Studiengänge anbieten, werden so benachteiligt. Zur vollen Entfaltung einer Anreizwirkung müssten hier die entsprechenden Teilnahmebarrieren am Wettbewerb fallen (Leszczensky und Orr 2004).<sup>15</sup>

In der Praxis der indikatorgestützten Leistungsbeurteilung von Hochschulen, wird für die herangezogenen Indikatoren vielfach beansprucht, zur Beschreibung und Bewertung von Outputs des Handelns von Hochschulen geeignet zu sein. Durch die Messung bestimmter Ergebnisse in den verschiedenen Leistungsbereichen, wie oben beschrieben, sei es möglich, Leistungen der Hochschulen in einer vergleichenden Perspektive zu beurteilen und hinsichtlich ihres diesbezüglichen Erfolges bzw. Misserfolges zu sanktionieren. Tatsächlich beschreiben die meisten gewählten Indikatoren aber die Inputs bzw. die Nachfrage nach den Leistungen einer Hochschule, wie z.B. Studierendenzahlen und Personalstellen (Leszczensky und Orr 2004, S. 8; Engels 2004). Der relevante Unterschied zwischen bei-

---

<sup>14</sup>Diese Schätzunsicherheit bezieht sich auf die Konfidenzintervalle der Stichprobenmittelwerte der einzelnen „gerankten“ Merkmale, sofern diese auf Verteilungen von Erhebungsdaten zurückgehen, die ihrerseits mit Bewertungsskalen gewonnen wurden. Eine wichtige Voraussetzung für das Vorliegen tatsächlicher Unterschiede der Stichprobenmittelwerte ist darin zu sehen, dass sich deren Konfidenzintervalle nicht überlappen. Andernfalls lässt sich die Möglichkeit nicht ausschließen, dass lediglich Zufallsvariationen vorliegen und die Transformation dieser Unterschiede in Ranginformationen eine Fehlinterpretation bedeuten würde (Engel und Krekeler 2001, S. 124 f.).

<sup>15</sup>Die Zuverlässigkeit von Rankings für die Informierung von Studieninteressierten über die unterschiedliche Qualität von verschiedenen Studienstandorten aber auch für die vergleichende Leistungsbeurteilung von Hochschulen insgesamt ist Gegenstand einer breiten, kontroversen Debatte, die im Kontext der vorliegenden Arbeit jedoch nicht vertieft werden soll. Es sei beispielhaft auf die entsprechenden Beiträge in Engel (2001) und Müller-Böling u. a. (2001) sowie auf Müller-Böling (2000) und Hornbostel und Daniel (1994) verwiesen.

den besteht darin, dass Leistungsindikatoren, sofern sie Leistungen im Sinne von Erfolgen messen sollen, ex-post - also als Überprüfung eines Erfolgs, der sich durch den Einsatz der zur Verfügung stehenden Mittel erklären lässt - gemessen werden müssen. Input-Indikatoren werden jedoch lediglich ex-ante gemessen und beschreiben somit bestenfalls die unterschiedlichen Voraussetzungen der einzelnen Hochschulen in dem eigentlich zu induzierenden Wettbewerb. Darüber hinaus fehlen in dem als Leistungsvergleich angekündigten „Input-Output-Vergleich“ Zahlen oder Kennziffern, die den *Outcome* des Hochschulhandelns beschreiben. In diesem spiegelt sich das Maß wider, zu dem durch den (objektiven) Output, der sich in Quantitäten ausdrücken lässt, die übergeordneten (Qualitäts-)Ziele der Hochschule erreicht wurden.

Bezogen auf den Leistungsbereich Lehre ließe sich beispielhaft anführen, dass die Information über die Zahl der Absolventen einer Hochschule erst durch einen Hinweis auf den beruflichen Erfolg dieser Absolventen aussagefähig gemacht wird. Beruflicher Erfolg, der den Outcome der Hochschule beschreibt, lässt sich seinerseits bspw. anhand einer ausbildungsadäquaten Beschäftigung, einer angemessenen Arbeits- und Erwerbssituation, angemessenen Karriereperspektiven, der Fähigkeit und Gelegenheit, akademisch zu arbeiten u.ä. empirisch feststellen. Entsprechende qualitative Daten zu erheben, verursacht jedoch einen vergleichsweise hohen Aufwand. In den von den Ländern der Bundesrepublik angewandten oder erprobten Modellen zur indikatorgestützten Mittelvergabe finden sie bislang im Grunde keine Berücksichtigung (Leszczensky und Orr 2004, S. 2), so dass die Leistungsbeurteilung der Hochschulen in diesem Sinne unvollständig bleiben muss. In dem Maße, wie es den Modellen zur indikatorgestützten Mittelvergabe nicht gelingt, ein umfassendes, valides und neben Quantitäten auch Qualitäten einbeziehendes Bild der Leistungen der Hochschule zu zeichnen, wird die Akzeptanz des Verfahrens durch die den Messungen ausgesetzten Hochschulangehörigen möglicherweise nur schwer herzustellen sein.

Zur Vermeidung von Sanktionen, die sich in einer geringeren Mittelzuweisung ausdrücken, werden dadurch möglicherweise sogar opportunistische Anpassungsreaktionen (nachsichtigere Benotung von Studienleistungen, Anpassung

der Publikationsstrategien, etc.) provoziert. Derartige Effekte würden - zumindest aus akademischer Sicht - zu Qualitätsverlusten führen und dementsprechend hauptsächlich gegenteilige, bzw. nicht intendierte Effekte haben.

Die leistungsorientierte Mittelvergabe fokussiert in ihrer derzeitigen Praxis stark auf die Verwendung rein quantitativer Leistungsindikatoren. Da der Leistungserbringung in der Hochschule jedoch grundsätzlich eine qualitätsbezogene Dimension innewohnt, bleibt zu prüfen, inwieweit quantifizierende Indikatoren geeignet sind, neben Quantitäten auch die Qualität des Hochschulhandelns adäquat abzubilden. Die hier aufgezeigten potenziellen Bedrohungen der Validität der Leistungsbeurteilungen beziehen sich direkt auf die beschriebenen quantitativen Indikatoren (Studierendenzahlen, Zahlen zum Umfang der Publikationstätigkeit der Wissenschaftler einer Hochschule, etc.) die ihrerseits zugleich zur Beschreibung der Qualität und darüber hinaus sogar zu Vergleichszwecken bei der Leistungsbeurteilung herangezogen werden. Im Sinne einer tragfähigen und Gerechtigkeitskriterien entsprechenden Leistungsbeurteilung sind die sich potenziell ergebenden Fehlschlüsse als ernstzunehmender Einwand gegen die Verwendung ausschließlich auf Quantitäten bezogener Indikatoren zu sehen. „Betriebliche Effizienz drückt sich in Input-Output-Relationen und entsprechenden Renditen aus. Im universitären Bereich können wir Effizienz hingegen nicht als Resultat von Input-Output-Relationen betrachten; Universitäten sind keine Organisationen zur Realisierung vorab definierter Effizienzziele. Sie sind Brutstätten für schöpferische Ideen. Die Qualität der Universitäten erweist sich nicht in Unterwerfung und Befolgung ministerieller Vorschriften; sie zeigt sich im [akademischen; P.P.] Wettbewerb (...)“ (Starbatty 2004).

Das Steuerungsziel, dass mit der Berücksichtigung der Absolventenzahlen verbunden wird, besteht in einer Steigerung der Effizienz des Hochschulstudiums, die sich in einer gesteigerten Anzahl derjenigen Studierenden ausdrückt, die das Studium (in möglichst kurzer Zeit) mit einem Examen abschließen. Diese Zahl ist bislang in Deutschland im Vergleich zu anderen OECD-Nationen unterdurchschnittlich (Stange 2003). Die Würdigung der Absolventenzahl soll einen Anreiz zur Verbesserung der Studienstruktur und Studienorganisation liefern. Auch mit

diesem Indikator sind jedoch verschiedene Probleme verbunden. So ist die Verantwortung für das Phänomen des Studienabbruchs nicht nur bei den Hochschulen anzusiedeln, sondern lässt sich auch auf Gründe zurückführen, die in persönlichen Dispositionen der Studierenden oder in Aspekten der Studienrealität zu finden sind, die von den Hochschulen nicht direkt beeinflussbar und entsprechend nicht zu verantworten sind (vgl. z.B. Pohlenz und Tinsner 2004; Pohlenz u. a. 2007). Weitere denkbare verzerrende Einflussfaktoren bestehen in der regional und fachbezogen sehr unterschiedlichen Arbeitsmarktsituation, die in Teilen nur sehr bedingt einen Anreiz bietet, das Studium zügig zu beenden und in das Berufsleben einzutreten. Außerdem werden auf diese Weise Hochschulen benachteiligt, die stark von Hochschulwechsel betroffen sind. Deren Beitrag zum erfolgreichen Absolvieren eines Studiengangs durch Studierende bleibt unberücksichtigt (Leszczensky und Orr 2004, S. 58). Generelle Einwände betreffen die Gefahr, dass Reaktionen der Hochschulen auf festgestellte Defizite sich nur langfristig in einer Veränderung der Absolventenquote niederschlagen und dementsprechend nicht in die unmittelbare Leistungsbeurteilung Eingang finden. Organisationsverbesserungen auf dem Feld der Studiensituation der Verkürzung der Studienzeit anzustreben, findet bei der leistungsabhängigen Mittelzuweisung also keine zuzuordnende Berücksichtigung, so dass die Anreizwirkung begrenzt bleibt (a.a.O.: S. 58).

Schließlich ist in der Qualität der herangezogenen Daten bzw. in Problemen der landläufig eingesetzten Routinen ihrer Auswertung ein schwer wiegender Einwand gegen die Verwendung von Studierendenzahlen und Absolventen- oder Abbruchquoten begründet. Diese sind mit Problemen behaftet, die sich aus der Tatsache ergeben, dass Daten der amtlichen Hochschulstatistik generell als Bestandsstatistik, nicht aber als Verlaufsstatistik geführt werden (Arnold 1997):

- Abbrecher- und Absolventenquoten beziehen sich immer auf jeweils unterschiedliche Gruppen von Studierenden. (Es wird die Zahl der Studienanfänger eines Jahres mit der Zahl Studierender verglichen, die nach einer bestimmten Zeit noch in diesem Studiengang eingeschrieben sind. Aus der entsprechenden Verhältniszahl werden bspw. Abbrecher- oder Schwun-

quoten berechnet. Nicht berücksichtigt werden dabei aber Zugänge und Abgänge in und aus höheren Fachsemestern, so dass die berechneten Quoten verzerrt und mithin in ihrer Validität bedroht sind. Insbesondere der hochschulbezogene Vergleich von derartigen Quoten wird auch dadurch erschwert, dass vielfach völlig unterschiedliche Berechnungsgrundlagen und -verfahren herangezogen werden (Ziegele 1997).

- Weil die Informationen nur auf der Ebene der Zahlen von Studierenden, die in einem Semester immatrikuliert sind vorliegen, sind bestenfalls Trendaussagen möglich, der reale Nachvollzug von Studienverläufen auf der Ebene von Studienfällen ist dagegen auf diese Weise nicht erreichbar. Derartige Daten sind aber zur Errechnung von exakten, auf Studierenden- bzw. Studienanfängerkohorten bezogene Abbrecherquoten unerlässlich.
- Neben der fehlenden Information über das Ausmaß von Studienabbruch, Studienerfolg und Wechselgeschehen ist die Frage, in welchem Maß Studiengänge durch Wechsler in höhere Fachsemester nachgefragt werden, nicht einwandfrei zu beantworten. Die entsprechenden Informationen sind aber für die Planung des Studienangebots (inhaltliche Ausrichtung und Orientierung an den Ausbildungsbedürfnissen der Studierenden mit unterschiedlichen fachlichen Spezialisierungen und Hintergründen) unverzichtbar.

Die amtliche Hochschulstatistik sammelt semesterweise Daten zu Immatrikulation, Rückmeldung, Exmatrikulation, Studienfachwechsel, etc. Diese Daten werden jedoch nicht im Zeitablauf verknüpft (Arnold 1997, S. 339). „Man kann daher nichts über die individuellen Verläufe sagen, also nicht feststellen, was aus einem einzelnen Studierenden im Folgejahr und später geworden ist. Damit ist es zum Beispiel nicht möglich, anhand der Amtlichen Statistik konkret zu sagen: Die Studienanfänger etwa aus dem Studienjahr 1985 waren zu  $x$  Prozent erfolgreich, haben zu  $y$  Prozent Fach oder Hochschule gewechselt und zu  $z$  Prozent ihr Studium abgebrochen“ (a.a.O.). Es ist unter diesen Bedingungen lediglich möglich, die Bestände immatrikulierter Studierender in zwei aufeinander folgenden Zeiträumen zu vergleichen. „Wir können beispielsweise feststellen, dass sich im Winterse-



mester 1995/96 insgesamt 100 Studierende im Fach Physik immatrikuliert haben und dass im folgenden Wintersemester 80 Studierende im 3. Fachsemester eingeschrieben waren. Ob die fehlenden 20 Studierenden ihr Studium abgebrochen haben, an eine andere Universität gegangen sind oder das Fach gewechselt haben, ist nicht bekannt. Es ist nicht einmal bekannt, ob tatsächlich 20 Studierende das Fach Physik verlassen haben, oder ob es vielleicht 30 waren und 10 andere dazu gekommen sind“ (Arnold 1997, S. 339). Voraussetzung für die Ermittlung belastbarer Quoten hinsichtlich des Studienabbruchs und der erfolgreich abgeschlossenen Examen besteht im Vorliegen einer Studienverlaufsstatistik, die die Studienbiographien auf der Ebene der Personen zum Ausgangspunkt nimmt.

Eine relevante empirische Frage, die sich mit der Anwendung von lehrbezogenen Leistungsindikatoren zur Mittelvergabe in Hochschulen beschäftigen, besteht daher in derjenigen nach etwaigen Einschränkungen der Aussagefähigkeit der gewählten Indikatoren und der Möglichkeit von Fehlschlüssen auf Grund einer beeinträchtigten Validität durch verzerrte (Kenn-)Zahlen und Quoten. Die Validität der hochschulstatistischen Daten bzw. Indikatoren bemisst sich nach vorliegendem Verständnis an dem Maße in dem die entsprechenden Messungen verzerrungsfrei die angezielten Sachverhalte (bspw. Absolventen- oder Studienabbrucherquoten) wiedergeben.<sup>16</sup>

Einen Beitrag zur Klärung der Frage nach der Validität der aus der amtlichen Hochschulstatistik errechneten Abbrecher- und Absolventenquoten sollen die Sekundäranalysen der Daten der Hochschulstatistik der Universität Potsdam leisten, die unten vorgestellt werden. Dort werden die Ergebnisse eines Pilotprojekts zur Analyse von Studienverläufen mittels der Daten der Hochschulstatistik beschrieben.

Auf den Ergebnissen der Datenanalysen aufbauend, ist es darüber hinaus Ziel der vorliegenden Arbeit, auch Vorschläge für die zukünftige Weiterentwicklung der Beurteilungsverfahren von Hochschulleistungen zu unterbreiten. Als ei-

---

<sup>16</sup>In Kapitel 6 wird das Konzept der Testgütekriterien wie Validität und Reliabilität, insbesondere im Kontext der Qualitätsbeurteilung von Hochschulen näher erläutert.

nes der wichtigen Argumente für eine sorgfältige methodologische Prüfung dieser Verfahren und der dementsprechend erhobenen Daten zur Leistungsbeurteilung, wurden oben Gerechtigkeitsaspekte genannt. Ebenfalls aus Gerechtigkeits-erwägungen ist jedoch durchaus einleuchtend, eine leistungsbezogene Beurteilung von wissenschaftlichem Personal einzuführen (z.B. Schimank 2006). Eine solche muss jedoch über die quantitativen Leistungsindikatoren hinaus eine qualitätsorientierte Beschreibung der individuellen Leistungen in die entsprechenden Beurteilungen einbeziehen. Nur so ist die Akzeptanz bei den Betroffenen sicherzustellen, die ihrerseits die Voraussetzung ist, für eine sachgerechte und von opportunistischen Anpassungseffekten freie Leistungserbringung. Zukünftig stellt daher die Einbeziehung qualitativer Aspekte in die Datenerhebung zur Leistungsbeschreibung der Hochschulen eine Herausforderung für die Methodenentwicklung in der noch vergleichsweise jungen wettbewerbsorientierten Hochschulsteuerung dar.

Eine Möglichkeit, die verschiedenen Qualitätsvorstellungen der beteiligten Akteure zu vereinbaren, besteht in der Formulierung von Zielvereinbarungen (bspw. zwischen Land und Hochschulen). In Zielvereinbarungen, in denen Aktionspläne hinsichtlich zu erreichender Ziele festgeschrieben werden, können Zielerreichungskriterien fixiert werden, die Raum für das Verfolgen verschiedener Strategien zum Erreichen unterschiedlicher Ziele lassen. So kann eine Erweiterung der rein quantitativen Merkmale der Leistungen um qualitative Aspekte erreicht werden (vgl. z.B. Müller 2004). Eine Kontrolle der Zielerreichung ist in den entsprechenden Verfahren allerdings mit einem höheren methodischen Aufwand, bspw. durch die Anwendung von Evaluationsverfahren, in denen qualitätsbezogene Leistungsbeschreibungen von Hochschulen oder Hochschuleinheiten vorgenommen werden, verbunden. Schließlich stellt sich die Frage nach der tatsächlichen Steuerungs- und Anreizwirkung leistungsorientierter Mittelvergabe. Eines der wichtigsten Probleme, das im Zusammenhang mit der Anwendung von indikatorgestützten Steuerungsmodellen gesehen wird, besteht in der Vielgestaltigkeit und potenziellen Widersprüchlichkeit von Qualitätsvorstellungen, die mit den Leistungen der Hochschulen verbunden werden können. Indikatorgestützte Steuerungsmodelle müssen Leistungen zu Vergleichszwecken quantifizieren und

laufen dabei Gefahr, relevante qualitative Informationen und Leistungsparameter in nicht hinreichendem Maße zu berücksichtigen (Gutjahr-Löser 2000, S. 123; Pasternack 2004, S. 107 f.). Gleichwohl zeichnet es sich ab, dass insbesondere die indikatorgesteuerte Mittelvergabe kombiniert mit dem Instrument der Zielvereinbarung „zunehmend das traditionelle Zuweisungs- und Budgetbemessungsverfahren, das kameralistische Haushaltsführung mit der Fortschreibung von Budgets verbindet“ (Ziegele 2001) ablöst. Inwieweit diese grundlegenden Änderungen positive Reformimpulse setzen können oder aber im wesentlichen nicht-intendierte Effekte hervorrufen, wird durch zukünftige Forschungen untersucht werden müssen.

### **5.2.3 Analyse von Studienbiographien mit Daten der Hochschulstatistik**

Nachfolgend werden Aspekte der oben diskutierten Kennzahlen daraufhin untersucht, ob diese geeignet sind, die durch sie beschriebenen Sachverhalte tatsächlich adäquat darzustellen. Insbesondere die Studienabbruchquote soll hier zum Gegenstand der Analysen gemacht werden. Dieser Parameter steht im Mittelpunkt der öffentlichen Diskussion um die Leistungs(un)fähigkeit der Hochschulen und wird als Beleg für ihre Ineffizienz gesehen (Schröder-Gronostay 2003). Unbestritten ist die Tatsache, dass vorzeitiger Studienabbruch ein wichtiges Problem des Hochschulalltags ist, auch wenn in der Diskussion vielfach unerwähnt bleibt, dass dieser sich vielfach auch auf rationale Entscheidungen zurückführen lässt, wie bspw. die studentische Erkenntnis, eine falsche Ausbildungsentscheidung getroffen zu haben und diese (frühzeitig) korrigieren zu wollen (vgl. z.B. Heublein u. a. 2003). In solchen Fällen ist ein schneller Studienabbruch bzw. Hochschul- oder Fachwechsel aus der individuellen, studienbiographischen Perspektive vermutlich die vernünftigste Entscheidung.

Gleichwohl ist mit dieser persönlichen Entscheidung aus der Sicht der „hochschulstatistischen Bilanz“ ein wichtiges Problem verbunden. Dieses besteht darin, dass die einzelnen Ereignisse des Studienverlaufs in der Hochschulstatistik unverbunden als Einzeldaten aufgenommen werden (s.o.). Das Verlassen eines

Studiengangs durch Studierende taucht entsprechend als Information auf, die als Schwund, also als Misserfolg des „verlassenen“ Studienfachs bzw. Studiengangs interpretiert werden muss. Hochschul- und Fachwechsel (innerhalb einer Hochschule) bleiben unberücksichtigt, auch wenn in dem neuen Studiengang in den jeweils gewechselt wurde, das Studium erfolgreich abgeschlossen wird. Durch die Schwundquote eines Studienfachs oder Studiengangs, die die weiteren ausbildungsbiographischen Schritte und Entscheidungen der betroffenen Studierenden nicht berücksichtigt, wird das Ausmaß des Studienabbruchs in diesem Fach dementsprechend überschätzt. Insbesondere wenn dieser Effekt auf Fachwechsel innerhalb einer Hochschule zurückgeht, ist dies für diese misslich. In diesem Fall schlägt der „Ausbildungsmisserfolg“ des verlassenen Fachs in der entsprechenden, öffentlich wahrgenommenen Studienabbruchquote zu Buche. Ein etwaiger Ausbildungserfolg in einem neu gewählten Fach der gleichen Hochschule und mit hin ihr Ausbildungserfolg als Institution insgesamt findet dabei zumeist nicht in ebenso deutlicher Weise Eingang in die Diskussion bzw. Bilanzierung. Aus diesem Grund ist es für Hochschulen von großer Bedeutung, mehr über das Ausmaß des Fachwechsels innerhalb der von ihr angebotenen Studiengänge in Erfahrung zu bringen. Entsprechende Informationen können helfen, die Debatte um die Schwundquote, die mit der Studienabbruchquote üblicherweise gleichgesetzt wird, zu versachlichen. Die folgenden Reanalysen der Hochschulstatistik der Universität Potsdam stellen ein entsprechendes Instrumentarium und die unter seiner Anwendung gewonnenen Ergebnisse zur Diskussion.

Die oben beschriebenen Indikatoren beruhen auf Schätzverfahren, die wie alle Schätzungen mit Schätzunsicherheiten und möglichen Verzerrungen der Schätzergebnisse verbunden sind. Im vorliegenden Fall besteht die Gefahr der fehlerhaften Zuordnung von Studienfällen in die verschiedenen einbezogenen Gruppen. Dadurch würde bspw. eine errechnete Absolventenquote möglicherweise durch zugewanderte Studierende, die in höhere Fachsemester eines Studienganges gewechselt haben und entsprechend nicht der „virtuellen“ Studienanfängerkohorte zugeordnet werden können sowie durch die im Allgemeinen individuell sehr unterschiedlich lange Studiendauer unterschätzt.

Gleichzeitig werden Quoten zum Studienabbruch auf der Ebene von Fächern durch die nicht berücksichtigte Information über den Fachwechsel (abwandernde Studierende) innerhalb der Hochschule möglicherweise zum Nachteil der beurteilten bzw. betrachteten Fächer überschätzt.

Belastbare Quoten lassen sich nur dadurch errechnen, dass die Studienbiographien der einzelnen Studienfälle betrachtet werden und die relevanten Ereignisse (Abbruch, Examen, Fachwechsel, etc.) in den einzelnen Studienverläufen berücksichtigt und im Aggregat verschiedener Gruppen von Studierenden quantifiziert werden. Aggregationen auf verschiedenen Ebenen (Hochschulen, Studienfächer, Studiengänge, etc.) können dabei für die unterschiedlichen Analyseziele vorgenommen und ein entsprechend fein gegliedertes Bild über die relevanten Leistungskennzahlen gezeichnet werden.

Studienverlaufsstatistiken ermöglichen es, auf der Basis der einzelnen Studienfälle und deren Studienbiographien während der Immatrikulationszeit den Fach- und Hochschulwechsel sowie Studienabbruch und Studienerfolg verschiedener Studierendekohorten zu quantifizieren. Sie sind für den Kontext der vorliegenden Arbeit insofern relevant, als sie die kontroverse Diskussion um die auf Abbruch- und Absolventenquoten bezogenen Leistungsindikatoren bereichern.

### 5.2.3.1 Datenschutzrechtliche Regelungen

Trotz dieses großen Potenzials wird auf den Einsatz von Studienverlaufsstatistiken auf Bundesebene derzeit weitgehend verzichtet. Der Grund besteht in datenschutzrechtlichen Regelungen, deren rechtsgeschichtliche Wurzeln in der Bewertung der Volkszählung der 1980er Jahre und dem daraus folgenden Volkszählungsurteil zu finden sind (Arnold 1997).

Die kritische Bewertung von Studienverlaufsstatistiken ist in der notwendigen Verknüpfung personenbezogener Daten über lange Zeiträume begründet, die ganze Lebens- bzw. Bildungswege und -entscheidungen nachzuzeichnen erlauben und missbräuchlich für verschiedene Interessen genutzt werden könnten, die nicht mit dem ursprünglichen Erhebungszweck der Daten übereinstimmen (a.a.O.). Die Erfassung und Speicherung personenbezogener Daten für Zwecke

der amtlichen (Hochschul-)Statistik unterliegt einem bestimmten Verwendungszweck. Daten dürfen nur zu dem Zweck, zu dem sie erhoben wurden, verwendet werden. Eine Analyse von Studienverläufen setzt deshalb voraus, dass vor der Einbeziehung der benötigten Daten in die entsprechenden Analysen geprüft wird, ob der Auswertungszweck mit dem ursprünglichen Erhebungszweck übereinstimmt bzw., ob Voraussetzungen dafür gegeben sind, dass die Daten außerhalb dieses ursprünglichen Verwendungszwecks verarbeitet werden können. Dies trifft bspw. zu, wenn aufgrund kleiner Gruppengrößen in Studiengängen/Studienkohorten ein Deanonymisierungsrisiko besteht. In solchen Fällen dürfen erhobene personengebundene Daten nicht an Dritte weitergegeben werden, sofern dies nicht durch eine gesetzliche Vorschrift erlaubt wird oder die Betroffenen ihr Einverständnis erklärt haben.

In der jüngsten Zeit wird jedoch auf Grund des steigenden Informationsbedürfnisses und Legitimationsdrucks (vgl. Kapitel 2) dem Hochschul- und Fachbereichsleitungen ausgesetzt sind, der Verzicht auf Studienverlaufsstatistiken neu bewertet und in Richtung eines Verzichts auf den Verzicht diskutiert. Studienverlaufsstatistiken sollen wieder für die Hochschulplanung und -entwicklung nutzbar gemacht werden können. Insbesondere die Studienfächer würden von einer regelmäßig vorgelegten, verlässlichen Information zur Studienbiographie und zum Verbleib der Studienanfänger profitieren, indem sie besser informiert durch Anpassungen des Studienangebots auf Studienabbruch und Fach- bzw. Hochschulwechsel reagieren können. Mit dem Aufbau einer zentralen Studienverlaufsstatistik sowie einer entsprechenden Anwendung auf der Ebene der Fachbereiche oder Hochschulen ist also zukünftig zu rechnen. Auf der Ebene der einzelnen Hochschulen und zum Zweck der Hochschulplanung ist die verlaufsbezogene Betrachtung von Statistiken ohnedies statthaft.

Der allgemeine diesbezügliche Politikwechsel drückt sich sowohl in den Empfehlungen des Ausschusses für die Hochschulstatistik an den Deutschen Bundestag für den Zeitraum 1.6.2000-31.5.2004 (Bundestag 2005; Bundestagsdrucksache 15/5400, Seite 5) sowie in verschiedenen Pilotprojekten zum Aufbau einer effektiven Studienverlaufsstatistik aus. So wurden - gefördert durch Mittel des BMBF -

von der HIS GmbH Analysen zum Studienverlauf ausländischer Studierender an vier ausgesuchten Hochschulen durchgeführt.

Aus Mitteln des DAAD wurde die Anwendung der in diesem Zusammenhang erprobten Analyseinstrumente an weiteren Hochschulen ermöglicht, so auch an der Universität Potsdam. Die Ergebnisse der entsprechenden Analysen werden unten berichtet.<sup>17</sup> Nachfolgend soll zunächst kurz auf die Konzeption der Studienverlaufsstatistik an der Universität Potsdam eingegangen werden.<sup>18</sup>

### 5.2.3.2 Studienverlaufsstatistiken an der Universität Potsdam

Im Rahmen eines vom DAAD geförderten Forschungsprojekts konnte im Wintersemester 2005/06 der Aufbau der Studienverlaufsstatistik an der Universität Potsdam realisiert werden.<sup>19</sup> Mit Unterstützung der zuständigen Abteilungen der Universitätsverwaltung wurden Daten der Hochschulstatistik (zum belegten Studiengang bei Studienbeginn, Studiengangwechseln im Studienverlauf, dem Verlassen der Hochschule und den dafür ursächlichen Gründen) so aufbereitet, dass

---

<sup>17</sup>Bei diesen Analysen wurde sicher gestellt, dass sie einzig zum Zweck der Ermittlung von Abbruch- und Studienerfolgsquoten auf der Ebene der Hochschule und der Fächer durchgeführt wurden. Eine Verwendung der personenbezogenen Informationen für andere Zwecke wird schon dadurch ausgeschlossen, dass andere als die genannten, sich unmittelbar auf das Studienabbruch-/Studienwechselgeschehen beziehenden Informationen nicht abgefragt werden. Auch werden die personenbezogenen Daten nicht an Dritte (bspw. Studiengangsverantwortliche in den Fächern) weitergegeben. Die Fächer werden nur über die Auswertungsergebnisse hinsichtlich der Abbruch- und Absolventenquote informiert, nicht jedoch über die personenbezogenen Daten zum Studienverlauf von Studierenden.

<sup>18</sup>Diese wurde in Absprache mit dem Datenschutzbeauftragten der Universität durchgeführt.

<sup>19</sup>Gemäß dem Anliegen des DAAD wurde die Förderung eingesetzt, um insbesondere die Studienverläufe *internationaler* Studierender zu analysieren. Für die vorliegende Arbeit wurde das entwickelte Instrumentarium für die Betrachtung der Studienbiographien deutscher Studierender angewendet. Langfristig ist eine vergleichende Analyse der Studienabbruch- und Absolventenquoten unterschiedlicher Studierendengruppen angestrebt, bspw. um verschiedene, durch Gruppenzugehörigkeiten bedingte Dispositionen zu ermitteln. Aus diesen kann spezifischer Förderungs- bzw. Steuerungsbedarf abgeleitet werden, so z.B. zur Frage nach Risikogruppen hinsichtlich des Studienabbruchs.

die Studienfälle unter den genannten Gesichtspunkten auf der Ebene von Studienanfängerkohorten aggregiert und verglichen werden konnten.

Durch dieses Vorgehen wurde es ermöglicht, dass nicht die mit den oben genannten Unsicherheiten belasteten statistischen Schätzverfahren zur Analyse von Abbruchsquoten benutzt werden müssen, sondern reale Studienverläufe auf der Individualebene ausgewertet werden konnten.<sup>20</sup> Für Auswertungen, die sich auf die Frage der interessierenden Quoten beziehen, wurden die Daten auf verschiedenen Ebenen (Studiengänge, Fächer) aggregiert. So kann bspw. für die Ebene der Studiengänge eine belastbare Quote zur Beschreibung des tatsächlichen Ausmaßes des Studienabbruchs sowie des Zugangs in höhere Fachsemester angegeben werden.

In dem veränderten Auswertungsmodus wurde es möglich, Informationen, die sich auf die Studienbiographie von Studierenden beziehen, in einer Studienverlaufsperspektive zu analysieren. So können Aussagen darüber getroffen werden, in welchem Ausmaß über die Semester hinweg, Studierende einer Studieneingangskohorte das Studium aufgegeben bzw. erfolgreich abgeschlossen haben. Dafür wurde - ähnlich dem üblichen Verfahren der Hochschulstatistik - eine Datei mit den relevanten Daten mit Studienfallbezug aufgesetzt. Die notwendigerweise in die Auswertungen einzubeziehenden Daten beschränkten sich auf den Studiengang, in dem der bzw. die Studierende in den einzelnen Fachsemestern eingeschrieben war/ist sowie die semesterweise eingeholte Information über den Rückmeldestatus (zurückgemeldet, exmatrikuliert, etc.) und schließlich - bei Verlassen der Universität - eine Information über den Grund der Exmatrikulation (Abbruchs- bzw. Prüfungsdaten). Die entwickelte Datendatei enthielt die entsprechenden Daten aller Studienfälle, die an der Universität Potsdam seit ihrer Gründung im Jahre 1991 studiert haben. Dabei war die Dauer des Studienaufenthalts unerheblich, es wurden bspw. auch Austauschstudierende, die sich ohne die Absicht, das Studium an der untersuchten Hochschule abzuschließen, für ein oder zwei Semester

---

<sup>20</sup>Gleichwohl wurden die notwendigen Anonymisierungen der Daten vorgenommen, so dass nicht auf individuelle Identitäten der Studierenden zurück geschlossen werden konnte.



eingeschrieben hatten, in die Ausgangsdatendatei einbezogen. Gruppierungen und Sortierungen der Daten wurden später je für spezifische Analysezwecke vorgenommen.

Die Ausgangsdatendatei enthielt 49.601 Datensätze von Studierenden, die im Laufe des Bestehens der Universität Potsdam seit 1991 bis zum Wintersemester 2005/06 für mindestens ein Semester an ihr eingeschrieben waren.<sup>21</sup> Das Instrumentarium zur Abfrage der relevanten Daten aus den Datenbanken der Hochschulstatistik war für die Datenbanksoftware der HIS-GmbH (HIS-SOS) optimiert.<sup>22</sup> Diese wird an der Universität Potsdam seit dem Jahr 1999 flächendeckend eingesetzt. Aus diesem Grund konnten Studienanfängerkohorten, die das Studium vor diesem Zeitpunkt an der Universität Potsdam aufgenommen hatten, in einigen Auswertungen nicht berücksichtigt werden. Die Auswertungen zu den Studienverläufen beziehen sich auf Studienanfängerkohorten aus drei aufeinander folgenden Semestern: Sommersemester 1999 bis Sommersemester 2000. Spätere Studienanfängerkohorten bleiben ebenfalls weitgehend unberücksichtigt, weil hier der Zeitpunkt der Studienaufnahme nur kurz zurückliegt, die Studienverläufe entsprechend noch nicht weit fortgeschritten oder gar abgeschlossen sind. Die Betrachtung der drei genannten Studienverläufe stellt eine Betrachtung über eine hinreichende Zahl von 12-14 Semestern sicher. Tabelle 5.1 weist die in den genannten Studienanfängerkohorten erzielten Fallzahlen aus. Wie bereits erwähnt, weist die Information zum Schwund in den Fächern lediglich alle diejenigen Studierenden aus, die das jeweils untersuchte Fach nach einem bestimmten Semester verlassen haben. Auf der *Ebene der betreffenden Fächer* ist dieser Vorgang zunächst ausschließlich als Ausbildungsmisserfolg zu werten, die Studierenden haben das Fach verlassen und haben sich gegen ein erfolgreich fortgesetztes Studium ent-

---

<sup>21</sup>Der Betrachtungszeitraum endet mit dem Wintersemester 2005/06. Um die folgenden Semester in die Betrachtungen einzubeziehen, müssen erneute Ausgangsdatendateien aus der Hochschulstatistik ausgelesen werden.

<sup>22</sup>Für die Datenauswertungen kam ein von der HIS-GmbH entwickeltes Manual „zur Analyse des Studienverlaufs ausländischer Studierender“ zum Einsatz (Heublein und Sommer o. J.).

schieden. Sofern aber in ein anderes Fach der gleichen Universität gewechselt und hier das Studium erfolgreich beendet wird, kann die entsprechende Studienbiographie auf der *Ebene der Universität* durchaus als Ausbildungserfolg verbucht werden.

Studienanfängerkohorte	Fallzahl (N)
Sommersemester 1999	889
Wintersemester 1999/00	2987
Sommersemester 2000	892

Tabelle 5.1: Fallzahlen in den einbezogenen Studienanfängerkohorten

Der Fachwechsel innerhalb einer Hochschule ist noch vergleichsweise leicht zu quantifizieren und zu interpretieren. Fachwechsel, der über die Grenzen einer Hochschule hinweg verläuft, ist dagegen nicht notwendigerweise als solcher erfassbar, weil einerseits die nötigen Informationen nicht vorliegen<sup>23</sup> und andererseits die betreffenden Studierenden das Hochschulsystem möglicherweise zunächst ganz verlassen und bspw. eine Berufsausbildung absolvieren, bevor sie sich erneut für ein Studium entscheiden. Entsprechend gebrochene Ausbildungsbiographien sind auf der Ebene der Datenerhebung einer einzelnen Hochschule nicht nachvollziehbar.<sup>24</sup> Es wird aber im Falle eines an der zweiten Hochschule erfolgreich abgeschlossenen Studiums die hochschulbezogene Studienabbruchquote verfälscht und darüber hinaus der Beitrag der „entsendenden“ Hochschule zum erfolgreich an der „empfangenden“ Hochschule unterschlagen. Die Abbruchquote wird in die-

---

<sup>23</sup>Die relevanten hochschulstatistischen Daten über die Einschreibung in einen Studiengang werden jeweils von der diesen anbietenden Hochschule erfasst. Die Studienbiographie insgesamt, also einschließlich der Wechsel in einen Studiengang an einer anderen Hochschule, ist dagegen bislang nicht Gegenstand der Hochschulstatistik.

<sup>24</sup>Da für das Hochschulsystem als Ganzes auf die Durchführung von Studienverlaufsstatistiken derzeit verzichtet wird (s.o.), lässt sich der Studienerfolg von Studierenden, die zwischen Hochschulen wechseln, auf dieser Ebene nicht nachvollziehen.

sem Fall überschätzt, weil die betreffenden Studierenden zwar für die zunächst verlassene Hochschule Schwund darstellen und als solcher in der betreffenden Hochschulstatistik in Erscheinung treten, das Hochschulsystem aber mit dem angestrebten Ausbildungsziel von einer anderen Hochschule aus verlassen. Wechsel zwischen den Fächern bzw. Studiengängen einer Hochschule lässt sich dagegen mit den Mitteln der Studienverlaufsstatistik analysieren und quantifizieren. Anhand des Beispiels der Volkswirtschaftslehre (VWL) soll im folgenden auf die quantitative Bedeutung des Fachwechsels für die Interpretation von Schwundquoten insgesamt hingewiesen werden. Das Fach wurde für die Analysen ausgewählt, weil sich an diesem Beispiel die Auswirkungen für die an den Wechselbewegungen beteiligten (entsendenden und empfangenden) Fächer prägnant darstellen lassen.

Mittels einer Serie von Kreuztabellen wurde der Fachwechsel Studierender der VWL über die jeweils aufeinander folgenden Semester nachgezeichnet. Abbildung 5.1 zeigt schematisch und unter Verwendung fiktiver Zahlen die Vorgehensweise bei der sequenziellen Analyse der Kreuztabellen. In den Spalten der Tabellen wurden die Hauptfächer der Studierenden im ersten (bzw. zweiten, dritten, usw.) Semester und in den Zeilen die Fächer, die im zweiten Semester (bzw. dritten, vierten usw.) Semester an der Universität Potsdam belegt wurden, ausgewiesen. Dadurch versammelt sich auf den Diagonalen der Tabellen die Zahl derjenigen Studierenden, die in jeweils zwei aufeinander folgenden Semestern das gleiche Hauptfach belegt, bzw. ihre Hauptfächer nicht gewechselt haben. Jenseits der Diagonalen finden sich die Fachwechsler wieder.

In die Auswertungen wurden die Studierenden einbezogen, die im Wintersemester 1999/00 ein Studium der VWL an der Universität aufgenommen haben. Die Datenbasis umfasst alle Studienfälle, auf die diese Bedingung zutrifft und dementsprechend bspw. auch Austauschstudierende, die von vornherein beabsichtigen, die Universität nach einem Semester wieder zu verlassen. Durch diese und andere Studierendengruppen ist der Schwund bereits nach dem ersten Semester vergleichsweise hoch. Die hier herangezogene Stichprobe wurde jedoch bewusst nicht um derartige Fälle bereinigt, um ein möglichst umfassendes Bild des Abbruch- und Wechselgeschehens zu zeichnen. Über die Semester hinweg lässt

		Sommersemester 2000					
		VWL	BWL	Soziologie	Verwaltungswiss.	Rechtswiss.	...
Wintersemester 2000/2001	exmatrikuliert	5	10	0	5	0	10
	VWL	45	0	2	1	1	0
	BWL	10	82	1	2	0	0
	Soziologie	2	0	57	0	0	0
	Verwaltungswissenschaft	0	0	0	34	0	0
	Rechtswissenschaft	0	0	0	0	66	0
	...	0	0	0	0	0	45
	$\Sigma$	62	92	60	42	67	55

		Wintersemester 2000/2001					
		VWL	BWL	Soziologie	Verwaltungswiss.	Rechtswiss.	...
Sommersemester 2001	exmatrikuliert	5	2	7	4	6	5
	VWL	35	0	0	0	0	0
	BWL	5	80	0	0	0	0
	Soziologie	0	0	50	0	0	0
	Verwaltungswissenschaft	0	0	0	30	0	0
	Rechtswissenschaft	0	0	0	0	60	0
	...	0	0	0	0	0	40
	$\Sigma$	45	82	57	34	66	45

Abbildung 5.1: Schematische Darstellung sequenzieller Kreuztabellen

sich jeweils ein Anteil Studierender feststellen, die zwar das Fach VWL verlassen, nicht aber die Universität Potsdam, sondern ihr Studium in einem anderen Fach fortsetzten. Dieser Anteil variiert über die Semester, beträgt im Wechsel vom ersten ins zweite Semester moderate fünf Prozent, um auf einen Spitzenwert von nahezu 50 Prozent zwischen dem vierten und fünften Semester anzusteigen und danach wieder auf geringere Werte abzufallen. Das bedeutet, dass im Wechsel vom vierten zum fünften Semester die Hälfte des fachbezogenen Schwundes aus solchen Studierenden gespeist wird, die in einem anderen Fach der Universität ihr Studium fortsetzen. Dies ist im Fall der VWL umso bemerkenswerter, als der

Großteil der Fachwechsler sich der Betriebswirtschaftslehre (BWL) an der Universität Potsdam zuwendet (vgl. Tabelle 5.2). Zur Erläuterung der Bedeutung der in Tabelle 5.2 präsentierten Ergebnisse dient das folgende Lesebeispiel: Von den ursprünglich im ersten Fachsemester eingeschriebenen 210 Studierenden der VWL verlassen 36 nach diesem ersten Semester das Fach. Von diesen 36 Studierenden, die den fachbezogenen Schwund darstellen, wechseln zwei (oder 5,5 Prozent) in ein anderes Fach der Universität Potsdam. In diesem Fall wechselt keiner dieser beiden Studierenden zur Betriebswirtschaftslehre. Keiner der Studierenden legt ein Examen ab, dementsprechend setzen 174 Studierende ihr Studium fort. Diese und sieben zugewanderte Studierende bilden die Gruppe der im folgenden Semester eingeschriebenen Studierenden im Fach VWL (181 Studierende).<sup>25</sup> In den folgenden Semestern lässt sich deutlich erkennen, dass Studierende der VWL bevorzugt in Richtung der Betriebswirtschaftslehre aus ihrem ursprünglich gewählten Studienfach abwandern. Die Anteile betragen bis zu 80 Prozent, wie in der Abfolge des vierten und fünften Fachsemesters.

Zwar muss betont werden, dass die Zahlen (und insbesondere relativen Zahlen) bzw. die ausgewiesenen Anteile in ihrer Höhe aufgrund der insgesamt kleinen Gruppengröße nicht überschätzt werden dürfen. Der gezeigte Zusammenhang des „massenhaften“ Wechsels von VWL-Studierenden in die Betriebswirtschaftslehre zeigt sich aber auch, wenn andere Stichproben zugrunde gelegt werden (bspw. andere Studienanfängerkohorten). Während also die Volkswirtschaftslehre als entsendendes Fach charakterisiert werden kann, kommt der BWL der Status eines aufnehmenden Faches zu, mit den jeweils gravierenden Folgen, auf die unten weiter eingegangen wird.

Der Studienverlauf der Studierenden im Fach VWL wurde hier unter dem Gesichtspunkt des Fachwechsels betrachtet. Aus diesem Grund wurde auch nur ein Ausschnitt über sechs Fachsemester betrachtet. nach diesen sechs Semestern

---

<sup>25</sup>Bei den Studierenden, die bereits nach fünf bzw. sechs Fachsemestern als Absolventen ausgewiesen werden, handelt es sich vermutlich um Wechsler, die ehemals in ein höheres Fachsemester an der Universität Potsdam zugewandert sind.

	I	II	III		IV		V	VI	VII
	Studierende	Schwund	davon Wechsel		davon zu BWL		Absolventen	Fortsetzung des Studiums	Zugänge
Nach dem...	(n)	(n)	(n)	(%)	(n)	(%)	(n)	(n)	(n)
... 1. Fachsemester	210	36	2	5,5	0	0	0	174	7
... 2. Fachsemester	181	49	12	24,5	9	75,0	0	132	6
... 3. Fachsemester	138	43	10	23,3	6	60,0	0	95	3
... 4. Fachsemester	98	32	15	46,9	12	80,0	0	66	0
... 5. Fachsemester	66	21	8	38,1	3	37,5	1	45	1
... 6. Fachsemester	46	18	3	16,7	0	0	1	28	0

Tabelle 5.2: Fachwechsel und Fachschwund: das Beispiel VWL (Studienanfängerjahrgang 1999/2000)

nimmt das Wechselgeschehen sowohl auf der genannten Hauptrichtung zwischen den beiden wirtschaftswissenschaftlichen Fächern, als auch insgesamt stark ab. Offensichtlich ist die Studienfachwahl in höheren Semestern endgültig gefestigt, so dass Wechsel in andere Studienfächer nur noch sehr selten vorkommen. Auch der Schwund nimmt in höheren Semestern ab, während die Zahl der Absolventen zunimmt.

In der Diskussion um Fachwechsel und Studienabbruch wird vielfach argumentiert, dass diese ein Ergebnis einer zu wenig rationalen Studienfachwahl seien und zum Ausdruck bringen, dass die mit der ursprünglichen Studienfachwahl verbundenen Erwartungen vor den Eindrücken der tatsächlich vorgefundenen Studiensituation nicht bestehen. Dies sei auf eine schlechte Informierung der Studieninteressierten vor Studienbeginn genauso zurückzuführen, wie auf schlechte Studienbedingungen, mangelnde Studierfähigkeit u.ä. (vgl. z.B. Pohlenz und Tinsner 2004). Durch den vielfach auftretenden Studienabbruch und Fach- bzw. Hochschulwechsel sei die schlechte Qualität von Lehre und Studium belegt, da die

von ihnen betroffenen Studierenden nicht an das Fach bzw. die Universität gebunden werden konnten, durch schlechte Studienbedingungen vertrieben wurden und sich abgewendet haben. Nun lassen jedoch die Befunde zur Wechseltätigkeit zwischen den Fächern VWL und BWL auch andere Interpretationen zu. Diese legen nämlich die Vermutung nahe, dass die Studienwahl für die Volkswirtschaftslehre durchaus einer rationalen Entscheidung folgte, die bspw. durch die bestehenden Zulassungsrestriktionen in dem ursprünglich gewünschten Fach (BWL) beeinflusst worden sein können. Die Wahl der Volkswirtschaftslehre als Studienfach könnte in diesem Zusammenhang als das Ergebnis eines Kompromisses gesehen werden, mit dem die Erwartung verbunden wird, dass ein späterer Wechsel in ein höheres Fachsemester des eigentlich gewünschten Studienfachs Betriebswirtschaftslehre vergleichsweise leicht zu bewerkstelligen ist. Die Kompromisslösung bestünde darin, zunächst ein Studium in einem thematisch verwandten Studienfach aufzunehmen und so die befürchtete aufzubringende Wartezeit auf bzw. mehrfache Bewerbung für einen Studienplatz im gewünschten Studienfach zu vermeiden. Darüber hinaus sind auch Zusammenhänge denkbar, die sich auf die konkrete Studienplanung beziehen. So ist zu vermuten, dass bewusst zunächst Studienleistungen in der VWL absolviert werden, die durch die Studienordnung auch als Studienleistung für die Betriebswirtschaftslehre anerkannt werden. Im Falle einer gemeinsamen Politik der Leistungsbeurteilung und -anerkennung der beiden Fächer ließen sich Studienbiographien und Studienschwerpunkte über die Grenzen der Fächer hinweg planen. Der Schluss, dass entsprechende strategische Überlegungen hinsichtlich der Studienplanung eine Rolle spielen, wird durch die hohe Zahl Studierender nahe gelegt, die nach dem vierten Semester, also dem Zeitpunkt, für den nach den seinerzeit einschlägigen Diplomprüfungs- bzw. Studienordnungen die Zwischenprüfung vorgesehen war, einen Wechsel vornehmen. Unabhängig von der individuellen Bedeutung des Studienfachwechsels für die einzelnen Studierenden, hat dieser für die Fächer gravierende Folgen in zweifacher Hinsicht:

- Für die Studiengangspannung bedeutet der Wechsel zwischen den beiden Fächern, dass „außerplanmäßig“ Studierende verloren werden, bzw. aufge-

nommen werden müssen. Durch die zusätzlich in die Lehrveranstaltungen aufzunehmenden Studierenden verschlechtern sich die Betreuungsrelationen weiter. Darüber hinaus müssen vom aufnehmenden Fach mehr Studierende zu einem Examen geführt werden. Anhand der in Tabelle 5.2 präsentierten Fallzahlen kann zwar keineswegs die Rede davon sein, dass sich aufgrund der Wechsler die Studienbedingungen im aufnehmenden Fach prekär verschärfen. Betrachtet man aber die Zahlen nicht nur einer Studienanfängerkohorte, sondern diejenigen, die sich auf alle Studienfälle beziehen, werden (absolute) Zahlen erreicht, die das Phänomen des Wechsels zwischen den Fächern durchaus zu einer relevanten Planungsgröße werden lassen (vgl. Tabelle 5.3). Langfristig ist also durch eine seitens der Fächer wenig gesteuerte Wechseltätigkeit der Studierenden durchaus mit einer Verschlechterung der Studienbedingungen zu rechnen.

- Da die Zahl und insbesondere die Quote der Studienabbrecher auch für Legitimations- bzw. Beurteilungszwecke auf verschiedenen Ebenen genutzt wird, ist die üblicherweise unberücksichtigte Information über das Ausmaß des Fachwechsels ein gravierender Nachteil für Fächer, die - wie im vorliegenden Beispiel die Volkswirtschaftslehre - Studierende an andere Fächer abgeben. Dieser Sachverhalt wurde bereits geschildert. Zusätzlich verschlechtert sich aber in den landläufig herangezogenen Verfahren zur Berechnung von Absolventenquoten (die Zahl der Absolventen wird mit der Zahl der Studienanfänger in einem der Regelstudienzeit entsprechenden Zeitraum ins Verhältnis gesetzt) die entsprechende Leistung des *aufnehmenden* Fachs: Im Verhältnis zu den Absolventen steigt nachträglich die Zahl der Studierenden, die zur Berechnung der Absolventenquote zugrunde gelegt wird.

Zur Beurteilung der Leistungen dieser Fächer werden dementsprechend verzerrte Daten zugrunde gelegt, so lange die Information über den Wechsel zwischen Fächern nicht berücksichtigt wird. Auf diese Weise wird die Außendarstellung und die Legitimation der Leistungen, die in den betroffenen Fächern erbracht werden, erschwert. Dies kann unmittelbare Folgen für ihre Position im Wettbewerb um



Ressourcen haben. Insbesondere trifft dies zu, wenn sich die Leistungsbeurteilung auf die oben beschriebenen Indikatoren stützt. Um entsprechende Verzerrungen zu vermeiden, ist es nötig, studienverlaufsbezogene Informationen, wie die zum Fachwechsel innerhalb der Universität einzubeziehen. Ein noch größerer Informationsgewinn zur Frage der Erfolgsquoten und damit letztlich der Effizienz des Hochschulsystems, wäre durch den Aufbau einer bundesweiten Studienverlaufsstatistik zu erzielen, die es ermöglichen würde, Wechselbewegungen nicht nur zwischen den Fächern einer Hochschule, sondern über deren Grenzen zu anderen Hochschulen hinweg nachzuvollziehen.

	I	II	III	IV	V	VI	VII
	Studierende	Schwund	davon Wechsel	davon zu BWL	Absolventen	Fortsetzung des Studiums	Zugänge
Nach dem...	(n)	(n)	(n) (%)	(n) (%)	(n)	(n)	(n)
... WS 1999/00	1.365	440	40 9,1	22 55,0	3	922	47
... SS 2000	969	209	45 21,5	27 60,0	0	760	54
... WS 2000/01	814	199	33 16,6	24 72,7	1	614	17
... SS 2001	631	158	65 41,2	56 86,2	2	471	10
... WS 2001/02	481	141	26 18,4	18 69,2	2	338	3
... SS 2002	341	94	25 26,8	11 44,0	3	244	4
... WS 2002/03	248	66	9 13,6	5 55,6	3	182	3

Tabelle 5.3: Fachwechsel und Fachschwund: das Beispiel Volkswirtschaftslehre (alle Studienfälle)

Die vorgelegten Analysen beschränken sich auf das Beispiel der Beziehungen zwischen Volkswirtschaftslehre und Betriebswirtschaftslehre. Diese Fächer wurden mit Bedacht zur Illustration des Effekts, den der Fachwechsel auf die Berechnung von Quoten ausübt, ausgewählt. So sind die Wechselbewegungen bspw. zwischen den Fächern der Philosophischen Fakultät deutlich disparater und ent-

sprechend schwieriger zu interpretieren, während im Fall der wirtschaftswissenschaftlichen Fächer deutlich weniger in thematisch entferntere Fächer gewechselt wird. Schon in die unmittelbaren Nachbardisziplinen innerhalb der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Potsdam, wie die Verwaltungswissenschaft oder die Politikwissenschaft wird nur sehr begrenzt gewechselt.

Zukünftige Forschungen werden sich mit der Quantifizierung der Wechselbewegungen innerhalb der Hochschule(n) zu beschäftigen haben sowie mit der Frage, inwieweit diese sich auf die studienabbruch- und absolventenbezogenen Leistungsbilanzen der Fächer auswirken.

# Kapitel 6

## Kriterien für die Aussagekraft von Evaluationsergebnissen

In der Diskussion der Bedeutung von Evaluation für die Konzeption von Qualitätsentwicklungs- oder -sicherungsinstrumenten (vgl. Kap. 4) wurde bereits ausgeführt, dass aus vorliegender Sicht in der Verfügbarkeit von relevanten Informationen eine wesentliche Voraussetzung für eine nachhaltige entsprechende Hochschulsteuerung zu sehen ist. Informationen in diesem Sinne sind einerseits das Ergebnis von Datenerhebungen zu Leistungskennzahlen, wie sie im vorangegangenen Kapitel beschrieben wurden. Andererseits werden sie aber auch - und insbesondere - aus Befragungen von Studierenden zu deren Qualitätseinschätzungen über Lehre und Studium, gewonnen, also aus Lehrevaluation im gängigen Begriffsverständnis.<sup>1</sup> Die Relevanz der entsprechenden Informationen für ihre Nutzung in Steuerungsprozessen leitet sich zunächst aus inhaltlichen Aspekten, wie ihrer Neuheit, Detailliertheit, Generalisierbarkeit, Nützlichkeit, etc. ab (s.o.). Daneben bestehen aber auch methodologische Kriterien, die an Evaluationsdaten angelegt werden müssen, damit davon gesprochen werden kann, dass durch sie eine

---

<sup>1</sup>Zur begrifflichen Unschärfe bei der Gleichsetzung von Befragungen oder allgemeiner, von Umfrageforschung mit Evaluation vgl. die Ausführungen in Kap. 3 und insbesondere Kromrey (2001) sowie Kromrey (2001a).

---

relevante Aussage ermöglicht wird, die ihrerseits bspw. in der Durchführung einer Verbesserungsstrategie zur Hochschulorganisation, also in politisches Handeln mündet.

Diese Kriterien bestehen insbesondere in einer Datenqualität der erhobenen Daten, die einer Prüfung hinsichtlich zentraler Testgütekriterien Stand hält. Dabei sind in erster Linie die Validität und Reliabilität der erhobenen Daten zu nennen. Deren Einhaltung stellt aus vorliegender Sicht eine zentrale Voraussetzung für eine nachhaltige Nutzung von Evaluationsdaten für hochschulpolitische Entscheidungsprozesse dar. Bereits verschiedentlich wurde angedeutet, dass der wichtigste Grund für diese Festlegung in der Annahme besteht, dass die Nutzung von potenziell in ihrer Validität und Reliabilität bedrohten Daten für Entscheidungs- oder Steuerungsprozesse zu Fehleinschätzungen und Fehlsteuerungen führen kann, durch die u.U. auch Gerechtigkeitsaspekte berührt sind. Dies trifft zu, wenn bspw. Entscheidungen über Mittelallokationen auf eine invalide und unreliable Datenbasis und entsprechenden Fehlinterpretationen von Informationen aufgesetzt werden. Die Frage nach der Qualität erhobener Evaluationsdaten steht entsprechend am Anfang einer zu planenden Qualitätssicherungsmaßnahme.

In der Betonung dieses eher messtheoretischen Aspekts der Durchführung von Evaluationen spiegelt sich der oben diskutierte Einfluss verschiedener Evaluationsparadigmen auf die praktische Durchführung von Qualitätssicherungsmaßnahmen wider. So ist die Frage nach den genannten Testgütekriterien bei der Erhebung von Daten ganz unabhängig davon zu stellen, inwieweit mit dem auf diesen Daten basierten Evaluationsdesign ein eher rationalistischer oder naturalistischer Ansatz verfolgt wird. Sofern mittels der Erhebung von Daten Messungen (bspw. zu Einstellungsparametern von Befragten) beabsichtigt sind, ist an diese Daten die Erwartung zu richten, dass sie den zu messenden Gegenstand verzerrungsfrei abbilden. Entsprechend wird hier davon ausgegangen, dass die verschiedenen Evaluationsparadigmen, die oben diskutiert wurden (vgl. Kapitel 3) diesbezüglich in gleichberechtigter Weise nebeneinander existieren.

## 6.1 Testgütekriterien

Gütekriterien von Messungen werden in vielen Lehrbüchern der Sozialwissenschaften (z.B. Schnell u. a. 1999; Bortz und Döring 2002) sowie der Psychologie (z.B. Kubinger 1995; Kubinger und Jäger 2003; Tent und Stelzl 1993) bezüglich ihrer Bedeutung für die Konstruktion und Durchführung wissenschaftlicher Messungen und die Interpretation der entsprechenden Ergebnisse ausführlich beschrieben. Hier sollen sie daher nur kurz zusammenfassend dargestellt werden. Die Hauptkriterien für die Güte von Tests oder Messungen -Objektivität, Reliabilität und Validität- lassen sich in eine Vielzahl von Unterkriterien gliedern, die ihrerseits aus vorliegender Sicht für die Planung, Durchführung und Nutzung von Evaluationen und Evaluationsergebnissen von unterschiedlicher Bedeutung und Wichtigkeit sind.

Die im Kontext der vorliegenden Arbeit relevanten (Unter-)Kriterien werden im vorliegenden Kapitel beschrieben. Die Analysen, die unten dargestellt werden, beschäftigen sich zu einem wesentlichen Teil mit der Überprüfung von Daten der Lehrevaluation hinsichtlich der Frage, inwieweit die entsprechenden Erhebungen den Ansprüchen gerecht werden, die an sie mit Blick auf die genannten Testgütekriterien zu stellen sind. Sind die erhobenen Daten in der Lage, das mit ihrer Erhebung verbundene Ziel zu erreichen, nämlich eine sachgerechte Beurteilung der Qualität von Lehre und Studium zu ermöglichen oder ist von Limitationen der Aussagefähigkeit der Daten auszugehen, die ihrerseits auf Bedrohungen von Reliabilität und Validität zurückzuführen sind? An die entsprechenden Befunde kann sich die Diskussion der Frage anschließen, inwieweit die Datenerhebungen zur Evaluation von Lehre und Studium insgesamt anschlussfähiges, für die Hochschulssteuerung verwertbares Wissen liefern und welche Adjustierungen der Konzeption von entsprechenden Evaluationsdesigns ggf. vorgenommen werden müssen. Das Kriterium einer Basierung von Entscheidungen in der Hochschulplanung auf zuverlässige und gültige Daten, ist eine grundlegende Anforderung, die an diese Entscheidungen sowie die entsprechenden Entscheidungsprozesse zu stellen ist. Zu begründen ist dies mit der Akzeptanz, die Steuerungsentscheidungen

von den Hochschulmitgliedern insbesondere dann entgegengebracht wird, wenn diese Entscheidungen auf einer sachgerechten und transparenten Beurteilungsgrundlage basieren. Die Akzeptanz von Entscheidungen und Entscheidungsprozessen wird ihrerseits wiederum als Vorbedingung für die aktive Beteiligung der Hochschulmitglieder an Reformprozessen und damit als Voraussetzung für deren Erfolg angesehen.

## 6.2 Objektivität, Reliabilität, Validität

Den genannten Testgütekriterien ist gemein, dass sie Ableitungen aus den Annahmen der Klassischen Testtheorie darstellen (z.B. Bühner 2004, S. 21; Tent und Stelzl 1993, S. 43).<sup>2</sup> Deren Anwendung bei der Konstruktion von Mess- oder Datenerhebungsinstrumenten sowie der Interpretation von Messergebnissen ermöglicht es, diese sinnvoll zu deuten und hinsichtlich ihrer Güte und damit Aussagefähigkeit und Reichweite zu beurteilen (Schnell u. a. 1999, S. 143).<sup>3</sup> Die Klassische Testtheorie nimmt als grundlegendes Kennzeichen von Messungen deren Zusammensetzung aus einem wahren Wert und einem Messfehler an. Das Grundmodell der Klassischen Testtheorie postuliert, „dass ein realisierter Messwert aus der Summe eines wahren Wertes und einem Messfehler (resultiert; P.P.); Messfehler sind also Differenzen zwischen wahren Werten und beobachteten Werten. Ein wahrer Wert kann als Mittelwert einer großen Zahl *unabhängiger* Messungen des-

---

<sup>2</sup>Zum Zweck der vertieften Einführung in die Klassische Testtheorie wird von Schnell u. a. (1999), Tent und Stelzl (1993) sowie Bortz und Döring (2002) das Standardwerk von Lord und Novick (1968) genannt.

<sup>3</sup>Daraus ergibt sich die Einschränkung, im vorliegenden Zusammenhang Testtheorien hauptsächlich als Theorien über solche Tests aufzufassen, die Eigenschaften, Merkmale oder Zustände von Personen erfassen. Bei diesen Tests kann es sich bspw. auch um Fragen (oder formal formuliert: Konstrukte) handeln, die den beobachteten Personen mittels Fragebögen zur Beurteilung vorgelegt werden (Bühner 2004, S. 15). Das Verständnis von Testtheorien als Theorie über statistische Schlüsse von Stichprobendaten auf Merkmale einer Population wird in diesem Zusammenhang ausgeklammert (a.a.O.).

selben Objekts aufgefasst werden“ (Schnell u. a. 1999, S. 144, Kursivsetzung im Original). Darüber hinaus werden vier weitere grundlegende zentrale Axiome der Klassischen Testtheorie beschrieben:

- Der Mittelwert der Messfehler ist gleich Null. Der Mittelwert der wiederholt an einem Objekt durchgeführten Messungen repräsentiert demnach den wahren Wert.
- Die Messfehler und der wahre Wert sind unkorreliert, die Höhe des Messfehlers ist unabhängig von der Ausprägung des getesteten Merkmals.
- Die Messfehler zweier Messreihen sind unkorreliert, Messfehler sind daher unabhängig von der Ausprägung bspw. in anderen Persönlichkeitsmerkmalen der beobachteten Person.
- die Messfehler einer Messreihe sind mit den Messfehlern einer zweiten Messreihe unkorreliert, es gibt keine zeitliche Stabilität von Messfehlern (vgl. z.B. Bortz und Döring 2002, S. 193).

Durch diese Annahmen wird die Bedingung formuliert, dass es sich bei den vorliegenden Messfehlern um zufällige und nicht systematische Fehler handelt. Es wird der Tatsache Rechnung getragen, dass Testergebnisse (im Falle von Wiederholungsmessungen) über verschiedene Testzeitpunkte variieren, bspw. beeinflusst durch Lerneffekte und dass unsystematische innere (z.B. durch die Erhebungssituation bedingte) und äußere (z.B. auf die Motivation der Testpersonen bezogene) Einflüsse die Testergebnisse verzerren (Bühner 2004, S. 21).

Die Formulierung dieser Bedingungen stellt gleichzeitig den Ansatzpunkt für die Kritik an der Klassischen Testtheorie dar. So scheint es in der Praxis der empirischen Forschung unrealistisch zu sein, Einflüsse auf Test- oder Befragungsergebnisse grundsätzlich als unsystematisch auftretend anzunehmen.<sup>4</sup> „In

---

<sup>4</sup>Weitere Kritikpunkte beziehen sich auf die Stichprobenabhängigkeit und die in der Praxis vielfach nicht gegebene Bedingung der Normalverteilung von Stichprobendaten (z.B. Tent und Stelzl 1993, S. 52 ff.).

letzter Konsequenz ist damit sowohl die Annahme eines fehlenden Zusammenhangs zwischen wahren Wert und Messfehler zu bezweifeln, als auch die Konstanz des wahren Werts über verschiedene Messwiederholungen“ (Bühner 2004, S. 27). Die Klassische Testtheorie stellt eine reine Messfehlertheorie dar, in der die Unterschiedlichkeit, mit der Testergebnisse tatsächlich zu Stande kommen, nicht berücksichtigt wird (a.a.O.). Diese Frage ist Gegenstand der Probabilistischen Testtheorie (Bühner 2004, S. 35). Deren grundlegende Annahme besteht darin, dass die Wahrscheinlichkeit einer bestimmten Antwort befragter Personen auf die einzelnen Items eines Fragebogens von der jeweiligen Ausprägung einer bestimmten mit dem Item verbundenen latenten Merkmalsdimension abhängt (Bortz und Döring 2002, S. 192). Obwohl die Probabilistische Testtheorie als Beitrag zur Überwindung der Schwächen der Klassischen Testtheorie angesehen wird (Bühner 2004, S. 21), hat sich diese jedoch - nicht zuletzt auf Grund ihrer vergleichsweise einfachen Handhabbarkeit - in der Praxis bewährt und wird bei der Konstruktion des größten Teils von Tests angewendet (vgl. Rost 1999; zit. nach Bortz und Döring 2002, S. 211). Gründe sind u.a. darin zu sehen, dass Ergebnisse von Tests, die nach der Klassischen Testtheorie konstruiert wurden, in hohem Maße mit denen vergleichbar sind, die auf einem probabilistischen Testmodell basieren und das dementsprechend für einen vergleichsweise geringen zusätzlichen Nutzen durch deren Einsatz ein hoher zusätzlicher Aufwand in Kauf genommen werden muss (a.a.O., S. 211 f.).<sup>5</sup>

---

<sup>5</sup>Die Frage, welche Bedeutung der Probabilistischen Testtheorie im Verhältnis zur Klassischen Testtheorie beizumessen ist und welche Implikationen für ihren Einsatz bei der Konstruktion von Tests und Messverfahren sowie der Beurteilung von Testgütekriterien sich daraus ergeben, ist Gegenstand einer Auseinandersetzung verschiedener Positionen in der Psychologischen Diagnostik (vgl. z.B. Rost (1999) und die Replik von Kubinger (2002)). Da die in der vorliegenden Arbeit berücksichtigten Daten bzw. die Konstruktion der entsprechenden Datenerhebungsinstrumente auf den Annahmen der Klassischen Testtheorie beruht und deren (klassischen) Testgütekriterien ein relevanter Gegenstand der Analysen sind, wird die Probabilistische Testtheorie hier nicht weiter diskutiert.



Basierend auf den Axiomen der Klassischen Testtheorie lassen sich die drei (Haupt-)Testgütekriterien Objektivität, Reliabilität und Validität näher beschreiben.

**Objektivität:** Das erste der drei Testgütekriterien hat für die nachfolgend berichteten Analysen aus methodologischer Sicht eine eher nachgeordnete Bedeutung und wird daher nur am Rande beschrieben.<sup>6</sup> Dieses Vorgehen ist unmittelbar in der Definition der Objektivität von Tests begründet. Nach dieser sind Tests dann objektiv, wenn die Ergebnisse unabhängig von der Durchführung der Messung sind. In der psychologischen Literatur, in der die Definition von Testgütekriterien insbesondere auch auf experimentelle Forschungssituationen unter Laborbedingungen rekurriert, ist damit die Unabhängigkeit der Testergebnisse von der Person des Untersuchungsleiters gemeint (z.B. Lienert und Raatz 1994; zit. nach Kubinger 1995, S. 26). Beeinflussungen bzw. Beeinträchtigungen der Objektivität würden sich bspw. aus unterschiedlichen Anweisungen des Untersuchungsleiters oder aus dem Einfluss besonderen Expertenwissens auf die Auswertung der Daten (Interpretationsobjektivität) ergeben (Bortz und Döring 2002, S. 194). Im Falle standardisierter, quantitativer Erhebungsverfahren mittels Fragebögen ist daher die Objektivität ein vergleichsweise unproblematisches Testgütekriterium (a.a.O., S. 195).

**Reliabilität:** Die Reliabilität oder Zuverlässigkeit eines Tests gibt darüber Auskunft, mit welcher Genauigkeit ein Merkmal durch einen Test gemessen wird. „Es geht also um die *formale Exaktheit* der Merkmalerfassung, um die Zuverlässigkeit, mit der das Ergebnis richtig, im Sinne von exakt ist“ (Kubinger 1995, S. 36; Kursivsetzung im Original). Kromrey (1998) definiert die Reliabilität eines Tests in Anlehnung an Esser u. a. (1977) als intertemporale, intersubjektive und interinstrumentelle Stabilität. in der Terminologie der klassischen Testtheorie ließe sich definieren, dass „die Reliabilität (...) umso höher (ist), je kleiner der zu einem Messwert X gehörende Fehleranteil E ist. Perfekte Reliabilität würde

---

<sup>6</sup>Gleichzeitig stellt Objektivität von Messungen die „inhaltliche Voraussetzung für Reliabilität und Validität dar“ (Tent und Stelzl 1993, S. 43), weshalb sie vorab genannt wird.

bedeuten, dass der Test in der Lage ist, den wahren Wert  $T$  ohne jeden Messfehler zu erfassen ( $X = T$ )“ (Bortz und Döring 2002, S. 195). Dieser ideale Fall tritt in der Realität aufgrund von Störungen der entsprechenden Messvorgängen nicht auf. Um die Qualität der Messung hinsichtlich des Kriteriums der Reliabilität beurteilen zu können, werden verschiedene Verfahren zu deren Ermittlung herangezogen: Mittels rechnerischer Verfahren werden die Retest-Reliabilität, die Paralleltest-Reliabilität, die Testhalbierungs-Reliabilität oder die interne Konsistenz von Tests ermittelt. Diese zielen darauf, die Stabilität von Messungen bzw. die Invarianz der Messinstrumente gegenüber zeitlichen oder persönlichkeitsbezogenen Effekten zu ermitteln. Die Reliabilität gibt entsprechend Auskunft darüber, wie viele (bzw. wie wenige) zufällige Fehler die Messung beeinflussen.

**Validität:** Die Validität, als das zentrale Gütekriterium eines Messinstruments beschreibt, inwieweit dieses tatsächlich misst, was es messen sollte bzw. zu messen beabsichtigt (Schnell u. a. 1999; Bortz und Döring 2002). So kann die Validität eines Messinstruments beeinträchtigt sein, wenn in einer Befragung durch die Fragestellung eine generelle Zustimmungstendenz der Befragten provoziert wird (Schnell u. a. 1999, S. 148). Am genannten Beispiel der Zustimmungstendenz illustrieren Schnell u. a. (1999) den Zusammenhang zwischen Validität und Reliabilität. Das in seiner Validität eingeschränkte Messergebnis kann durchaus eine hohe Reliabilität haben, etwa, wenn über die Items einer Skala hinweg, die beschriebene Zustimmungstendenz in konsistenter Weise auftritt. „Es ist möglich, dass wiederholte Messungen zwar dasselbe Resultat erbringen (also reliabel sind), aber dennoch etwas anderes messen, als beabsichtigt ist (und somit nicht valide sind). Andererseits ist es nicht möglich, dass ein Instrument tatsächlich das misst, was es messen soll, ohne dass zugleich wiederholte Messungen nahezu dasselbe Ergebnis zeigen“ (Schnell u. a. 1999, S. 148). Während die Reliabilität das Auftreten bzw. Ausbleiben zufälliger Fehler bei einer Messung beschreibt, zeigt die Validität entsprechend das Ausmaß des Auftretens systematischer Fehler bei einer Messung an (a.a.O.).

Das Konzept der Validität lässt sich in eine Reihe von Unterkategorien aufteilen, die ihrerseits verschiedene Validitätsaspekte in sich vereinen (s.o.). Neben

der unten ausführlicher beschriebenen Konstruktvalidität handelt es sich dabei insbesondere um die Kriteriumsvalidität, durch die verschiedene Aspekte einer prognostischen Relevanz bzw. Validität von Messungen angesprochen werden: Voraussagen, die auf einer Messung mit einem Instrument beruhen und durch (spätere) Messungen mit einem anderen Instrument bestätigt werden, zeigen die prädiktive Validität oder Übereinstimmungsvalidität an. Die Kriteriumsvalidität lässt sich an der Übereinstimmung von Messungen mit einem äußeren Beurteilungskriterium ablesen. So ließe sich im vorliegenden Kontext von einer positiven studentischen Einschätzung der Lehrqualität bspw. auch auf ein erfolgreiches Studium (hohe Lernzuwächse, erfolgreiche Studienabschlüsse, gute Abschlussnoten, etc.) schließen. Gleichzeitig ist aber zu bedenken, dass Lernerfolg nicht nur von einer guten Lehrqualität, sondern darüber hinaus von den Lerndispositionen der Studierenden abhängig ist (Fleiß, Vorwissen, Zeitbudget, etc.). Ein direkter Rückschluss von Lernerfolgen auf die Studienqualität verbietet sich also, weshalb der Validitätsaspekt der Kriteriumsvalidität im Zusammenhang mit Fragestellungen der Lehrevaluation vglw. schwer zu handhaben ist (Rindermann 1995a; Kulik 2001). Aufgrund der Tatsache, dass die in den unten berichteten Analysen verwendete Datenbasis eine Auswertungsperspektive wie die beschriebene nicht eröffnet, die Kriteriumsvalidität in diesem Zusammenhang also ohnehin nicht untersucht werden kann, wird auf eine weitere diesbezügliche Diskussion hier verzichtet.<sup>7</sup>

## 6.3 Testgütekriterien und Lehrevaluation

Aus der Fülle verschiedener Unterasspekte der beschriebenen Testgütekriterien stehen einige in einem konkreten inhaltlichen Zusammenhang mit der Frage nach

---

<sup>7</sup>Neben der Kriteriumsvalidität sind auch andere Validitätsaspekte von untergeordneter Bedeutung für die Anliegen der vorliegenden Arbeit. So liegen bspw. für die Inhaltsvalidität keine objektiven Kriterien zu ihrer Bestimmung vor. Sie beschreibt vielmehr, ob ein in Frage bzw. zur Messung stehendes theoretisches Konzept umfassend, mit allen seinen relevanten Ausprägungen operationalisiert wurde (Schnell u. a. 1999, S. 149).

der Datenqualität bei der Beurteilung von Hochschulleistungen. Es wurden daher verschiedene Unter Aspekte der genannten Testgütekriterien für die Durchführung entsprechender Datenanalysen ausgewählt, bei der aus vorliegender Sicht sowohl eine unmittelbare Relevanz für hochschulpolitische Entscheidungs- und Steuerungsprozesse, als auch die Möglichkeit, das Vorliegen bzw. die Bedrohung dieser Gütekriterien zu überprüfen, besteht. Bei diesen handelt es sich um die Konstruktvalidität, die zeitliche Stabilität von Qualitätsurteilen und deren Sachbezogenheit.

#### 6.3.1 Konstruktvalidität

Die Qualität von Lehre und Studium ist ein mehrdimensionales Konstrukt (Jackson u. a. 1999; Pellert 2002; Marsh und Roche 1997). Befragungen Studierender zu ihrer Einschätzung der Qualität von Lehre und Studium werden deshalb (idealer Weise) nicht auf die Erhebung eines oder weniger Globalindikatoren beschränkt. Vielmehr werden verschiedene Dimensionen von Lehrqualität vorab theoretisch begründet konstruiert und den Studierenden zur Bewertung vorgelegt. Bei diesen kann es sich - abhängig von der Ebene auf der befragt wird, also zur Qualität einzelner Lehrveranstaltungen, ganzer Studiengänge, etc. - beispielsweise um die Strukturierung der Lehrpräsentation, die Organisationsqualität des Studiums oder andere Qualitätsaspekte handeln.

Gemein ist diesen Kategorien, dass es sich bei ihnen um (mehr oder weniger) komplexe theoretische Konstrukte handelt, die einer unmittelbaren Beurteilung meist nicht zugänglich sind. Diesen Konstrukten, die gemäß den Absichten und Anliegen der Befragenden bzw. „Konstrukteure“ der entsprechend verwendeten Erhebungsinstrumente zur Erfassung der Lehrqualität als relevant erachtet werden, müssen zunächst „direkt erfahrbare Sachverhalte zugeordnet werden, die als beobachtbare Hinweise (...) auf das in seiner (...) Abstraktheit nicht unmittelbar beobachtbare Phänomen dienen können“ Kromrey (1998). Sie werden daher mit Fragen verknüpft, die sich auf die Bestandteile der gewählten Konstrukte beziehen und diese allgemein (bzw. intersubjektiv) verständlich widerspiegeln. So wird bspw. die Frage nach der Organisationsqualität des Studiums mit Fragen

nach dem Abstimmungsgrad von Pflichtveranstaltungen, der Transparenz von Prüfungsanforderungen, etc. verbunden, mit der Absicht, nicht direkt messbare Sachverhalte einer Messung zugänglich zu machen. Methodologisch gesprochen, werden (beobachtbare) Indikatoren bzw. Indikatorvariablen jeweils mit einem latenten (nicht direkt beobachtbaren) Konstrukt verknüpft.<sup>8</sup> Zur Untersuchung von erhobenen Daten hinsichtlich des Vorliegens von Konstruktvalidität - also der Frage danach, ob durch das Messinstrument tatsächlich das gemessen wird, was zu messen beabsichtigt wurde (vgl. z.B. Schmellerleh-Engel und Schweizer 2003) - ist zu analysieren, inwieweit die Indikatorvariablen auch empirisch in Beziehung zu den Konstrukten stehen, denen der Entwickler des Fragebogens sie zugeordnet hat. Eine entsprechende Übereinstimmung würde signalisieren, dass die theoretischen Dimensionen durch die empirisch erhobenen Qualitätsurteile der befragten Personen bestätigt werden, mithin Konstruktvalidität dieser Dimensionen vorliegt. Umgekehrt könnte ein fehlender Zusammenhang der theoretischen Qualitätskonzepte mit den ihnen jeweils zugeordneten Indikatorvariablen andeuten, dass eine Interpretation der Daten, bei der das Zugrundeliegen der entsprechenden Qualitätsdimensionen unterstellt wird, überdacht werden muss.

Das Konzept der Konstruktvalidität lässt sich weiter in zwei relevante Validitätsaspekte untergliedern. Bei diesen handelt es sich um diskriminante und konvergente Validität (vgl. z.B. Kubinger 1995, S. 45).<sup>9</sup>

Diskriminanz oder *diskriminante Validität* liegt vor, wenn sich für ein Konstrukt nachweisen lässt, dass es andere Sachverhalte misst, als andere Konstrukte, wenn also durch das Konstrukt spezifische Sachverhalte erfasst werden, die nicht

---

<sup>8</sup>Durch Korrespondenzregeln wird die Verknüpfung von theoretischer und Beobachtungsebene hergestellt (z.B. Falter 1977; Kromrey 1998).

<sup>9</sup>Zur Überprüfung von Daten auf das Vorliegen von diskriminanter und konvergenter Validität wurde von Campbell und Fiske (1959) die Multitrait-Multimethod Matrix (MTMM) zur Überprüfung und Beschreibung der korrelativen Zusammenhangsstruktur der verschiedenen Indikatorvariablen mit ihren jeweiligen Konstrukten eingeführt (vgl. z.B. Marsh 1982; Schmellerleh-Engel und Schweizer 2003. Die mit der Verwendung dieser Methode verbundenen Probleme werden unten näher erläutert (vgl. Kap. 8).)

gleichzeitig auch durch andere Konstrukte erfasst werden könnten. Dem würde entsprechen, dass die Indikatoren eines Konstrukts mit diesem und nur mit diesem, nicht aber mit einem anderen Konstrukt in einem engen Zusammenhang stehen (vgl. Schmellerleh-Engel und Schweizer 2003).

Konvergenz oder *konvergente Validität* zeigt an, ob die verschiedenen Operationalisierungen eines Konstrukts miteinander in Zusammenhang stehen. Das Vorliegen konvergenter Validität würde durch einen hohen Zusammenhang der verschiedenen Indikatorvariablen eines theoretischen Konstrukts angezeigt werden. Ein Fehlen von Konvergenz ließe sich aus einer Datenstruktur ablesen, in der die einem Konstrukt zugeordneten Indikatorvariablen nur in geringem (korrelativen) Zusammenhang stünden. In diesem Fall würden die dem Konstrukt zugeordneten Indikatoren andere Sachverhalte als den eigentlich angestrebten messen. Dies hätte zur Folge, dass aus Interpretationen der Daten (die bspw. lediglich Häufigkeitsauszählungen zu einzelnen Indikatorvariablen, nicht aber deren weitaus komplexere korrelative Struktur einbezieht) Fehleinschätzungen resultieren können.

#### 6.3.2 Zeitliche Stabilität von Qualitätsurteilen

In der zeitlichen Konsistenz - oder der intertemporalen Stabilität - von Qualitätsdimensionen und Qualitätsbeurteilungen drückt sich ihre Reliabilität aus. Allgemein wird als Kriterium für das Vorhandensein dieses Aspekts von Reliabilität verstanden, dass die Messung der gleichen Konstrukte, die zu verschiedenen Zeitpunkten durchgeführt werden, das gleiche bzw. vergleichbare Ergebnisse produzieren (Bühner 2004, S. 24). Dabei können die Messungen (bzw. Befragungen und Datenerhebungen) sowohl an verschiedenen (aber hinsichtlich bestimmter Aspekte vergleichbaren) als auch an den gleichen Personen durchgeführt werden. Weisen die Qualitätseinschätzungen über die zur Beurteilung vorgelegten Qualitätsdimensionen zeitliche Inkonsistenzen auf, in dem Sinne, dass die Beurteilungen des gleichen Sachverhaltes über einen Zeitraum hinweg stark variieren (ohne dass erkennbare signifikante Veränderungen hinsichtlich dieses Sachverhaltes stattgefunden haben), so ist die Reliabilität als eingeschränkt zu betrach-

ten. Werden also bspw. von Studierenden in aufeinander folgenden Semestern Qualitätseinschätzungen zur Studiensituation erbeten, die zum zweiten Erhebungszeitpunkt ein gegenüber dem ersten stark abweichendes diesbezügliches Bild zeichnen, ohne dass sich die Studienbedingungen objektiv verändert haben, sind die entsprechenden Messungen nicht zuverlässig für die Beurteilung der Studienbedingungen einzusetzen. Leistungsbeurteilungen von Hochschulen, die sich auf derartige Daten stützen, wären in ihrer Aussagekraft eingeschränkt und insbesondere dann diskussionswürdig, wenn sich an diese Beurteilungen eine politische Entscheidung über die Allokation von Mitteln anschließen soll (vgl. auch die Diskussion in Kapitel 4).

Über die Bedeutung als Kriterium für die Überprüfung erhobener Evaluationsdaten bezüglich ihrer Aussagekraft für Steuerungsentscheidungen hinaus, ist der Reliabilitätsaspekt der intertemporalen Stabilität aber auch von einer weiteren wichtigen inhaltlichen Bedeutung. Im Kontext der Evaluation von Lehre und Studium ist zu beachten, dass die verschiedenen Studienabschnitte jeweils unterschiedliche Anforderungen und damit potenziell unterschiedlich relevante Beurteilungskriterien mit sich bringen, die ihrerseits spezifische Beurteilungskompetenzen erfordern (Hofmann und Stiksrud 1994, S. 173). Mittels der Untersuchung zeitlicher Stabilitäten bzw. Instabilitäten lässt sich dem prozesshaften Charakter des Studiums Rechnung tragen. In diesem stehen bspw. die Eingangsvoraussetzungen der Studierenden, Lehr- und Lernprozesse sowie deren Ergebnisse (Lernerfolge und -misserfolge) in einem interdependenten Verhältnis zueinander und können in einer längsschnittlichen, studienbiographischen, kurz: eine die zeitlichen (In-)Stabilitäten von Qualitätseinschätzungen berücksichtigenden Perspektive analysiert werden. Dadurch können unterschiedliche Studienverläufe rekonstruiert werden, die sich bspw. bezüglich der verschiedenen Anforderungen, Voraussetzungen, Leistungsfähigkeiten, Krisen, etc. unterscheiden und sich zu einer Typologie verdichten bzw. aggregieren lassen. Von einer solchen Typologie ausgehend, ließen sich Strategien für die gezielte, gruppenorientierte Betreuung von Studierenden ableiten, die ihrerseits als möglicher Beitrag zur Qualitätssicherung von Lehre und Studium wirken könnte, indem sie die Erfolgchancen

größerer Gruppen Studierender sichert. Aus dieser doppelten Bedeutung für die Überprüfung der Datenqualität hinsichtlich der Reliabilität einerseits und für die Analyse prozessbezogener Einschätzungen der Studienqualität andererseits ergibt sich die Bedeutung, die der Analyse zeitlicher (In-)Stabilitäten von Qualitätsurteilen im Kontext der vorliegenden Arbeit beigemessen wird.<sup>10</sup>

#### 6.3.3 Sachbezogenheit von Qualitätsurteilen

Ein weiterer wichtiger Validitätsaspekt zur Beurteilung der Datenqualität im Kontext der Lehrevaluation bezieht sich auf die Frage, inwieweit die studentischen Qualitätsurteile sich rational auf Eigenschaften des zu beurteilenden Gegenstandes (die Lehrveranstaltung, den Studiengang, etc.) beziehen, bzw. inwieweit sie durch sachfremde Urteilskomponenten, die nicht in dem beurteilten Gegenstand begründet sind, verzerrt werden (vgl. z.B. Rindermann 1995a; Engel und Pohlenz 2001; Greimel und Geyer o.J.). Potenzielle Validitätsverletzungen ergeben sich insbesondere aus der Subjektivität studentischer Befragungsdaten (Engel und Pohlenz 2001). Werden die studentischen Qualitätseinschätzungen bspw. durch ihren persönlichen Beurteilungshintergrund mitbestimmt (Geschlecht, Al-

---

<sup>10</sup>Längsschnittlich erhobene Daten eignen sich darüber hinaus prinzipiell zur Überprüfung der Kriteriumsvalidität. Diese drückt sich u.a. in der prognostischen Validität aus (Kubinger 1995; Schnell u. a. 1999, S. 150). Bezogen auf die Lehrevaluation ließe sich bspw. an eine Überprüfung der Kriteriumsvalidität mittels der Korrelation von gängigerweise abgefragten Beurteilungsdimensionen wie einer „hohen Studienmotivation“ mit einem Außenkriterium (zeitlich nachgelagerte Prüfungsnoten, reibungsloser Übergang in den Arbeitsmarkt) denken. Zu klären ist dabei allerdings die Frage nach der Validität des als relevant erachteten Außenkriteriums selbst. So lässt sich bspw. die Frage des Übergangs in den Arbeitsmarkt vermutlich nicht allein durch die (zeitlich vorgelagerte) Studienmotivation erklären, sondern auch durch andere, externe Faktoren, wie die allgemein angespannte Arbeitsmarktsituation (vgl. a.a.O.). Aufgrund dieser Probleme sowie aufgrund der Tatsache, dass keine längsschnittlichen Daten vorliegen, die sämtliche der genannten, für die entsprechenden Auswertungen nötigen Informationen und Dimensionen einbeziehen, wird im vorliegenden Zusammenhang von einer weiteren Diskussion der Kriteriumsvalidität abgesehen.



ter, Einkommen, wöchentlich aufgewendete Zeit für Erwerbstätigkeit, elterlicher Bildungshintergrund, etc.), wäre davon auszugehen, dass die Validität der auf die Studienqualität bezogenen Urteile potenziell bedroht ist. Darüber hinaus sind a priori subjektiv gefärbte Beurteilungskomponenten, wie die individuelle Studienmotivation und das Interesse zu nennen (Engel und Pohlenz 2001). Die entsprechenden Angaben erbringen zwar relevante Informationen, allerdings mehrheitlich eher über die urteilende Person, als über die Qualität des zu beurteilenden Gegenstandes. In der Studienmotivation oder dem Studieninteresse drücken sich persönliche Präferenzen und Erwartungen an ein Studium aus, die aber nicht notwendigerweise mit dem übereinstimmen müssen, was aus (fach-)wissenschaftlicher oder didaktischer Sicht mit diesem verbunden sein sollte. Die mannigfaltigen individuellen (und vermutlich teilweise widersprüchlichen) Interessen aller Studierenden zu befriedigen, liegt außerhalb dessen, was eine Universität zu leisten im Stande ist. Gleichwohl werden studentische Qualitätseinschätzungen durch das Maß, in dem individuelle Interessen befriedigt und Erwartungen erfüllt werden, mitbestimmt (Engel und Pohlenz 2001). Von der Erwartungsbilanz der Studierenden (die mit ihrer Qualitätseinschätzung kovariiert) auf die tatsächliche Qualität des Studiums zurückzuschließen, würde entsprechend bedeuten, Leistungsbeurteilungen auf eine in ihrer Validität bedrohte Urteilkategorie aufzusetzen. Derartige Bedrohungen der Validität studentischer Befragungsdaten waren in den letzten drei Jahrzehnten immer wieder Gegenstand intensiver Forschungen und kontroverser Debatten (Greenwald 1997). Dabei stand die Position derjenigen, die eine beeinträchtigte Validität studentischer Befragungsdaten als erwiesen ansahen und diese deshalb als mehr oder weniger ungeeignet für die Funktion eines hochschulpolitischen Steuerungsinstruments erachteten, der Auffassung gegenüber, dass eine entsprechende Skepsis unbegründet sei.<sup>11</sup> Die wichtigsten Positionen und

---

<sup>11</sup>Vgl. für die eher skeptische Position z.B. (Kromrey 2001a, 2001c); Esser (1997); Abrami (2001); d'Apollonia und Abrami (1997). Die optimistische Position wird dagegen bspw. von Daniel (1994); Rindermann (2001); Rindermann (1995); Rindermann und Amelang (1994); Marsh (1987); Marsh und Roche (1997); Spiel (2001a); Spiel und Gössler (2000); Wolf u. a.

Befunde zu den potenziell verzerrenden, subjektiven bzw. sachfremden Einflussgrößen sollen im folgenden dargestellt werden. Die Datenauswertungen, die in unten (vgl. Kapitel 10) zu diesem Aspekt dargestellt werden, sollen in erster Linie bereits vorliegende Analyseergebnisse mit den verschiedenen Positionen dieser Debatte konfrontieren.

Studentische Urteile zur Qualität der Lehre (insbesondere auf der Ebene einzelner Lehrveranstaltungen) haben im anglo-amerikanischen Raum eine wesentlich längere Tradition als in Deutschland und stellen dort eine etablierte Praxis der Leistungsbeurteilung dar (z.B. Rindermann 1997). Traditionell werden diese Qualitätsurteile dort für hochschulpolitische Steuerungsentscheidungen eingesetzt, die die einzelnen Lehrenden unmittelbar betreffen (bis hin zu Berufsentscheidungen). Entsprechend der größeren politischen Reichweite und potenziell bedeutsameren individuellen Auswirkungen der studentischen Qualitätsurteile wurde hier auch die Diskussion zu ihrer Validität eher und überaus kontrovers geführt (z.B. Ory 2001). Diese Diskussion setzte bereits in den 1970-80er Jahren ein, um in den 1990er Jahren bereits wieder abzuebben, mit dem Ergebnis, dass die Mehrheit der vorgelegten Forschungen sich die Auffassung zu eigen machte, dass Validitätsbedrohungen nicht in einem Ausmaße vorlägen, das einen Verzicht auf studentische Qualitätsurteile für die Hochschulsteuerung indizieren könnte (Greenwald 1997, S. 1182 f.). In Deutschland wurde diese Debatte mit einiger Verzögerung auch geführt. Insbesondere wurde die Frage nach der Datenqualität studentischer Lehrevaluationen virulent, als eine Basierung von Leistungsbeurteilungen auf studentische Qualitätsurteile auch hier als probates Steuerungsmittel in Erwägung gezogen wurde und die Hochschulen durch die (weitaus meisten) Landeshochschulgesetze zur Durchführung von Evaluationen der Lehre unter Beteiligung der Studierenden verpflichtet wurden (z.B. Hage 1996).

In der Diskussion um die die Validität studentischer Qualitätsurteile potenziell bedrohenden Aspekte ist zwischen verschiedenen Gruppen zu unterscheiden.

---

(2001); vertreten.

Zunächst wird an Lehrevaluationen generell kritisiert, dass diese sich aus den Individualurteilen einer heterogenen Gruppe von Beurteilern zusammensetzen und solche individuellen Urteile nicht zu einem Gesamturteil summiert werden könnten, sondern vielmehr individuelle Akzeptanzurteile darstellten (Kromrey o.J.; Kromrey 1993).

Als ein weiterer Kritikpunkt wird vielfach angemerkt, dass davon auszugehen ist, dass persönliche, soziodemographische Hintergrundmerkmale der Studierenden einen verzerrenden Einfluss auf das Qualitätsurteil ausüben, diese also durch das Geschlecht, das Alter, das eigene und elterliche Einkommen sowie den elterlichen Bildungshintergrund beeinflusst werden (Rindermann 1995a).

Ferner kreist die Diskussion um potenzielle Bias-Variablen, also solche Einflussgrößen, die das Urteil beeinflussen, jedoch nicht rational auf die Eigenschaften des zu beurteilenden Gegenstandes zurückzuführen sind (s.o.). In der Diskussion und Forschung zur diesbezüglichen Beurteilung der Daten der Lehrevaluation wurden insbesondere die individuellen Interessen und Teilnahmemotivationen (Spiel und Gössler 2000), die Benotungspraxis durch die Lehrenden, der Schwierigkeitsgrad von Lehrveranstaltungen, (sachfremde) Attribute der Lehrenden (bspw. Attraktivität, Humor) sowie äußere Kennzeichen und Bedingungen der Lehr- und Lernsituation untersucht (vgl. z.B. Hage 1996; Greenwald und Gillmore 1997; Feldman 1978; Abrami 1980; Abrami u. a. 1982; Abrami u. a. 1990; Hofmann 1988; DuCette und Kenney 1982; Greimel und Geyer o.J.; d'Apollonia und Abrami 1997). Über diese auf die eher inhaltlichen Urteilskomponenten der studentischen Urteile abhebenden Validitätsaspekte hinaus, wurden auf der methodologischen Ebene deren interne Konsistenz sowie zeitliche Stabilität als Kriterien für ihre Validität erforscht (Cranton und Smith 1990; Feldman 1977).

Wie bereits angedeutet wurde (s.o.), sind die Positionen in dieser Debatte äußerst widersprüchlich und werden jeweils durch Forschungsergebnisse gestützt, die für bzw. gegen die Möglichkeit sprechen, studentische Qualitätsurteile als valides und verzerrungsfreies Bewertungsinstrument zu nutzen. Angesichts der umfangreichen und vielfältigen Forschungen in diesem Bereich stellt sich Marsh (1987) die Frage, wie es zu so diametral unterschiedlichen Befunden kommen

kann und schließt, dass diese Situation durch unterschiedliche Forschungsinteressen sowie die Vielzahl an unterschiedlichen, mehr oder weniger sachgemäßen Forschungsansätzen zu Stande kommt: “How can opinions vary so drastically in an area which has been the subject of thousands of studies? Part of the problem lies in the preconceived biases of those who study student ratings; part of the problem lies in the unrealistic expectations of what student evaluations can and should be able to do; part of the problem lies in the plethora of ad hoc instruments based upon varied item content and untested psychometric properties; and part of the problem lies in the fragmentary approach to the design of both student-evaluation instruments and the research based upon them“ (Marsh 1987, S. 260).

Als Ergebnis der Diskussion und als grundlegender Anspruch, der an die Durchführung von Lehrevaluationsstudien und die Erforschung ihrer Validität zu stellen ist, wird von Marsh (a.a.O.) einmal mehr die Berücksichtigung der Multidimensionalität sowohl der Lehrqualität als auch der sich auf diese beziehenden studentischen Urteile sowie eine sachgerechte, theoriegeleitete Konzeptualisierung von Lehrqualität gefordert.

Kennzeichnend für die deutsche, wie für die internationale Diskussion ist, dass sich die Kontroverse hauptsächlich auf studentische Qualitätsurteile auf der Ebene der konkreten Lehrpräsentation in den Lehrveranstaltungen bezieht. Diesbezüglich liegt wie oben beschrieben ein umfänglicher Literaturkörper vor, der sich mit der Klärung der Frage nach der Validität der erhobenen Daten studentischer Lehrveranstaltungskritiken beschäftigt.<sup>12</sup>

Zur Frage der Validität institutioneller Qualitätsbeurteilungen, bspw. auf der Ebene von ganzen Studiengängen, wurde dagegen noch nicht in vergleichbarem Umfang geforscht (Cannon 2001). Gerade in diesem Zusammenhang verstehen sich die unten vorgestellten Sekundäranalysen als ein Beitrag zur Weiterführung der entsprechenden Diskussion.

---

<sup>12</sup>Vgl. zusammenfassend bspw. Rindermann (1997) sowie Spiel und Gössler (2000).

# Kapitel 7

## Datenbasis für die Analysen

Die Analysen, die in den folgenden Kapiteln berichtet werden, basieren auf Daten, die im Rahmen des „Potsdamer Modells der Lehrevaluation“ seit 1998 erhoben wurden. Zunächst soll diese Datenbasis genauer beschrieben werden.

### 7.1 Das Potsdamer Modell der Lehrevaluation

Lehrevaluation an der Universität Potsdam blickt inzwischen auf eine mehrjährige Tradition zurück. Fußend auf einem Beschluss des Akademischen Senats der Universität wurde eine zentrale Servicestelle für Lehrevaluation eingerichtet, die mit der Durchführung der Lehrevaluation beauftragt wurde. Grundlage für die Arbeit dieser Servicestelle stellen die Empfehlungen des Wissenschaftsrates zur „Stärkung der Lehre in den Hochschulen durch Evaluation“ (Wissenschaftsrat 1996) dar. Seit ihrem Bestehen konnte die Servicestelle eine Reihe von Studien zur Studiensituation in den Fächern sowie zu Fragen nach den Gründen für Studienabbruch und zum Verbleib von Absolventen realisieren (Engel und Krekeler 2001; Engel und Pohlenz 2001). Darüber hinaus ist sie mit der regelmäßigen Durchführung studentischer Lehrveranstaltungsbeurteilungen beauftragt (s.u.). Evaluation hat sich an der Universität zu einem Instrument der Studienreform entwickelt, welches Hochschulleitung, Fachbereiche und Fächer sowie Lehrende mit relevanten Informationen zur Studiensituation versorgt. Evaluation wird dabei im wesentlichen als Informationsquelle und internes Steuerungsinstrument

der Hochschulplanung angesehen, Leistungsbeurteilungen und -belohnungen einzelner Lehrender oder Fächer sind in dem gewählten und durch den genannten Senatsbeschluss geregelten Modell nicht vorgesehen. Dieser Verzicht wird nicht zuletzt aus den oben geschilderten Problemen, die sich aus potenziellen Bedrohungen der Validität von erhobenen Daten ergeben, abgeleitet (vgl. Kapitel 6).

Die zentrale Informationsquelle für die Untersuchungen der Servicestelle für Lehrevaluation stellen studentische Befragungsdaten zur Qualitätseinschätzung von Lehre und Studium dar, die auf unterschiedlichen Ebenen erhoben werden. Bei diesen handelt es sich gemäß der in Kapitel 5 beschriebenen eingesetzten Verfahren der Lehrevaluation an Hochschulen um die konkreten Lehrveranstaltungen („studentische Lehrveranstaltungskritik“) und die Studiengänge in den einzelnen Studienfächern („Fach- und Fachbereichsevaluation“) (Engel und Krekeler 2001; Engel und Pohlenz 2001, S. 131 ff.).<sup>1</sup> Darüber hinaus werden retrospektive Qualitätsurteile von Studienabbrechern, Hochschulwechslern und Absolventen erhoben (Pohlenz 2002; Pohlenz 2003; Pohlenz und Tinsner 2004; Pohlenz u. a. 2007).<sup>2</sup>

### 7.1.1 Evaluation der Lehrpräsentation

Ziel der studentischen Lehrveranstaltungskritik ist auch im Rahmen des Potsdamer Modells der Lehrevaluation, die veranstaltungsinterne Diskussion von Lehrenden und Studierenden zu unterstützen und durch die Auswertungsergebnisse auf eine informierte Basis zu stellen. Lehrende und Lernende sollen gleichermaßen in ihrer Selbstreflexion über etwaige Verbesserungspotenziale unterstützt werden. Deshalb beziehen sich die gestellten Fragen neben verschiedenen Aspekten des Dozentenhandelns (Strukturierung der Lehrveranstaltung, Unterstützungs-

---

<sup>1</sup>Diese entsprechen den Erhebungen von Befragungsdaten, die im Rahmen interner Evaluationen auf Studiengangsebene durchgeführt werden (vgl. Kap. 4).

<sup>2</sup>Das Instrument der peer-review konnte bislang erst in sehr begrenztem Maße eingesetzt werden. Nichtsdestotrotz wird es als wichtiger Bestandteil des Modells in künftige Evaluationsaktivitäten einbezogen.

leistungen außerhalb der Seminar- und Vorlesungstermine, etc.) auch auf den Beitrag der Studierenden zum Gelingen der Lehrveranstaltung. Dieser besteht aus vorliegender Sicht bspw. in einer sorgfältigen Vorbereitung verabredeter Literatur und zu haltender Referate sowie einem erkennbaren Engagement für die Veranstaltungsziele.

Die Durchführung der studentischen Lehrveranstaltungskritik ist sowohl für Lehrende als auch für Studierende freiwillig. Diese Festlegung folgt der allgemeinen Erkenntnis, dass Evaluationsdaten insbesondere dann Aussicht darauf haben, zu Verbesserungen der jeweiligen Situation beizutragen, wenn die Beteiligten auf freiwilliger Basis die Gelegenheit der Partizipation nutzen (vgl. z.B. König 2002).

Der bewusste Verzicht auf Leistungsbeurteilungen von Lehrenden, die sich auf studentische Lehrveranstaltungsbeurteilungen stützen, trägt der Möglichkeit subjektiv verzerrter und damit in ihrer Validität bedrohter Qualitätsurteile Rechnung. Durch den Verzicht auf eine Nutzung der Auswertungsergebnisse studentischer Veranstaltungskritiken für Leistungsbeurteilungen und weiterreichende Steuerungsentscheidungen (insbesondere im Zusammenhang mit der leistungsorientierten Mittelvergabe) trägt das vorliegende Evaluationsmodell den Kritikpunkten hinsichtlich potenzieller subjektiver Verzerrungen studentischer Qualitätseinschätzungen Rechnung. Vielmehr konzentriert es sich auf seine Funktion als „sozialkommunikatives Steuerungsinstrument“. Um diesen Charakter der Lehrveranstaltungsevaluation als internes Optimierungsinstrument nicht aufzuheben, werden die Auswertungsergebnisse in erster Linie den Lehrenden zurückgemeldet, damit diese sie in die Diskussion mit den Studierenden einbringen können.<sup>3</sup> Berichte an die Fakultäts- oder Hochschulleitung werden nur in anonymisierter und aggregierter Form vorgenommen, um Leistungsvergleiche und entsprechende Beurteilungen, die auf studentische Qualitätseinschätzungen auf-

---

<sup>3</sup>In der im Wintersemester 2003/04 eingeführten Online-Version der Lehrveranstaltungskritik haben die Lehrenden jedoch die Möglichkeit, die Ergebnisse den Studierenden online zur Einsicht zur Verfügung zu stellen (Pohlentz und Hagenmüller 2005). Von dieser Möglichkeit wird in steigendem Maße Gebrauch gemacht.

setzen und auf Grund etwaiger subjektiver Verzerrungen zu ungerechtfertigten Malus einzelner Lehrender führen, zu vermeiden.

Wie in Kapitel 6 bereits dargestellt wurde, waren Daten der studentischen Lehrveranstaltungskritik in der Vergangenheit bereits vielfach Gegenstand der Forschung zur Validität und der daran geknüpften Aussagefähigkeit für Leistungsbeurteilungen. Die teils widersprüchlichen Ergebnisse dieser Forschungen haben zu der beschriebenen Debatte geführt. Diese soll im Rahmen dieser Arbeit nicht erneut aufgegriffen werden. Vielmehr soll die entsprechende Diskussion um eine Untersuchung von Daten ergänzt werden, die auf der Ebene von Studiengängen und zur Frage der allgemeinen Studienqualität erhoben wurden. Auf eine Sekundäranalyse der Daten aus der studentischen Lehrveranstaltungskritik wird aus diesem Grund im Rahmen der vorliegenden Arbeit verzichtet.

### **7.1.2 Qualitätseinschätzungen auf der Ebene von Studiengängen**

Die Empfehlungen des Wissenschaftsrates, auf denen das hier beschriebene Evaluationsmodell aufbaut, konzeptualisieren die Qualität von Lehre und Studium u.a. als das Ergebnis einer gelungenen Studienorganisation und setzen die Qualitätsdefinition damit auch oberhalb der Ebene einzelner Lehrveranstaltungen an. Fächer bzw. Fachbereiche stellen die wichtigste Bezugseinheit der internen und externen Evaluation von Lehre und Studium dar (Wissenschaftsrat 1996). Zentraler Evaluationsgegenstand im Potsdamer Modell ist demgemäß auch die Studiensituation auf der Ebene der Studiengänge in den verschiedenen Studienfächern. Wie bereits in der Diskussion verschiedener Qualitätsverständnisse und -definitionen in Kapitel 3 dargestellt wurde, entzieht sich die Qualität von Lehre und Studium einer eindeutigen Definition. Vielmehr wird diese durch unterschiedliche, durchaus in Zielkonflikten zueinander stehende Aspekte ausgemacht (Hennen 2002, S. 101). Beispielsweise unterscheidet die Qualitätsdefinition des Wissenschaftsrates fünf Aspekte für die Qualitätsbestimmung „guter“ Lehre. Diese zeichnet sich demnach durch die folgenden Aspekte aus:



- „Qualität kann sich insbesondere auf das Ausbildungsprofil der Absolventen eines Studiengangs, ihre Kenntnisse, Fähigkeiten und Qualifikationen, die als Ergebnisse der Ausbildung bewertet oder beurteilt werden, beziehen. Die Beurteilung von Qualität erfolgt auf der Grundlage von (...) Vorstellungen, Standards oder Normen, die Qualitätsmessung erlauben und ein Urteil darüber ermöglichen, ob ein Absolventenprofil als „gut“ oder „schlecht“ einzuschätzen ist.
- Qualität als Resultat eines hohen Maßes an Konsistenz und Kohärenz eines (Ausbildungs-)Prozesses hebt auf die Organisation des Lehr- und Studienbetriebs, dessen Schlüssigkeit und Stimmigkeit im Hinblick auf gesetzte und zu erreichende Ausbildungsziele ab.
- Qualität als Kongruenz zwischen der von einem Fachbereich getragenen Ausbildungspraxis einerseits und den Erwartungen und Ansprüchen andererseits, die an eine wissenschaftlich fundierte Hochschulausbildung herangetragen werden, konzentriert sich vor allem auf die Studienziele und angestrebten Ausbildungsprofile. Dabei steht ihre Relevanz im Hinblick auf mögliche Tätigkeitsfelder von Absolventen im Vordergrund. Die an die Lehre gerichteten Qualitätserwartungen können dabei von verschiedenen Interessenlagen geprägt werden und unterschiedliche Ausprägungen erfahren.
- Qualität als Ausdruck eines günstigen Verhältnisses zwischen eingesetzten Mitteln und dem damit erzielten Ergebnis orientiert sich an Effizienzgesichtspunkten. Von Bedeutung sind dabei Aspekte der Finanzierung von Ausstattung von Fachbereichen und Hochschulen. Qualität ist in diesem Fall das Ergebnis eines möglichst effizienten und akzeptablen Einsatz dieser Mittel im konkreten Studienbetrieb sowie im Hinblick auf die Realisierung vorgegebener oder selbst gesteckter Ausbildungsziele.
- Qualität als Prozess der Qualifizierung zwischen Beginn und Abschluss eines Studiums stellt Aspekte von Bildung und Ausbildung, Persönlichkeitsentwicklung und Wissenserwerb in den Vordergrund“ (Wissenschaftsrat 1996).

Aus diesen konzeptionellen Vorgaben, denen sich das Potsdamer Modell der Lehrevaluation verpflichtet fühlt, wird einmal mehr deutlich, dass die Qualität von Lehre und Studium sich nicht nur auf Effizienzkriterien reduzieren lässt. Vielmehr sind einerseits gerade auch die angestrebten (akademischen) Ziele der Ausbildung mit einzubeziehen (Engel und Krekeler 2001, S. 5). Andererseits wird auf den *Prozess* der Ausbildung, namentlich den Lehrbetrieb und die Studienwirklichkeit rekurriert (Willems u. a. 1994, S. 7 ff.). Da jedoch die Ausbildungspraxis den jeweiligen Zielen und Vorstellungen der Fächer einer Universität verpflichtet ist, ergibt sich dort auch die Anforderung, Qualitätsbeschreibungen, die sich auf konkrete Standards beziehen, vorzunehmen. „Die inhaltliche Bestimmung von Qualität entsteht in Verhandlungen zwischen allen Beteiligten“ (Meinhold 2000). Diesen Grundlegungen entsprechend, wird die Studienqualität in ihren verschiedenen Dimensionen mittels der Erhebung studentischer Befragungsdaten zu ihrer Einschätzung hinsichtlich verschiedener Qualitätskriterien für Lehre und Studium, beschrieben. Die Themen der Befragungen bestehen dabei aus Fragen, die sich einerseits auf die individuellen Studiendispositionen (Studienmotivation, Lernbiographien, soziodemographische Daten einschließlich des familiären Bildungshintergrundes, etc.) der Studierenden beziehen. Andererseits wird auf ihre Einschätzung verschiedener Aspekte der Studien- und Prüfungsbedingungen sowie bspw. im Fall der Absolventen auf ihre rückblickende Einschätzung der Studienqualität vor dem Hintergrund bereits gemachter Berufserfahrungen eingegangen. Ein besonderes Kennzeichen der Datenerhebung nach dem Potsdamer Modell der Lehrevaluation ist die Konzeption der Befragungen Studierender in einem Längsschnittdesign (Panel-/Kohortendesign).

### 7.1.3 Datenerhebungen im Längsschnitt

Der erfolgreiche berufliche Werdegang von Absolventen ist ein zunehmend zentrales Kriterium für die Beurteilung der Qualität von Studiengängen. Im Mittelpunkt des Interesses steht dabei das erfolgreiche Studium als wichtige Voraussetzung für beruflichen Erfolg und insbesondere für einen möglichst reibungslosen Übergang von der Hochschule in das Berufsleben. Längsschnittdaten zur Beschreibung der

Studienqualität in ihrer zeitlichen Entwicklung sowie zur Analyse individueller Studien- und Berufsverläufe werden daher nicht nur in den Wissenschaftsrats-Empfehlungen zur Evaluation von Lehre und Studium (Wissenschaftsrat 1996) eine besondere Bedeutung beigemessen. Auch anderenorts wird die Wichtigkeit einer längsschnittlichen Datenerhebung zur Analyse von studienbiographischen Entwicklungen und Prozessen betont (vgl. z.B. Ginsburg und Rhett 2003). Aus *evaluationsmethodologischer* Sicht ist in der Erhebung und Analyse von Daten, die Studienbiographien im zeitlichen Verlauf begleiten, ein innovatives Moment zu sehen. Wie in Kapitel 4 bereits diskutiert wurde, ist ein zentrales Merkmal der Evaluationsforschung in der Hochschulevaluation bei der Entwicklung von Untersuchungsplänen nicht hinreichend realisierbar. Namentlich handelt es sich bei diesem um die Anwendung von experimentallogischen Forschungsdesigns mit Kontrollgruppen, die eine Voraussetzung für die Abschätzung von Wirkungen eines Programms auf die dem Programm ausgesetzten Personen ist (vgl. Kapitel 4). Die Analyse von längsschnittlich erhobenen Daten eröffnet zwar nicht die gleichen Analyseperspektiven wie eine Analyse von Daten, die unter (feld-)experimentellen Bedingungen erhoben worden sind. Sie stellen aber gegenüber punktuell erhobenen Querschnittsdaten, die bspw. als Befragungsdaten zu singulären Ereignissen gesammelt werden insofern eine Verbesserung und Weiterentwicklung dar, als sie z.B. erlauben, auf der Individualebene der befragten Personen Stabilitäten oder Instabilitäten in deren Urteilen über den zu beurteilenden Gegenstand - nämlich Studiengänge und Studienbedingungen - im zeitlichen Verlauf abzuschätzen. Über derartige Informationen lassen sich durchaus Schlussfolgerungen über längerfristige Auswirkungen bspw. der Studienbedingungen auf die Lernergebnisse in späteren Studienphasen ziehen.

Der Gedanke der Analyse von Studienverläufen hat in das Potsdamer Modell der Lehrevaluation insofern Eingang gefunden, als die Befragungen Studierender zu ihrer Einschätzung der Studienqualität in einem längsschnittlich angelegten Panel-/Kohortendesign durchgeführt werden (Engel und Krekeler 2001). Bei Daten im Paneldesign handelt es sich um solche, bei denen gleiche Personen zu verschiedenen aufeinander folgenden Zeitpunkten zu bestimmten (gleichen)

Sachverhalten befragt werden (z.B. Engel und Reinicke 1994). Dies eröffnet die beschriebene Analyseperspektive der Betrachtung von Studienverläufen und der Beschreibung der Studiensituation in verschiedenen Studienphasen vor dem Hintergrund der Betrachtung von (In-)Stabilitäten der jeweiligen Einschätzungen befragter Personen (a.a.O.).

Konkret werden Befragungen zu den in Tabelle 7.1 aufgeführten Zeitpunkten im Studienverlauf durchgeführt. Die Wahl dieser Erhebungszeitpunkte leitet sich aus den nachfolgend beschriebenen konzeptionellen Annahmen des Potsdamer Modells der Lehrevaluation ab.

Befragung im...	Befragungsgegenstand
...1./2. Semester	Organisation der Studieneingangsphase
...5./6. Semester	Grundstudium, Zwischenprüfungen
...8./9. Semester	Hauptstudium, Prüfungsphase
Absolventen/ Exmatrikulierte	Einstieg in das Berufsleben

Tabelle 7.1: Thematische Schwerpunkte der Befragungen

Die jeweiligen Spezifika der verschiedenen Studienphasen bestimmen dabei die Schwerpunkte für die studentischen Befragungen, so dass in der Betrachtung von Studienverläufen die unterschiedlichen Entwicklungsstufen berücksichtigt werden können.<sup>4</sup>

- Die *Studieneingangsphase*, die im Regelfall die Phase des Übergangs von

---

<sup>4</sup>An dieser Stelle muss angemerkt werden, dass die hier beschriebene Verfahrensweise sowie die entsprechenden Analysen, die Gegenstand der Kapitel 8 - 10 sind, sich auf die bisher gängige Studienstruktur mit Diplom- bzw. Masterabschlüssen bezieht. Durch die flächendeckende Einführung neuer Bachelor- und Masterstudiengänge sind grundlegende Anpassungen des Evaluationsmodells durchaus vorzunehmen. Gleichwohl sind auch die „neuen“ Studiengänge, zumindest dann, wenn sie als konsekutive Studienprogramme konzipiert sind, durch Studienabschnittsphasen gekennzeichnet, die als einschneidende Zäsuren im Verlauf der Studienbiographie identifiziert werden können.

der Schule in die berufliche bzw. akademische Ausbildung darstellt, spielt eine äußerst wichtige bildungsbiographische Rolle. Hier sind Entscheidungen zu treffen, die für die persönliche Erwerbsbiographie der späteren Arbeitmarktteilnehmer von einschneidender Bedeutung sind. Die angehenden Studierenden sehen sich dabei einer Berufswelt gegenüber, die immer komplexer werdende Anforderungsprofile hervorbringt. Fachliche Qualifikationen, Flexibilität, Teamfähigkeit sowie die Fähigkeit, selbstständig Probleme zu lösen, sind nur einige der relevanten Kompetenzen, die die Basis eines erfolgreichen Berufslebens bilden. Entsprechend weit gefächert sind die Erwartungen, die an eine akademische Ausbildung und die durch sie vermittelten Fähigkeiten gestellt werden sowie die Motivationen aus denen heraus die Entscheidung für die Aufnahme eines Studiums getroffen wird. Der Studieneinstieg kennzeichnet zudem die Phase der Orientierung in der Universität als neuem Lebensraum, mit eigenen, den Studienanfängern möglicherweise bislang unbekanntem Arbeits- und Verhaltensweisen. Die Studierenden müssen zunächst lernen, mit den an sie gestellten Anforderungen zurecht zu kommen. Zu den wichtigsten Aspekten zählt hier sicherlich die zu erwerbende Fähigkeit, den Studienalltag eigenständig zu strukturieren und das vielfältige Angebot einer Hochschule - auch über das Curriculum des jeweiligen Studiengangs hinaus - für sich zu nutzen.<sup>5</sup> Gleichzeitig kann an die Adresse der Hochschulen gerichtet der Anspruch formuliert werden, die neu eingeschriebenen Studierenden zu begleiten und durch eine gute Organisation des Studieneinstiegs etwaige Startschwierigkeiten abzufedern. Eine gute Organisation des Studieneinstiegs und Beratungsleistungen seitens der Hochschule kann dazu beitragen, die Zahl der Studienabbrecher zu senken (Griesbach u. a. 2000) und dem wichtigen Qualitätsmerkmal universitärer Lehre, nämlich dem, dass mit „der Lebenszeit junger Menschen sorgsam

---

<sup>5</sup>Inwieweit durch die angestrebte zeitliche Straffung des Studiums in diesem Zusammenhang ein studienkultureller Wandel herbeigeführt wird, ist noch nicht abzusehen, stellt aber zukünftig vermutlich eine wichtige Frage der Hochschulforschung dar.

umgegangen wird“ (Hochschulrektorenkonferenz 2000b) gerecht zu werden. Bei den Befragungen von Studienanfängern im Rahmen des Potsdamer Modells der Lehrevaluation stehen demgemäß Fragen im Mittelpunkt, die die Aspekte der Studienmotivation und Studierenerwartungen sowie Aspekte der wahrgenommenen Qualität der Organisationsleistungen universitärer Einrichtungen beleuchten.

- In der Phase des *Übergangs vom Grund- ins Hauptstudium*<sup>6</sup> können die Studierenden erstmalig Bilanz ziehen. Es gilt nun bspw. zu beurteilen, ob die an das Studium gestellten Erwartungen, gemessen an dem bislang Erlebten, als erfüllt angesehen werden können. Wurden die Wünsche, die mit der Studienfachwahl verbunden worden waren erfüllt und würde man sich, vor die gleiche Wahl gestellt, erneut für das gleiche Studienfach entscheiden? Gleichzeitig gibt diese Studienphase rückblickend die Gelegenheit, selbstkritisch die eigenen Studienleistungen, die z.B. in Form der Zwischenprüfungsnote messbar sind, mit den gestellten Studienanforderungen zu konfrontieren und so den eigenen Studienfortschritt, aber auch Überforderungen durch das Studium einzuschätzen. So lassen sich bspw. aus persönlichen Potenzialen spezifische Interessen für bestimmte zukünftige Studienschwerpunkte ableiten und weiterentwickeln. In der Formulierung und der Konzentration auf diese Studienschwerpunkte besteht nämlich die besondere Bedeutung des folgenden Studienabschnitts Hauptstudium.<sup>7</sup> Bei sich selbst festgestellte Überforderungen können dazu anregen, die Wahl des Studienfachs zu überdenken und ggf. abzuändern. Die Zäsur des Übergangs zwischen Grund- und Hauptstudium erfüllt also in zweierlei Hinsicht für die Studienbiographie insgesamt eine wichtige Funktion. Sie ist einerseits Meilenstein bei der Überprüfung der Frage, mit der Wahl des Studiums richtig gehandelt zu haben und andererseits - sollte dies der Fall sein - Grundlage für die Ent-

---

<sup>6</sup>Vgl. Fußnote 5.

<sup>7</sup>Die Enstsprechung zu der genannten Übergangsphase in der Struktur konsekutiver Studiengänge wäre in der Abfolge von Bachelor und Masterphase zu sehen (vgl. Fußnote 5).

scheidung über die weitere inhaltliche Ausrichtung und Spezialisierung des Studiums. Diese Entscheidung ist ihrerseits von großer Bedeutung für die spätere berufliche Einsetzbarkeit der im Studium erlernten Fähigkeiten und Wissensinhalte. Der Schwerpunkt der entsprechenden Befragungen liegt daher auch auf der Beurteilung von Lehre und Studium sowohl rückblickend vor dem Hintergrund bereits gemachter Erfahrungen, als auch prospektiv hinsichtlich der geplanten nächsten Schritte in der Studienbiographie.

- Analog zu den Studierendenbefragungen zum Übergang zwischen verschiedenen Studienphasen wird im vorliegenden Modell der Lehrevaluation auch die Phase des *Übergangs in das Berufsleben* durch Befragungen begleitet. Befragt werden einerseits Studierende, die in der Abschlussphase ihres Studiums stehen und um eine rückblickende Beurteilung verschiedener Aspekte der Studiensituation gebeten werden. Andererseits mündet der beschriebene mehrphasige Erhebungszyklus in einer Absolventenbefragung. Studierende, die kurz davor stehen, ihr Studium zu beenden, werden gebeten, die Organisation der Prüfungsphase zu bewerten und einzuschätzen, ob es diesbezügliche Studienzeit verlängernde Faktoren gibt. Darüber hinaus wird den Studierenden die Gelegenheit gegeben, vor dem Hintergrund ihrer Studiererfahrungen zu prognostizieren, wie sich ihre berufliche Zukunft nach Beendigung des Studiums gestalten wird. Der Schwerpunkt der Befragung von Absolventen liegt auf der Beurteilung der Frage, inwieweit das Studium sich für die ersten Schritte ins Berufsleben bezahlt gemacht hat. Im Mittelpunkt steht hier die Frage, ob sich die ehemaligen Studierenden durch das Studium angemessen auf die Berufspraxis vorbereitet fühlen und ob sie den Anforderungen, die im Berufsleben an sie gestellt werden, gerecht werden können. Derartige Informationen stellen aus vorliegender Sicht eine äußerst wichtige Datengrundlage zur Beurteilung der Qualität von Lehre und Studium dar. Für diese wird landläufig die Absolventenquote herangezogen, die als Maß für den „Output“ der Hochschulen aufgefasst wird. Qualitative Daten über die berufliche Entwicklung der Absolventen und die Nützlichkeit der im Studium erlernten Wissensinhalte und Kompetenzen können

darüber hinaus über den „Outcome“, als Maß für den mit dem Output (hier: das erreichte Examen) erzielten Nutzen informieren.

- Quer zu den genannten Befragungen in den verschiedenen Studienphasen werden *Exmatrikulierte* unabhängig von der Zahl der studierten Semester nach den Gründen für die Studienabbruchs- bzw. Hochschulwechselentscheidung gefragt. Die Motive für einen derartigen Schritt sind vielfältig und können sich auf Ursachen beziehen, die in den Studienbedingungen (Organisationsdefizite, infrastrukturelle Rahmenbedingungen, etc.) zu finden sind, die sich aber bspw. auch als Ergebnis der persönlichen Dispositionen der Studienabbrecher und Hochschulwechsler darstellen (Über- oder Unterforderungen, Prüfungsängste, nachlassende Studienmotivation, etc.). Für die Hochschulen sind Informationen über die Determinanten des Studienabbruchs und Hochschulwechsels aus verschiedenen Gründen relevant. Zum einen lassen sich Abbruchgründe identifizieren, die außerhalb des Einflussrahmens der Hochschulen liegen und deshalb für die Beurteilung der Qualität von Lehre und Studium - auf deren Fehlen das Phänomen des Studienabbruchs ja landläufig zurückgeführt wird (BMBF 2003) - nicht geeignet sind (Pohlenz und Tinsner 2004). Durch eine Identifikation derartiger Determinanten und die Beschreibung ihres Einflusses auf die individuellen Abbruchentscheidungen wird den Hochschulen eine wichtige Informationsquelle für die Argumentation in den diesbezüglichen hochschulpolitischen Diskussionen an die Hand gegeben. Andererseits dienen Informationen über die Gründe für das vorzeitige Verlassen der Hochschule diesen als Anhaltspunkt für die interne Weiterentwicklung des Studienangebots, im Sinne der Qualitätssicherung und -entwicklung.

### 7.1.4 Realisierte Einzelstudien

Innerhalb des Modells der Lehrevaluation konnten die in Tabelle 7.2 ausgewiesenen Einzelstudien sowie eine Panelstudie, bestehend aus Befragungen von Studierenden zu Studienbeginn und zum Übergang ins Hauptstudium realisiert werden.



Die Befragungen stellten den Versuch einer Vollerhebung in den jeweiligen Studierendengruppen dar, mit dem Ergebnis selbst-selektierter Stichproben.<sup>8</sup> Es handelt sich also bei den erhobenen Daten nicht um repräsentative Zufallsstichproben im strengen methodologischen bzw. stichprobentheoretischen Sinne. Insofern stellen die für die unten dargestellten Analysemodelle durchgeführten Signifikanztests keine Überprüfungen der Repräsentativität der Ergebnisse für eine Grundgesamtheit Studierender dar. Sie wurden zwar formal durchgeführt, lassen aber lediglich Schlüsse über die Güte der Modelle für die jeweils einbezogenen Befragungsdaten der untersuchten Gruppen Studierender zu. Sie stellen insofern ein eher heuristisches Kriterium für die Beurteilung der berichteten Analyseergebnisse dar.

Studie	Befragung im...	...über	...von Studierenden im...	<i>n</i>
S1	WS 1997/98	Studieneingangsphase	1./2. Sem.	852
S2	WS 1998/99	Studieneingangsphase	1./2. Sem.	753
S3	WS 1998/99	Übergangsphase	5./6. Sem.	519
S4	WS 1998/99	Studienabschluss	8 + Sem.	640
S5	WS 1999/00	Übergangsphase	5./6. Sem.	484
S6	WS 1999/99	Absolventen		105
S7	WS 00/01-01/02	Exmatrikulierte		523
S8	Panel S1-S5		1./2. – 5./6. Sem.	220

Tabelle 7.2: Bislang realisierte Befragungen

Die Ergebnisse der Befragungen wurden der Hochschulleitung und den Fakultätsleitungen berichtet. Hier können sie in die Planung und Weiterentwicklung von Lehre und Studium einbezogen werden. Das Evaluationsmodell lehnt sich insgesamt an das dreistufige Modell der Lehrevaluation aus interner Selbstevaluation, externer Begutachtung durch peers und die Entwicklung von Zielvereinbarungen an (vgl. z.B. Reissert 1994; Carstensen und Reissert 1995). Die beschriebenen Befragungen von Studierenden sind eine wichtige Informationsquelle für die erste

---

<sup>8</sup>Gründe für eine „Nichtteilnahme“ Studierender an den Befragungen entzogen sich weitgehend einer statistischen Kontrolle.

Stufe der internen Evaluation auf der Ebene von Fächern und Fachbereichen. Entwicklungsbedarf besteht noch bei der Durchführung externer Evaluationen (peer reviews). Diese sind im Potsdamer Modell vorgesehen, jedoch konnten hier bislang nur in sehr begrenztem Maße entsprechende Erfahrungen gesammelt werden. Externe Evaluationen sind als Bestandteil des dreistufigen Modells unverzichtbar. Im Sinne einer problemorientierten Weiterentwicklung des Modells auch vor dem Hintergrund der geforderten Verknüpfung von Evaluation und Akkreditierung (z.B. Reil 2002), wird dieses Instrument zukünftig eine größere Rolle zu spielen haben.

# Kapitel 8

## Relevante Dimensionen zur Beurteilung der Studienqualität

In Kapitel 6 wurde der Aspekt der Konstruktvalidität als eines der zentralen Testgütekriterien für Datenerhebungen, die im Bereich der Lehrevaluation durchgeführt werden, genannt. Eng verbunden damit, ist die Frage danach, ob sich die unterschiedlichen Qualitätsdimensionen von Lehre und Studium als urteilswirksam für die entsprechenden Qualitätseinschätzungen befragter Personen herausstellen und sich in deren Urteilen empirisch wiederfinden lassen. Die Frage, inwieweit die als relevant unterstellten Qualitätsdimensionen von den diesbezüglich befragten Personen in der erwarteten Weise aufgefasst und zur Grundlage ihrer Qualitätsbeurteilung gemacht werden, ist deshalb wichtig, weil mit dem Gegenstand Studium und Lehre sehr unterschiedliche Qualitätskonzeptionen verbunden sein können. Generell werden in Untersuchungen zur studentischen Qualitätsbeurteilung von Lehre und Studium unterschiedliche Qualitätsaspekte - in Abhängigkeit vom jeweiligen Forschungsinteresse - in unterschiedlicher Form in den Vordergrund gestellt. Vielfach werden den Studierenden Fragen zu Qualitätsaspekten der Studienorganisation, zu dozentenbezogenen Aspekten, (sowohl zur Lehrpräsentation, als auch hinsichtlich der Betreuung), u.ä. zur Beurteilung vorgelegt (z.B. Engel und Pohlenz 2001; Berghoff u. a. 2004; Berghoff u. a. 2006). Die Multidimensionalität der Qualität von Lehre und Studium ist an sich zwar unstrittig, aber die Frage der Zahl und der Art dieser Dimensionen bleibt vorläufig ungeklärt

(Jackson u. a. 1999). Für eine verlässliche Qualitätsbeurteilung, die ihrerseits für Steuerungsentscheidungen genutzt werden kann, muss daher sichergestellt werden können, dass diese Beurteilungen auf geteilten Vorstellungen der an dieser Beurteilung beteiligten befragenden und befragten Personen beruhen. Die Urteilsrelevanz von theoretisch hergeleiteten Qualitätsdimensionen für die befragten Studierenden kann bspw. mittels (konfirmatorischer) Faktorenanalysen untersucht werden. Bevor die diesbezüglichen Analyseergebnisse berichtet werden, soll ein kurzer Abriss zur Methode der konfirmatorischen Faktorenanalyse zum besseren Verständnis dieser Ergebnisse beitragen.

### 8.1 Konfirmatorische Faktorenanalyse

Qualitätsdimensionen können formal auch als latente (nicht direkt messbare) Konstrukte bezeichnet werden, die über eine theoretisch begründete und/oder empirisch fundierte Zuordnung zu messbaren Indikatoren (manifeste Variablen) operationalisiert werden (z.B. Kromrey 1998). Das Konzept der (konfirmatorischen) Faktorenanalyse bietet sich zur Untersuchung des Zusammenhangs zwischen einem theoretischen Konstrukt und den ihm zugeordneten Indikatorvariablen in besonderer Weise an (vgl. z.B. Proojien und Kloot 2001; Hopko 2003; Schmellerleh-Engel und Keith 1999). Ein Vorliegen derartiger Zusammenhänge wurde als Anzeichen für das Vorliegen von Konstruktvalidität der Messinstrumente - hier: zur Erfassung der Qualitätsurteile über Lehre und Studium - aufgefasst (vgl. Kapitel 6). Auch zur Beschreibung und Erklärung der Zusammenhänge *zwischen* den theoretisch hergeleiteten Urteilsdimensionen wird vielfach auf den Einsatz von konfirmatorischen Faktorenanalysen zurückgegriffen (vgl. z.B. Howard u. a. 1985; sowie Jackson u. a. 1999; Race und Powell 2000).

Das Modell der konfirmatorischen Faktorenanalyse beruht auf der Annahme, dass latente Variablen (Konstrukte) mit den ihnen zugeordneten Indikatorvariablen korrelieren. Die Indikatoren können kausal auf die latenten Konstrukte zurückgeführt werden. Gleichzeitig besteht ein solcher Zusammenhang nicht zu solchen Indikatoren, die aufgrund der theoretischen Vorannahmen nicht mit

dem Konstrukt in einem Wirkungszusammenhang stehen. Anders als im Modell der exploratorischen Faktorenanalyse, welches häufig zur Aufdeckung von Faktorstrukturen und zur Datenreduktion (auf eben diese Faktorstrukturen) eingesetzt wird, werden im konfirmatorischen Faktormodell die Beziehungen zwischen Indikator- und latenten Variablen vorab aufgrund theoretischer Überlegungen festgelegt (Prooijen und Kloot 2001).<sup>1</sup> Das derart spezifizierte Modell wird dann mittels statistischer Schätzverfahren daraufhin getestet, ob die getroffenen Annahmen zutreffen, das theoretische Modell mithin durch die empirischen Daten bestätigt werden kann. (vgl. z.B. Bühner 2004; Kline 1998; Byrne 2001).<sup>2</sup>

Aufgrund von im Modell getroffenen Annahmen wird das zu rechnende Faktorenmodell spezifiziert und in die zu schätzenden Strukturgleichungen zerlegt. In linearen Strukturgleichungsmodellen, zu denen auch die konfirmatorische Faktorenanalyse zählt, wird zwischen einem Messmodell und einem Strukturmodell unterschieden:

- Im Messmodell werden die postulierten Beziehungen zwischen den beobachteten Indikatorvariablen und den jeweiligen latenten Konstrukten beschrieben (dabei werden Annahmen getroffen, wie „jeder Indikator misst nur das ihm zugeordnete Konstrukt und keine Fremddimension (Faktorladung)“; „jeder Indikator verfügt über einen unabhängigen Messfehler“.
- Im Strukturmodell werden die Beziehungen der latenten Variablen untereinander definiert. Dabei kann z.B. spezifiziert werden, ob eine Interkorrelation

---

<sup>1</sup>Nachteile der exploratorischen Faktorenanalyse für die Zwecke der Validitätsprüfung bestehen in der fehlenden Möglichkeit, die Unabhängigkeit bzw. Abhängigkeit der Faktoren zu testen. Diese Möglichkeit ist mit konfirmatorischen Modellen, die auf den LISREL-Ansatz (**L**inear-**S**tructural-**R**ELationship) von Jöreskog (1973) zurückgehen (vgl. z.B. Langer), gegeben.

<sup>2</sup>Dabei werden Kovarianz- oder Korrelationsmatrizen analysiert (Bühner 2004, S. 197). Für die Schätzung der Modellparameter kommen je nach Beschaffenheit der analysierten Daten - bspw. hinsichtlich der Normalverteilung - unterschiedliche statistische Schätzverfahren zum Einsatz, wie z.B. die Maximum Likelihood Methode, Generalised Least Squares (GLS), Unweighted Least Squares (ULS) u.a. (vgl. z.B. Bühner 2004, S. 200 f.).

der latenten Konstrukte angenommen wird.

Abbildung 8.1 zeigt schematisch ein Modell einer latenten Variablen mit zugeordneten Indikatorvariablen (zit. n. Bühner 2004, S. 199). Die latente Variable wird mit dem griechischen Buchstaben  $\xi$  (ksi) bezeichnet.<sup>3</sup> Dieses Modell könnte auch als einfaches, einfaktorielles konfirmatorisches Faktorenmodell bezeichnet werden.<sup>4</sup> Die Faktorladungen  $\lambda_{11}$  und  $\lambda_{21}$  (lambda) zeigen dabei die Stärke der korrelativen Zusammenhänge zwischen den Indikatorvariablen und der latenten Variablen an (die Korrelation zwischen A und B wird auf das Einwirken der latenten Variable  $\xi$  zurückgeführt). Schließlich werden durch die Terme  $e_1$  und  $e_2$  die Messfehlervarianzen der Indikatorvariablen angezeigt. Im Messmodell werden, der Terminologie linearer Strukturgleichungsmodelle folgend, die Matrizen der Faktorladungen der unabhängigen Indikatorvariablen auf ihren Konstrukten als LAMBDA X-Matrix ( $\Lambda_x$ ) bezeichnet.<sup>5</sup>

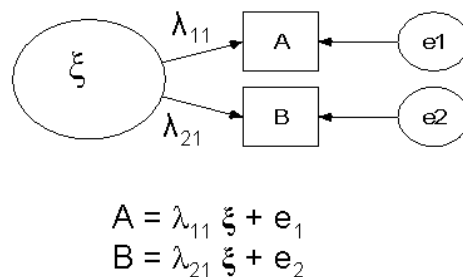


Abbildung 8.1: Modell der konfirmatorischen Faktorenanalyse

<sup>3</sup>Gemäß der Konvention werden für die Bezeichnung latenter bzw. zu schätzender Variablen in linearen Strukturgleichungsmodelle generell griechische Buchstaben verwendet.

<sup>4</sup>In der praktischen Anwendung konfirmatorischer Faktorenanalysen bedürfte es für die Berechnung eines einfaktoriellen Modells mindestens dreier Indikatorvariablen (Bühner 2004, S. 219), das Schema ist wie bereits erwähnt nur zur Verdeutlichung der Funktionsweise des Ansatzes gedacht.

<sup>5</sup>Manifeste Variablen werden mit lateinischen Buchstaben bezeichnet, in diesem Fall als x-Indikatoren.

Gleichung 8.1 zeigt die verallgemeinernde Matrixschreibweise<sup>6</sup> für ein Modell zur Schätzung der Beziehungen zwischen latenten und beobachteten Variablen (vgl. z.B. Backhaus u. a. 2000):

$$X = \Lambda_x * \xi + \delta \quad (8.1)$$

Für die Faktorladungen der Indikatorvariablen ( $\lambda$ , lambda), die latenten Konstrukte ( $\xi$ , ksi) und die Fehlervarianzen ( $\delta$ , delta) werden im Messmodell ebenfalls griechische Buchstaben verwendet.

Die Beziehungen der latenten Konstrukte untereinander werden in der PHI-Matrix ( $\Phi$ ) beschrieben. Gleichung (8.2) stellt schematisch eine PHI-Matrix für ein (Faktoren-)Modell mit drei latenten Variablen, für die Interkorrelation angenommen wird (die  $\varphi$ -Koeffizienten werden zur Schätzung freigegeben), dar (z.B. Langer):

$$\Phi = \begin{bmatrix} 1.00 & & \\ \varphi_{21} & 1.00 & \\ \varphi_{31} & \varphi_{32} & 1.00 \end{bmatrix} \quad (8.2)$$

Über statistische Schätzverfahren werden die Strukturgleichungen geschätzt. Zur Beurteilung, inwieweit die vom Modell unterstellten Zusammenhänge sich durch die empirische Realität bestätigen lassen, werden verschiedene Indizes herangezogen, die das Maß angeben, inwieweit das theoretische Modell durch die empirischen Daten bestätigt wird („Modell-Fit“; vgl. z.B. Bühner 2004; Backhaus u. a. 2000; Schmellerleh-Engel und Keith 1999; Schmellerleh-Engel und Keith 1998; Schmellerleh-Engel u. a. 2003). Als Kennwert zur Beurteilung der Modellgüte wird in der Regel zunächst der exakte Modell-Fit herangezogen. Dieser ist  $\chi^2$ -basiert, so dass die Stichprobengröße großen Einfluss auf das Ergebnis nimmt (Bühner 2004, S. 202). So führen im Falle großer Stichproben bereits kleine Abweichungen von der Korrelations-

---

<sup>6</sup>Die Strukturgleichungen zur Spezifikation der unten berichteten konfirmatorischen Faktoren- und linearen Strukturgleichungsmodelle werden dort dargestellt.

/Kovarianzmatrix zu einer Ablehnung des Modells (bzw. der Hypothese, dass dieses auf die Daten passt; vgl. Bühner 2004, S. 202). Aus diesem Grunde werden üblicherweise zusätzliche Kennwerte zur Beurteilung des Modell-Fit konsultiert, die eine näherungsweise Aussage über die Passung der empirischen Daten mit dem theoretischen Modell treffen (approximative Modell-Fit Indizes).<sup>7</sup> Dies führt dazu, dass Modelle ungeachtet der Tatsache, dass der exakte Modell-Fit eine Ablehnung nahe legen würde, u.U. mit Verweis auf die Werte für die approximativen Fit-Indizes, akzeptiert werden.

Die Anwendung der konfirmatorischen Faktorenanalyse als methodisches Instrumentarium für die Untersuchung des Vorliegens von Konstruktvalidität steht in Verbindung mit der Analyse von Korrelationsmatrizen nach dem Ansatz der Analyse von Multitrait-Multimethod Matrizen (MTMM) (Campbell und Fiske 1959, vgl. auch Schmellerleh-Engel und Schweizer 2003; Bortz und Döring 2002) und wird auch als deren Weiterentwicklung bezeichnet (Schmellerleh-Engel und Schweizer 2003). Diese gründet sich auf der Annahme, dass gültige Messungen dann vorliegen, wenn distinkte Konstrukte durch unterschiedliche Methoden nachgewiesen werden können (diskriminante Validität) und zugleich konvergente Validität vorliegt, die dadurch angezeigt wird, dass verschiedene Operationalisierungen des gleichen Konstrukts in Zusammenhang miteinander stehen (a.a.O.; vgl. z.B. auch Schnell u. a. 1999, S. 155 ff.).

Zur Überprüfung des Vorliegens von Konstruktvalidität, für das die beiden genannten Validitätsaspekte berücksichtigt sein müssen, werden in MTMM-Analysen verschiedene Konstrukte mit jeweils verschiedenen Methoden (unterschiedliche Skalen und Messverfahren) gemessen, wobei alle Konstrukte mit allen

---

<sup>7</sup>Dabei steht eine ganze Reihe von Koeffizienten zur Verfügung, die je nach gewähltem statistischen Schätzverfahren oder aufgrund ihrer Robustheit gegenüber Modellverstößen, wie einer fehlenden Normalverteilung der Daten eingesetzt werden (Bühner 2004, S. 205). Für die unten berichteten Modelle wurden der Comparative-Fit Index (CFI) und der Root-Mean-Square-Error-of Approximation Index (RMSEA) herangezogen, die im Falle des Vorliegens nicht normalverteilter Daten sowie bei Vorliegen fehlender Werte empfohlen werden (a.a.O.).



Methoden gemessen werden müssen (Schmellerleh-Engel und Schweizer 2003, S. 103). Dazu werden die Korrelationen analysiert, die „aus den Testwerten von  $n$  Personen bezüglich  $p$  Traits (Eigenschaften, Merkmalen, Konstrukten), gemessen mit jeweils  $q$  Methoden, resultieren. (...) Valide Messungen sollten einen möglichst geringen methodenspezifischen Anteil aufweisen. Dementsprechend sollten die Korrelationen zwischen Testwerten, die denselben Trait mit unterschiedlichen Methoden (konvergente Validität) erfassen, deutlich höher ausfallen als die Korrelation zwischen Testwerten, die dieselbe Methode verwenden, aber unterschiedliche Traits erfassen (diskriminante Validität)“ (a.a.O., S. 103 f.).<sup>8</sup>

Mit der MTMM-Analyse sind jedoch verschiedene Probleme verbunden, aufgrund derer für die unten dargestellten Datenauswertungen auf ihren Einsatz verzichtet wurde. Die Kritik richtet sich auf die Analyse auf der Ebene der Korrelationen einzelner Testwerte (Schmellerleh-Engel und Schweizer 2003, S. 106). Die Auswertung der Korrelationsmatrix beruht auf einer „Inspektion der manifesten Variablen (...), während die Interpretation der Kriterien [für das Vorliegen von Validität, *P.P.*] als Schlussfolgerung über latente Trait- und Methodenvariablen erfolgt“ (Schmellerleh-Engel und Schweizer 2003, S. 106). Für die Überprüfung der Beziehungen zwischen den latenten Variablen wird der Einsatz anderer statistischer Verfahren, wie der beschriebenen konfirmatorischen Faktorenanalysen ausdrücklich empfohlen (a.a.O.). Ein weiterer Grund für den Verzicht auf MTMM-Analysen besteht darin, dass durch diese insbesondere das Problem der Methodenspezifität von Befragungsergebnissen berücksichtigt wird (Steyer und Eid 1993).<sup>9</sup> Im vorliegenden Fall steht die Frage nach der Art der Beziehungen zwischen den verschiedenen latenten Variablen (Qualitätsdimensionen)

---

<sup>8</sup>In der Vergangenheit wurden vielfach MTMM-Analysen zur Überprüfung der Konstruktvalidität auch im Kontext studentischer Befragungen zu ihrer Einschätzung der Lehrqualität durchgeführt (z.B. Marsh 1982; Marsh 1984; Kremer 1990; zusammenfassend Greenwald 1997).

<sup>9</sup>Dementsprechend ist auch das Vorliegen von Daten unterschiedlicher methodischer Erhebungsarten vorauszusetzen. Diese Bedingung ist im Fall der in die Analysen der vorliegenden Arbeit einbezogenen Datenbasis nicht gegeben. Die analysierten Daten beruhen alle auf studentischen Befragungsdaten mittels standardisierter Fragebögen (s. Kapitel 7).

und den sie messenden Indikatorvariablen im Vordergrund. Insbesondere dieses Interesse begründet den Einsatz des nachfolgend berichteten konfirmatorischen Faktorenmodells.

## 8.2 Fragestellung und Untersuchungsansatz

Zur Untersuchung der Beurteilungsaspekte, die dem studentischen Urteil zur Qualität von Lehre und Studium zugrunde liegen, wurden Daten aus dem „Potsdamer Modell der Lehrevaluation“ herangezogen (vgl. Kapitel 7). Bei den im Zusammenhang mit der Untersuchung der Mehrdimensionalität studentischer Urteile verwendeten Daten handelt es sich um eine Befragung Studierender im fünften bzw. sechsten Fachsemester an der Universität Potsdam. Befragt wurden 519 Studierende verschiedener Fächer im Wintersemester 1998/99 (vgl. Tabelle 7.2, Studie S3). Befragungsgegenstand waren u.a. verschiedene Aspekte der Studienqualität, die jeweils auf einer vierstufigen Ratingskala abgefragt wurden.<sup>10</sup> Die einzelnen Items dieser Fragebatterie wurden mit dem Ziel erhoben, die Studienqualität hinsichtlich verschiedener übergeordneter Urteilskategorien im Spiegel des studentischen Urteils darzustellen. Bei diesen handelte es sich bspw. um den Aspekt der Studierbarkeit des Studienganges, die sich ihrerseits in einer angemessenen Nachvollziehbarkeit der gestellten Studienanforderungen sowie in der Unterstützung der Studierenden bei der Vorbereitung von Prüfungen ausdrückt. Ferner erschien die Studienstruktur relevant zu sein, als Kennzeichen für eine gelungene Organisationsqualität. Die Frage der Interdisziplinarität der Studienbestandteile sowie eine den akademischen Anspruch der Ausbildung widerspiegelnde Urteilskategorie („Wissenschaftlichkeit“) stellten weitere als urteilswirksam vermutete, grundlegende Determinanten des studentischen Urteils dar. Aufgrund inhaltlicher Vorentscheidungen wurden aus der genannten Variablenbatterie zur Erhebung studentischer Qualitätsurteile zu Lehre und Studium die in Tabelle 8.1

---

<sup>10</sup>Die Studierenden wurden in diesem Zusammenhang darum gebeten, sich zwischen den Polen „gerade richtig“ (1) und „nicht weit genug gehend“ (4) zu positionieren.

ausgewiesenen Einzelitems den genannten vier latenten Variablen als Indikatorvariablen zugeordnet.<sup>11</sup>

Latente Variable	Indikatorvariablen	Kürzel (Pfaddiagramm der KFA)
Studierbarkeit ( $\xi_1$ )	- Transparenz der Leistungsanforderungen ( $x_1$ ) - explizite Formulierung von Lehrzielen ( $x_2$ ) - Vorbereitung von Prüfungen ( $x_3$ )	- Transparenz - Ziele - Vorbereitung
Studienstruktur ( $\xi_2$ )	- Strukturierung des Lehrangebots ( $x_4$ ) - inhaltliche Abstimmung ( $x_5$ ) - zeitliche Abstimmung ( $x_6$ )	- Strukturierung - Inhalt - Zeit
Interdisziplinarität ( $\xi_3$ )	- Interdisziplinarität ( $x_7$ ) - Vermittlung fachübergreif. Qualifikation ( $x_8$ )	- Interdisziplin - Fachübergr.
Wissenschaftlichkeit ( $\xi_4$ )	- Verbindung von Lehre und Forschung ( $x_9$ ) - Pluralität von Lehrmeinungen ( $x_{10}$ ) - Vermittlung von Methodenkenntnissen ( $x_{11}$ ) - Theoriebezug ( $x_{12}$ )	- Verbindung - Pluralität - Methoden - Theoriebezug

Tabelle 8.1: Latente Variablen und Indikatoren (Konfirmatorische Faktorenanalyse; KFA)

### 8.2.1 Hypothesen

Mit der Durchführung der auf die genannten Qualitätsdimensionen bezogenen Analyse sind verschiedene untersuchungsleitende Hypothesen verbunden:

1. *Die Indikatorvariablen korrelieren hoch mit den latenten Konstrukten, denen sie jeweils zugeordnet sind und nicht mit den anderen latenten Konstrukten.* Ein entsprechender Befund würde für das Vorliegen diskriminanter Validität, als wichtigem Teilaspekt von Konstruktvalidität sprechen (vgl. Kapitel 6). Diese zeigt an, dass durch ein Konstrukt Sachverhalte gemessen werden, die nicht gleichzeitig auch durch andere Konstrukte gemessen werden.

<sup>11</sup>Neben den Bezeichnungen für die latenten Variablen und die Indikatorvariablen werden in Tabelle 9.1 auch deren im folgenden Pfaddiagramm (vgl. Abbildung 9.1 der konfirmatorischen Faktorenanalyse benutzten Kürzel ausgewiesen.

sen werden könnten (Schnell u. a. 1999, S. 155 ff.).

2. *Die Indikatorvariablen einer latenten Variable korrelieren untereinander hoch.* Dieser Zustand ist die Voraussetzung für das Vorliegen von konvergenter Validität, als zweitem wichtigem Teilaspekt von Konstruktvalidität. Durch die Korrelationen der Indikatoren untereinander wird angezeigt, dass die verschiedenen Operationalisierungen eines Konstrukts tatsächlich miteinander in Zusammenhang stehen und das gleiche Konstrukt messen.
3. *Die vier Faktoren (latenten Variablen) messen distinkte Sachverhalte und weisen deshalb untereinander bestenfalls moderate Korrelationen auf.* Diese Hypothese beschreibt das Verhältnis der latenten Variablen untereinander, wobei im Sinne der Frage nach dem Vorliegen von Konstruktvalidität davon ausgegangen wird, dass diese nur mäßig in einem korrelativen Zusammenhang stehen.

### 8.2.2 Ergebnisse

#### 8.2.2.1 Modellspezifikation

Gemäß der beschriebenen Hypothesen wurden die Modellannahmen des zu rechnenden Faktorenmodells spezifiziert und das entsprechende Modell in zu schätzende Strukturgleichungen zerlegt. Im *Messmodell* werden die Beziehungen zwischen den beobachteten und den latenten Variablen beschrieben (Faktorladungen, Messfehlervarianzen).<sup>12</sup> Das *Strukturmodell* definiert die Beziehungen der latenten Variablen untereinander (Annahme des Vorliegens von Interkorrelationen). In den folgenden Gleichungen werden das Messmodell ( $\Lambda_x$ -Matrix)(8.3) und das Strukturmodell ( $\Phi$ -Matrix)(8.4) der berechneten konfirmatorischen Faktorenan-

---

<sup>12</sup>Im Messmodell sind alle  $\lambda_x$ -Koeffizienten zur Schätzung freigegeben. Die notwendige Restriktion eines Parameters zur Skalierung der latenten Variablen wurde mittels der Festsetzung ihrer Varianzen (auf 1) vorgenommen (Bühner 2004, S. 211). Dieses Vorgehen wird auch in den linearen Strukturgleichungsmodellen der folgenden Abschnitte gewählt.

lyse wiedergegeben.

(a) Messmodell der latenten Variablen:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 \\ 0 & 0 & \lambda_{73} & 0 \\ 0 & 0 & \lambda_{83} & 0 \\ 0 & 0 & 0 & \lambda_{94} \\ 0 & 0 & 0 & \lambda_{104} \\ 0 & 0 & 0 & \lambda_{114} \\ 0 & 0 & 0 & \lambda_{124} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{bmatrix} \quad (8.3)$$

(b) Strukturkomponente:

$$\Phi = \begin{bmatrix} 1.00 & & & \\ \varphi_{21} & 1.00 & & \\ \varphi_{31} & \varphi_{32} & 1.00 & \\ \varphi_{41} & \varphi_{42} & \varphi_{43} & 1.00 \end{bmatrix} \quad (8.4)$$

### 8.2.2.2 Parameterschätzung

Das auf diese Weise spezifizierte konfirmatorische Faktorenmodell wurde mittels Maximum Likelihood (ML) Methode geschätzt.<sup>13</sup> Das Pfaddiagramm in Abbildung 8.2 gibt Auskunft über die geschätzten Parameter (quadrierte Korrelationen, Faktorladungen, Interkorrelationen der Faktoren).

Der Test für den exakten Modell-Fit erbrachte dabei kein zufriedenstellendes Ergebnis. Ein signifikanter  $\chi^2$ -Wert ( $p = .012$ ) von 71,44 bei 47 Freiheitsgraden

---

<sup>13</sup>Diese bietet sich hier insbesondere deshalb an, weil sie vergleichsweise robust gegen Verletzungen von Modellannahmen zur Normalverteilung der Werte ist (Bühner 2004, S. 201).

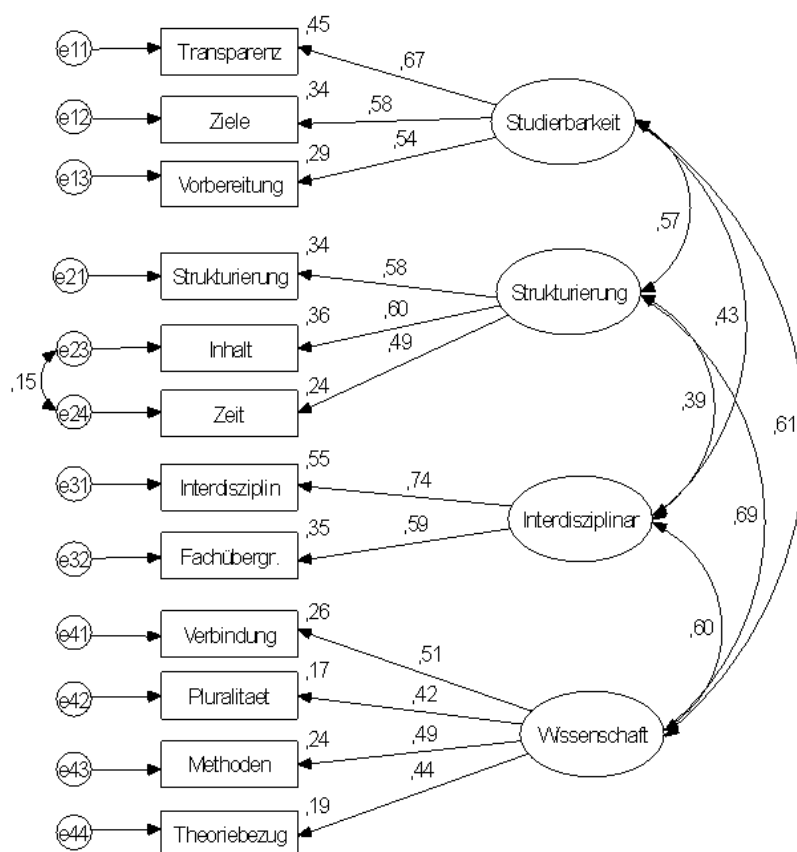


Abbildung 8.2: Pfaddiagramm der KFA; standardisierte Schätzer

spricht zunächst für die Zurückweisung des Modells. Angesichts der über 500 befragten Personen kann dieses Ergebnis allerdings auch auf die Abhängigkeit des exakten ( $\chi^2$ -basierten) Modell-Fits von der Stichprobengröße zurückzuführen sein (vgl. z.B. Bühner 2004; Byrne 2001).<sup>14</sup> Tatsächlich bestätigen die approximativen Modell-Fit Indizes, dass das Modell zumindest vorläufig angenommen werden kann: Die Werte für den CFI und für den RMSEA betragen .964 bzw. .032 (das Konfidenzintervall für den RMSEA liegt zwischen .015 und .046). Beide Fit-Indizes liegen damit innerhalb der für die Beurteilung der Modellgüte vorgeschlagenen Grenzen (Byrne 2001; Kline 1998; Hu und Bentler 1995).<sup>15</sup>

<sup>14</sup>Zudem wird ein zusätzliches, auf den exakten Modell-Fit bezogenes Kriterium erfüllt. Dieses besteht darin, dass der  $\chi^2$ -Wert kleiner als die verdoppelte Zahl der Freiheitsgrade sein soll, damit ein akzeptabler Modell-Fit angezeigt ist (vgl. z.B. Schmellerleh-Engel und Keith 1999).

<sup>15</sup>Die Berechnungen wurden mit dem AMOS-Programmpaket durchgeführt. Dieses bietet

Die Indikatorvariablen korrelieren durchweg hoch mit ihren Faktoren. Für die Dimensionen Studierbarkeit, Strukturierung und Interdisziplinarität ( $\xi_1$  bis  $\xi_3$ ) liegen die Korrelationen zwischen .49 und .74, was auf eine gute Reproduktion der Faktoren durch die zugeordneten Indikatorvariablen hindeutet. Lediglich im Fall des Faktors Wissenschaftlichkeit ( $\xi_4$ ) sind die Faktorladungen geringer, deuten aber mit Werten zwischen .42 und .51 ebenfalls noch auf eine angemessene Operationalisierung des Konstrukts hin. Durch dieses Ergebnis wird die erste der drei formulierten Hypothesen gestützt, dass die Indikatorvariablen jeweils mit ihrem (und nur mit ihrem) Faktor korrelieren und dass mithin diskriminante Validität der Messinstrumente angenommen werden kann.

Die quadrierten Korrelationen, durch die die Korrelationen der Indikatorvariablen untereinander angezeigt wird, weisen im Gegensatz zu den Faktorladungen eher kleinere bis mittlere Werte auf. Wiederum sind im Fall des Faktors Wissenschaftlichkeit die geringsten Werte (.17 - .26) aufzufinden, während die Indikatorvariablen der übrigen Faktoren mit Werten zwischen .24 und .55 hauptsächlich mittlere Korrelationen aufweisen. Die quadrierten multiplen Korrelationen geben die gemeinsame Varianz der Indikatorvariablen, die sich auf den Zusammenhang mit ihrem jeweiligen Faktor zurückführen lässt, an. Sie beschreiben also das Maß, zu dem die Indikatorvariablen eines Faktors auf die jeweils anderen zurückgeführt werden können und mithin reliable Messungen des entsprechenden Konstrukts sind (Bühner 2004, S. 237).

Die durchweg mittleren Werte für die quadrierten Korrelationen der Faktoren 1-3 (Studierbarkeit, Strukturierung, Interdisziplinarität) zeigen an, dass diese durchaus durch die ihnen zugeordneten Indikatorvariablen erklärt werden. Im Fall der Indikatorvariablen des Faktors 4 (Wissenschaftlichkeit) ist die Reliabilität der Messungen des Faktors, die durch die Indikatorvariablen erreicht

---

zur Behandlung fehlender Werte eine Rechenprozedur an, die ohne den fall- oder listenweisen Ausschluss fehlender Werte auskommt. Allerdings werden in diesem Verfahren einige der für die Beurteilung des Modell-Fit üblicherweise herangezogenen Indizes (z.B. GFI, SRMR) nicht ausgegeben (z.B. Byrne 2001).

werden, entsprechend skeptischer einzuschätzen. Insofern kann auch davon ausgegangen werden, dass die konvergente Validität zumindest potenziell bedroht ist und die gewählten Operationalisierungen des Konstruktes „Wissenschaftlichkeit“ dieses nicht hinreichend valide widerspiegeln. Entsprechende Schlussfolgerungen zur Leistungsbeurteilung der Hochschule, die sich auf eben diese „Wissenschaftlichkeit“ der Lehre beziehen, wären demnach nur sehr bedingt auf die entsprechenden studentischen Einschätzungen zu stützen. Die zweite Hypothese wird dementsprechend bestenfalls eingeschränkt als bestätigt angesehen.<sup>16</sup>

Die dritte der Hypothesen unterstellte geringe Interkorrelationen der Faktoren im Sinne ihrer Fähigkeit, inhaltlich distinkte Konstrukte zu messen (sehr hohe Korrelationen würden implizieren, dass die gewählten Konstrukte alle den gleichen Sachverhalt messen und demgemäß die durch das Modell implizierte Annahme einer 4-faktoriellen Struktur sich nicht empirisch bestätigen ließe). Tatsächlich sind aber durchweg mittlere bis hohe Korrelationen von .39 bis .69 vorzufinden. Auch hier ist insbesondere auffällig, dass der Faktor Wissenschaftlichkeit mit den anderen drei Urteilsdimensionen hoch korreliert (.60 - .69), während die Interkorrelationen der übrigen drei Faktoren geringer sind (.39 - .57). Die Operationalisierungen der Dimension „Wissenschaftlichkeit“ führen offensichtlich zu einer starken Kovariation mit den anderen Urteilsdimensionen, was den Schluss zulässt, dass die befragten Personen nicht die erwartete inhaltliche Trennung mit den in diesem Zusammenhang zur Beurteilung vorgelegten Sachverhalten vornehmen. Interpretationen, die die Items dieses Faktors zu einer gemeinsamen Urteils-kategorie verdichten, mit der die „Wissenschaftlichkeit“ der akademischen Ausbil-

---

<sup>16</sup>Als Kriterium für das Vorliegen von Validität und zur Beurteilung der Parameter, die in konfirmatorischen Faktormodellen geschätzt wurden, nennen bspw. Moosbrugger und Fischbach (1999) hohe Faktorladungen ( $\lambda \geq .50$ ) und genügend große multiple quadrierte Korrelationen ( $r \geq .25$ ). Diese Kriterien sind nahezu bei allen Faktoren erfüllt, lediglich die Parameter des vierten Faktors („Wissenschaftlichkeit“) liegen leicht unter den geforderten Werten. Dies deutet zusätzlich auf die tendenzielle Bedrohung der Validität der Operationalisierung dieses Faktors hin.



dung beschrieben werden soll, würden dementsprechend möglicherweise zu einer Fehlbeurteilung der erhobenen Daten gelangen. Demgemäß kann auch die dritte der drei Hypothesen nicht in vollem Umfang durch das Faktorenmodell bestätigt werden.<sup>17</sup> Nach Falter (1977) ist die Konstruktvalidität in zwei Komponenten aufzuteilen. Dabei handelt es sich um die Validität der Indikatoren (wird das Konstrukt von den Indikatoren adäquat abgebildet?) und die Validität des Konstrukts selbst (impliziert das Konstrukt eine theoretisch und empirisch sinnvolle Hypothese?). Die vorliegenden Analyseergebnisse zeichnen ein zweigeteiltes Bild. Während die Faktoren durch ihre jeweiligen Indikatoren durchweg valide abgebildet werden (*Hypothese 1*) ist die Validität der Konstrukte selbst offensichtlich eher skeptisch zu betrachten (*Hypothesen 2,3*).

Die Konstrukte waren sowohl aufgrund theoretischer Vorannahmen, als auch gestützt auf eine vorgeschaltete exploratorische Faktorenanalyse (Hauptkomponentenanalyse) gebildet worden. Gerade derartige Analysen stellen in der Lehrevaluation vielfach das gängige Vorgehen dar, um urteilswirksame Dimensionen des studentischen Urteils zu identifizieren (vgl. z.B. Race und Powell 2000).<sup>18</sup> Zumeist sind derartige Analysen jedoch in erster Linie datengestützt und legen kein a priori formuliertes theoretisches Modell der Dimensionierung der Studienqualität durch die Studierenden zugrunde (Proojien und Kloot 2001). Dies wird durch konfirmatorische Faktorenmodelle wie das oben berichtete ermöglicht. Die entsprechenden Ergebnisse geben dann bisweilen Anlass, die ursprünglich formu-

---

<sup>17</sup>Auf Modellmodifikationen, die bspw. in einer Neuordnung von Indikatorvariablen bestehen könnten, wird verzichtet, um den konfirmatorischen Charakter des Modells zu wahren. Aus statistischer Sicht kann das Modell wie oben erwähnt vorläufig angenommen werden. Gleichzeitig ist zu betonen, dass aus der Tatsache, dass das Modell „auf die Daten passt“ nicht darauf geschlossen werden darf, dass es sich um das einzig gültige Modell zur Beschreibung der Beziehungsstrukturen der einbezogenen Variablen handelt.

<sup>18</sup>Gleichzeitig wird durch dieses Vorgehen eine Dimensionenreduktion angestrebt, um die Interpretation der erhobenen Daten zu erleichtern. Auf Schwierigkeiten bei der Bestätigung von exploratorisch aufgedeckten Faktorstrukturen durch konfirmatorische Analysen weisen bspw. Proojien und Kloot (2001) hin.

lierten theoretischen Modelle bzw. die exploratorisch ermittelten Faktorenstrukturen zu überdenken oder gar zurückzuweisen (a.a.O.).

Insbesondere zeigen die vorliegenden Ergebnisse aber, dass eine Interpretation von studentischen Qualitätsurteilen hinsichtlich der verschiedenen (als urteilswirksam unterstellten) Urteilsdimensionen potenziellen Validitätsbedrohungen ausgesetzt ist und dementsprechend zu Fehlbeurteilungen von Hochschulleistungen führen kann.

# Kapitel 9

## Beziehungen der Qualitätsdimensionen

In einem nächsten Schritt ist es das Ziel, die Beziehungen der latenten Urteilsdimensionen untereinander, weiterführenden Analysen zu unterziehen. Dabei werden die im vorangegangenen Abschnitt ermittelten vier Kategorien (Studierbarkeit, Strukturierung, Interdisziplinarität und Wissenschaftlichkeit) beibehalten, auch wenn das oben berichtete konfirmatorische Faktorenmodell nahe legte, dass die gewählten Operationalisierungen zumindest teilweise potenziellen Validitätsbedrohungen ausgesetzt sind. Anliegen der hier berichteten Datenauswertungen ist es nämlich auch, anhand eines Beispiels durchgängig einen möglichen Ansatz zur Untersuchung der Datenqualität hinsichtlich relevanter Gütekriterien zu demonstrieren und die Implikationen der entsprechenden Ergebnisse zu diskutieren.

Vor dem Hintergrund der allgemeinen Forderung einer Steuerungswirksamkeit von Evaluationen und den entsprechend erhobenen Daten stellt sich die Frage, inwieweit die ermittelten Urteilsdimensionen die Qualitätseinschätzungen der Studierenden ihrerseits beeinflussen. Welche Bedeutung haben Interdisziplinarität und Wissenschaftlichkeit für die studentische Zufriedenheit, welche Rolle spielen eine gute Strukturierung und Studierbarkeit in diesem Zusammenhang und an welchen Stellen müssen bzw. können Studiengänge hinsichtlich ihrer Interdisziplinarität und ihres akademischen Anspruchs weiterentwickelt werden? Aus den Antworten auf derartige Fragen lassen sich konkrete Handlungsempfehlungen für

Hochschul- oder Fachbereichsleitungen ableiten. Derartigen Fragestellungen soll im Folgenden mittels des Einsatzes linearer Strukturgleichungsmodelle nachgegangen werden. Analog zur Vorgehensweise im vorangegangenen Kapitel wird zunächst eine kurze Zusammenfassung dieser Methode zum besseren Verständnis der berichteten Ergebnisse gegeben.

### 9.1 Lineare Strukturgleichungsmodelle

Der Ansatz linearer Strukturgleichungsmodelle ermöglicht es im vorliegenden Anwendungsfall, Aussagen über die Zusammenhänge der als urteilswirksam ermittelten Qualitätsdimensionen untereinander zu treffen. Dies ist für eine Planung und Steuerung von Hochschulen, die sich auf relevante Evaluationsergebnisse stützt, durchaus von praktischer Bedeutung: Über die Beziehungen zwischen einer Qualitätsdimension und den sie repräsentierenden Einzelaspekten von Qualität hinaus, kann ermittelt werden, inwieweit Reformanstrengungen bezüglich der einen Dimension von Qualität positive bzw. negative Effekte bezüglich einer anderen erwarten lassen. Dies ermöglicht eine nachhaltige und umfassende Planung, in der mögliche Auswirkungen des Handelns prognostiziert werden können und entsprechend Entscheidungen auf einer umfassend informierten Basis getroffen werden können. Im Falle eines vollständigen linearen Strukturgleichungsmodells wird dieses gegenüber dem Ansatz der konfirmatorischen Faktorenanalyse um eine (oder mehrere) abhängige (endogene) latente Variable(n) erweitert, die ihrerseits auf unabhängige (exogene) latente Variable(n) kausal zurückgeführt und ebenfalls durch Indikatorvariablen operationalisiert werden. Endogene latente Variablen werden mit dem griechischen Buchstaben  $\eta$  (eta) bezeichnet, ihre Fehlervarianzen werden mit  $\epsilon$  (epsilon) notiert. Abbildung 9.1 zeigt schematisch ein einfaches lineares Strukturgleichungsmodell mit einer unabhängigen ( $\xi$ ) und einer abhängigen latenten Variablen ( $\eta$ ) sowie deren Messfehler  $\zeta$  (zeta). Der unabhängigen latenten Variablen sind (wie im oben beschriebenen konfirmatorischen Faktorenmmodell) die x-Indikatoren zugeordnet, während die abhängige latente Variable durch die y-Indikatoren operationalisiert wird.

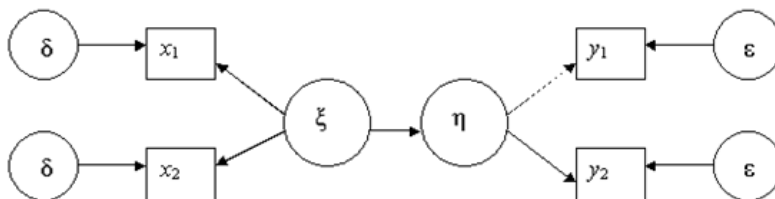


Abbildung 9.1: Modell eines linearen Strukturgleichungsmodells

Analog zum Ansatz der konfirmatorischen Faktorenanalysen werden auch lineare Strukturgleichungsmodelle für die Berechnung in Strukturgleichungen zerlegt. Das in Gleichung 9.1 dargestellte Messmodell der durch die  $y$ -Indikatoren gemessenen  $\eta$ -Variablen wird analog zur Spezifikation der  $x$ -Indikatoren (vgl. Gleichung 8.1) durch die  $\Lambda_y$ -Matrix gebildet (vgl. z.B. Backhaus u. a. 2000):

$$Y = \Lambda_y * \eta + \epsilon \quad (9.1)$$

Die Strukturkomponente zur Berechnung der abhängigen latenten Variablen wird durch die Beziehungen zwischen den in das Modell aufgenommenen  $\eta$ -Variablen (in der BETA-Matrix,  $B\eta$ ) untereinander sowie durch die Beziehungen zu den unabhängigen latenten  $\xi$ -Variablen (in der GAMMA-Matrix,  $\Gamma\xi$ ) und den Messfehler der  $\eta$ -Variablen ( $\zeta$ ) konstituiert. In Matrixschreibweise wird die allgemeine Form der Strukturkomponente wie in Gleichung 9.2 notiert (vgl. a.a.O.):

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta \quad (9.2)$$

Pfadkoeffizienten zwischen den unabhängigen und den abhängigen latenten Variablen werden mit  $\gamma$  (gamma) bezeichnet.<sup>1</sup>

---

<sup>1</sup>Werden mehrere  $\eta$ -variablen in das Modell aufgenommen, so werden die entsprechenden Pfadkoeffizienten zur Beschreibung der Beziehungen zwischen diesen Variablen mit  $\beta$  (beta) bezeichnet.

## 9.2 Kausalanalyse zur Studienzufriedenheit

Zur Analyse der Determinanten der Studienzufriedenheit Studierender in der Übergangsphase zwischen Grund- und Hauptstudium wurde ein lineares Strukturgleichungsmodell geschätzt. Gegenüber dem konfirmatorischen Faktorenmodell wurde dabei die einbezogene Datenquelle leicht variiert. Während in die Faktorenanalyse alle befragten Studierenden der genannten Studienphase einbezogen wurden, beschränkte sich das folgende Strukturgleichungsmodell auf die Studierenden des fünften Fachsemesters, um eine homogene Gruppe untersuchen zu können.<sup>2</sup> In diesem Strukturgleichungsmodell werden die vier bereits untersuchten Urteilsdimensionen als unabhängige (exogene) latente Variablen zur Erklärung der abhängigen (endogenen) latenten Variable „Studienzufriedenheit“ konzeptualisiert. Das Maß, in dem die Befragten Zufriedenheit mit dem Studium zu Protokoll geben, wird dabei als das Ergebnis einer Gesamtbilanz über das bisherige Studium sowie der Erwartungsbilanz der befragten Studierenden aufgefasst. Gesamturteil und Erwartungsbilanz wurden jeweils über eine 11stufige Skala mit den Polen „sehr schlecht“/Erwartungen wurden „überhaupt nicht erfüllt“ (1) und „sehr gut“/Erwartungen wurden „absolut erfüllt“ (11) abgebildet.<sup>3</sup> Diese beiden Variablen stellen dementsprechend die Indikatorvariablen der abhängigen latenten Variable „Studienzufriedenheit“ dar. Ferner wurde angenommen, dass die Studienzufriedenheit nicht nur auf die Wirkungen der einzelnen exogenen latenten Variablen zurückzuführen ist, sondern vielmehr ein Ergebnis ihres Zusammen-

---

<sup>2</sup>In dem zuvor einbezogenen Datensatz war eine gewisse Streuung der Studierenden über mehrere Semester zu finden, in die diese eingeschrieben waren. Die angestrebte Homogenität der Gruppe Studierender bezieht sich dabei auf die Zeit und damit einhergehenden Studienerfahrungen, die sie - unabhängig von der Tatsache, ob sie sich bereits im Hauptstudium befinden - an der Universität gesammelt haben. Im Falle Studierender, die bereits eine größere Anzahl Semester absolviert haben, liegen naturgemäß ähnliche Erfahrungshintergründe, wie bei Studierenden der später in einer vergleichenden Perspektive zu untersuchenden fortgeschrittenen Studienphasen vor. Die entsprechend bereinigte Stichprobe umfasste 315 Studierende.

<sup>3</sup>Für die Berechnung der folgenden Modelle wurden die Skalen umcodiert, um ihre Richtung mit denen der x-Indikatorvariablen der exogenen latenten Variablen zu harmonisieren.

wirkens ist. Ein solches Zusammenspiel der verschiedenen Einflussfaktoren der Studienzufriedenheit könnte als ein weiterer Beleg für die Mehrdimensionalität studentischer Qualitätsurteile zur Studienqualität gewertet werden.<sup>4</sup>

### 9.2.1 Hypothesen

Diese Annahmen lassen sich in folgenden Hypothesen zur Erklärung der Datenstruktur zusammenfassen:

1. *Die Studienzufriedenheit wird durch unterschiedliche Urteils-komponenten beeinflusst.* Das studentische Gesamturteil zur Studienzufriedenheit ist nicht als selbstständig zu erfassender „Globalindikator“ zu erfassen, sondern wird durch verschiedene Aspekte der wahrgenommenen Studienqualität bestimmt.
2. *Die Studienzufriedenheit wird durch das Zusammenwirken der vier verschiedenen Urteilkategorien beeinflusst.* Die Qualitätseinschätzungen sind nicht nur jeweils für sich urteilswirksam, sondern beeinflussen im Zusammenspiel die studentische Eindrucksbildung zur Studienqualität.
3. *Die Studienzufriedenheit korreliert positiv mit den vier exogenen latenten Variablen.* Je besser Studierbarkeit, Strukturierung, Interdisziplinarität und Wissenschaftlichkeit eingeschätzt werden, umso höher ist die Studienzufriedenheit.

---

<sup>4</sup>Gleichzeitig wird dieses Zusammenspiel im vorliegenden Modell nicht kausal interpretiert. Vielmehr wird auf die Frage der Erklärung der Studienzufriedenheit durch die genannten Variablen fokussiert. (Im Pfaddiagramm in Abbildung 9.2 werden diese kausal uninterpretierten Beziehungen durch die gekrümmten Pfeile zwischen den exogenen latenten Variablen dargestellt.)

## 9.2.2 Ergebnisse

Analog zur Darstellung der Ergebnisse der konfirmatorischen Faktorenanalyse werden nachfolgend zunächst die Modellspezifikation des linearen Strukturgleichungsmodells (Darstellung in Matrixschreibweise) und danach die Ergebnisse der Parameterschätzung beschrieben.

### 9.2.2.1 Modellspezifikation

Im vorliegenden kausalanalytischen Strukturgleichungsmodell werden anders als im oben beschriebenen Modell der konfirmatorischen Faktorenanalyse, die Wirkungen der latenten Variablen auf eine ebenfalls latente, abhängige Variable (Studienzufriedenheit) beschrieben. Die abhängige latente Variable ( $\eta$ ) wird ihrerseits durch y-Indikatorvariablen (Gesamtbilanz und Erwartungsbilanz) gemessen (vgl. Tabelle 9.1):

Latente Variable	Indikatorvariablen	Kürzel (Strukturgleichungsmodell)
Studienzufriedenheit ( $\eta$ )	- Erwartungsbilanz ( $y_1$ ) - Gesamturteil über das bisherige Studium ( $y_2$ )	- Erwartung - Gesamturteil

Tabelle 9.1: Endogene latente Variable, Indikatoren (Strukturgleichungsmodell)

Aus Gründen der Übersichtlichkeit werden die Korrelationskoeffizienten der Indikatorvariablen in den Messmodellen jeweils mit einem Superskript als Prädiktor der exogenen bzw. endogenen latenten Variablen gekennzeichnet ( $\lambda^{(x)}$  bzw.  $\lambda^{(y)}$ ; s. Gleichung 9.3 und 9.5). Die Messfehlervarianzen der y-Indikatorvariablen werden mit  $\epsilon$  (epsilon) bezeichnet. Die Spezifikation der Modellstruktur für das Messmodell der x-Indikatorvariablen zur Erklärung der unabhängigen (exogenen) latenten Variablen ( $\xi$ ) erfolgt entsprechend derjenigen für das in Kapitel 8 berichtete konfirmatorische Faktorenmodell:

(a) Messmodell der exogenen latenten Variablen:



$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(x)} & 0 & 0 & 0 \\ \lambda_{21}^{(x)} & 0 & 0 & 0 \\ \lambda_{31}^{(x)} & 0 & 0 & 0 \\ 0 & \lambda_{42}^{(x)} & 0 & 0 \\ 0 & \lambda_{52}^{(x)} & 0 & 0 \\ 0 & \lambda_{62}^{(x)} & 0 & 0 \\ 0 & 0 & \lambda_{73}^{(x)} & 0 \\ 0 & 0 & \lambda_{83}^{(x)} & 0 \\ 0 & 0 & 0 & \lambda_{94}^{(x)} \\ 0 & 0 & 0 & \lambda_{104}^{(x)} \\ 0 & 0 & 0 & \lambda_{114}^{(x)} \\ 0 & 0 & 0 & \lambda_{124}^{(x)} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{bmatrix} \quad (9.3)$$

Verallgemeinernd kann die  $\Lambda_x$ -Matrix auch als

$$x = \Lambda_x * \xi + \delta \quad (9.4)$$

notiert werden (vgl. auch Kap. 8). Im Messmodell für die abhängige (endogene) latente Variable  $\eta$  werden deren Beziehungen zu den y-Indikatoren spezifiziert (Gleichung 9.5):

(b) Messmodell der endogenen latenten Variable:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(y)} & 0 \\ 0 & \lambda_{21}^{(y)} \end{bmatrix} * [\eta] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (9.5)$$

Die allgemeine Notierung der  $\Lambda_y$ -Matrix lautet

$$y = \Lambda_y * \eta + \varepsilon \quad (9.6)$$

Das Strukturmodell gibt die Beziehungen zwischen den latenten abhängigen und unabhängigen Variablen an. Die  $\gamma$ -Koeffizienten repräsentieren dabei die Korrelationen zwischen den unabhängigen Variablen ( $\xi$ ) und der abhängigen Variable

( $\eta$ ).<sup>5</sup> Deren Residualkategorie (Fehlerterm) wird als  $\zeta$  (zeta) notiert:

(c) Strukturkomponente:

$$\eta = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \gamma_{31} & \gamma_{41} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} + \begin{bmatrix} \zeta \end{bmatrix} \quad (9.7)$$

bzw. allgemein:

$$\eta = \Gamma * \xi + \zeta \quad (9.8)$$

### 9.2.2.2 Parameterschätzung

Das Modell wurde nach der ML-Methode geschätzt. Abbildung 9.2 zeigt das entsprechende Pfaddiagramm mit der standardisierten Version der Parameterschätzungen. Wie im Fall der konfirmatorischen Faktorenanalyse legt der exakte Modell-Fit zunächst die Ablehnung des Modells nahe. Der signifikante  $\chi^2$ -Wert beträgt 108,57 bei 66 Freiheitsgraden ( $p = .001$ ).<sup>6</sup> Allerdings deuten auch hier die Werte für die approximativen Fit-Indizes (CFI = .922; RMSEA = .045, das Konfidenzintervall liegt zwischen .029 und .060), dass das Modell vorläufig zur Erklärung der Studienzufriedenheit durch die einbezogenen Variablen akzeptiert werden kann.

Die Parameter zu den Indikatoren der exogenen latenten Variablen bestätigen die Befunde der oben berichteten konfirmatorischen Faktorenanalyse. Die Werte für die quadrierten multiplen Korrelationen der x-Indikatorvariablen bleiben für die Variablen Studierbarkeit, Strukturierung und Interdisziplinarität nahezu unverändert, bezogen auf das Gesamtmodell kann daher auf insgesamt zufriedenstellend reliable Messungen der latenten Variablen durch die zugeordneten

---

<sup>5</sup>Im vorliegenden Fall entfällt die BETA-Matrix ( $\mathbf{B}\eta$ ), weil nur eine endogene latente Variable vorliegt.

<sup>6</sup>Auch hier ist allerdings das Kriterium  $\chi^2 \leq 2 * df$  erfüllt.

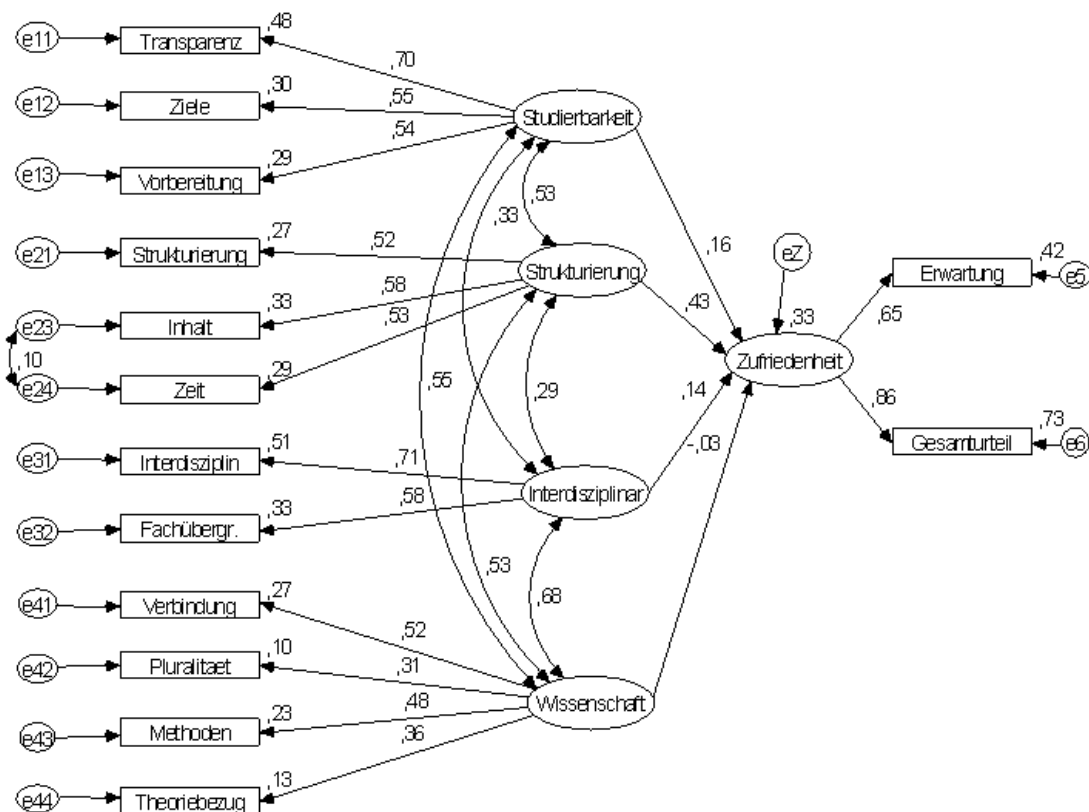


Abbildung 9.2: Kausalanalyse der Studienzufriedenheit; standardisierte Schätzer

Indikatorvariablen geschlossen werden. Auch die Werte für die Korrelationen der x-Indikatorvariablen mit den exogenen latenten Variablen sind im Vergleich zum konfirmatorischen Faktorenmodell nahezu gleichbleibend hoch. Im Falle der Variable „Wissenschaftlichkeit des Studiums“ liegen die Werte dagegen noch leicht unter denen des Faktorenmodells (.31 bis .52).

Auch die endogene latente Variable wird hinreichend gut durch die zugeordneten y-Indikatorvariablen („Erwartungen erfüllt“ und „Gesamturteil“) abgebildet. Deren quadrierte multiple Korrelationen (.42 bzw. .73) zeigen die Reliabilität der Messung des Konstrukts „Studienzufriedenheit“ durch diese Indikatoren an. Die Korrelationen der y-Indikatorvariablen mit der latenten Variable sind mit Werten von .65 und .86 sogar vergleichsweise hoch. Generell wird durch das Modell bestätigt, dass dem studentischen Gesamturteil zur Qualität von Lehre und

Studium (ausgedrückt durch die Studienzufriedenheit) unterschiedliche, inhaltlich distinkte Urteilskomponenten zugrunde liegen. Die Befunde zeigen einmal mehr, dass das studentische Qualitätsurteil nicht durch einen oder wenige Globalindikatoren erfasst werden kann. Dieser Befund, der sich aus der (empirischen) Bestätigung des Modells insgesamt ableiten lässt, kann als Stütze für die erste der drei oben formulierten Hypothesen gewertet werden, nach der die studentische Qualitätseinschätzung durch verschiedene urteilswirksame Aspekte konstituiert wird. Durch die vier exogenen latenten Variablen werden 67 % der Varianz in der Studienzufriedenheit erklärt.

Die Interkorrelationen der entsprechend als relevant identifizierten Urteils-komponenten (exogene latente Variablen) zeigen darüber hinaus an, dass die Determinanten der Studienzufriedenheit ihrerseits in einem Beziehungsgeflecht stehen. Für diese Interkorrelationen konnten Werte zwischen .29 und .68 ermittelt werden (vgl. Abbildung 9.2). Die der Studienzufriedenheit zugrunde liegenden Faktoren stellen also nicht nur jeweils einzelne Determinanten der Studienzufriedenheit dar, sondern beeinflussen diese auch vermittelt über ihr spezifisches Beziehungsgeflecht.

Durch diesen Befund wird auch die zweite der formulierten Hypothesen gestützt. Er ist insofern von Bedeutung, als durch Erkenntnisse der Beziehungen zwischen den Determinanten der Studienzufriedenheit, etwaige Entwicklungsstrategien gezielter auf identifizierte Entwicklungsbedarfe abgestimmt werden können. Auswirkungen von Steuerungsentscheidungen, die mit dem Ziel getroffen werden, Verbesserungen in einem Bereich der Studienqualität herbeizuführen, können so hinsichtlich der zu erwartenden Effekte in anderen Bereichen präziser prognostiziert werden.

Die dritte Hypothese betraf die Korrelationen zwischen den exogenen und der endogenen latenten Variable. Dabei wurde vermutet, dass die Studienzufriedenheit positiv mit den vier erklärenden Variablen Studierbarkeit, Strukturierung, Interdisziplinarität und Wissenschaftlichkeit korreliert ist, dass also die Studienzufriedenheit mit einer positiven Beurteilung dieser vier Dimensionen steigt. Wie die entsprechenden Koeffizienten im Pfaddiagramm in Abbildung 9.2 ver-

deutlichen, trifft diese Vermutung in unterschiedlichem Maße zu. Demnach ist der stärkste (positive) korrelative Zusammenhang zwischen der Studienzufriedenheit und der Strukturierung des Lehrangebots zu finden (der Korrelationskoeffizient beträgt .43), während für Studierbarkeit (.16) und Interdisziplinarität (.14) vergleichsweise moderat positiv mit der Studienzufriedenheit korrelieren. Der Aspekt der Wissenschaftlichkeit ist dagegen sogar schwach negativ (-.03) mit der Studienzufriedenheit korreliert. Diese sinkt also tendenziell, wenn das Studium als „zu akademisch“, im Sinne der diesem Faktor zugeordneten Indikatoren eingeschätzt wird. Diese Beziehungen zwischen der abhängigen und den unabhängigen latenten Variablen verdeutlichen die Gewichtungen, die den verschiedenen Qualitätsaspekten seitens der befragten Studierenden beigemessen werden. Während inhaltliche Aspekte des Studiums, wie eine interdisziplinäre Ausrichtung und eine Orientierung an akademischen Ausbildungszielen eher nachgeordnete Urteilkategorien sind, bzw. das Gesamturteil über das Studium sogar negativ beeinflussen, sind die eher „instrumentellen Aspekte“ des Studiums, wie eine gelungene zeitliche Organisation des Lehrangebots von größerer Relevanz für die studentische Eindrucksbildung.

Die für die Qualitätsbeurteilung einschlägigen Komponenten stehen darüber hinaus nicht - wie oben vermutet - in einem durchgängig positiven Zusammenhang mit der Studienzufriedenheit, sondern beeinflussen dieses teilweise auch negativ. Die dritte der oben formulierten Hypothesen kann dementsprechend nicht gänzlich durch das Modell bestätigt werden.

Studentische Qualitätseinschätzungen zu Lehre und Studium haben mittlerweile einen nicht zu vernachlässigenden Einfluss auf die (öffentliche) Leistungsbeurteilung von Hochschulen und Studiengängen. So werden bspw. für das vom Centrum für Hochschulentwicklung (CHE) regelmäßig durchgeführte Hochschulranking (vgl. Berghoff u. a. 2004; Berghoff u. a. 2006) studentische Gesamturteile zu ihrer Einschätzung der Studienqualität erhoben und als einschlägiger Indikator zur Beurteilung der Leistungsfähigkeit der entsprechend gerankten Hochschulen bzw. Studienfächer herangezogen.

Angesichts der hier vorgelegten Auswertungsergebnisse muss betont werden,

dass diese Gesamturteile i) möglichen Validitätsbedrohungen ausgesetzt sind und ii) studentische Qualitätseinschätzungen potenziell in einer Weise auf den Globalindikator „Gesamturteil“ verdichten, die der Vielschichtigkeit studentischer Beurteilungen nicht gerecht wird und relevante Informationen über die Art und Zusammenhänge dieser Qualitätseinschätzungen vernachlässigt.

Erkenntnisse über diese Vielschichtigkeit können im besten Fall dazu führen, dass Entwicklungsstrategien differenziert auf die jeweils unterschiedlich relevanten Aspekte des Studium eingehen. Im schlechtesten Fall könnten die Ergebnisse dazu genutzt werden, die Studiengangsplanung ausschließlich an der Maxime der Gefälligkeit der Lehre (Verzicht auf „Wissenschaftlichkeit“, Stärkung der Strukturierung, mit dem Ergebnis weiterer Verschulung) auszurichten, um in Rankings wie dem genannten, zu besseren Werten hinsichtlich des Globalindikators „studentisches Gesamturteil“ zu kommen und die Rangposition zu verbessern. Derartige Strategien würden das Ziel der Leistungsmessungen konterkarieren, indem sie durch eine Absenkung von Ansprüchen und Standards, zu einer Qualitätsverschlechterung beizutragen geeignet sind.

## 9.3 Stabilität von Qualitätseinschätzungen

Ein wichtiger Aspekt zur Beurteilung der Validität und Reliabilität von Messungen besteht in ihrer Stabilität über Zeiträume hinweg. Messungen, die sich durch eine Invarianz gegen zeitlich bedingte Einflüsse auszeichnen, weisen eine hohe Reliabilität auf. Gegenstand der folgenden Analyse ist es dementsprechend zu untersuchen, inwieweit die als urteilswirksam identifizierten Qualitätsaspekte auch für Studierende späterer Studienphasen leitend bei der Beurteilung des Studiums sind bzw. inwieweit sich die entsprechenden Gewichtungen im Zeitverlauf verschieben.

### 9.3.1 Variabilität im zeitlichen Verlauf der Studienphasen

Derartige Verschiebungen könnten insofern in Validitätsbedrohungen erhobener Daten resultieren, als etwaige unterschiedliche Gewichtungen, die dem studen-

tischen Urteil in verschiedenen Studienphasen zugrunde liegen, möglicherweise nicht durch die eingesetzten Erhebungsinstrumente sowie die Interpretation der entsprechenden Auswertungsergebnisse reflektiert werden. Anders formuliert: Die gleichen Sachverhalte, die Studierenden unterschiedlicher Studienphasen zur Beurteilung vorgelegt werden, können studienphasenspezifisch eine andere Gewichtung durch die befragten Personen erfahren. Dadurch besteht zumindest die Möglichkeit, dass die in den entsprechenden Erhebungsinstrumenten zur Beurteilung vorgelegten Konstrukte nicht durchgängig das messen, was sie zu messen beabsichtigten, das also ihre Validität bedroht ist.

Ähnlich wie die Erkenntnis, dass sich die studentische Qualitätseinschätzung in ihrer Mehrdimensionalität nicht durch einen oder wenige Globalindikatoren abbilden lässt, wäre entsprechend in Befragungen zu berücksichtigen, dass die Studierenden unterschiedlicher Studienphasen jeweils spezifische Auffassungen von Studienqualität haben. Insofern diese „studienphasentypische Färbung“ des studentischen Urteils nicht bei der Erhebung und Interpretation von Qualitätseinschätzungen gewürdigt wird, ist die Möglichkeit potenzieller Fehlinterpretationen zumindest gegeben. Die Unterscheidung spezifischer Sichtweisen Studierender bei der Erhebung ihrer Qualitätseinschätzungen wird in der Praxis der Lehrevaluation vielfach nicht durch die entsprechend eingesetzten Befragungsinventare aufgenommen.<sup>7</sup>

Daher soll im Folgenden mittels einer entsprechenden Analyse der im Panel-/Kohortendesign erhobenen Daten aus dem Potsdamer Modell der Lehrevaluation geklärt werden, inwieweit studienphasenspezifische Unterschiedlichkeiten tatsächlich urteilswirksam für die studentische Qualitätseinschätzung sind. Bei der für diesen Zweck herangezogenen Datenbasis handelt es sich ebenfalls um

---

<sup>7</sup>Dies betrifft bspw. die Daten, die vom CHE für die Ermittlung studentischer Qualitätseinschätzungen erhoben werden, auf denen wiederum das oben bereits erwähnte Hochschulranking basiert. Für die Ermittlung der studentischen Beurteilung von Lehre und Studium werden Studierende verschiedener Studienphasen um die Bearbeitung des gleichen Fragebogens gebeten (vgl. Berghoff u. a. 2004; Berghoff u. a. 2006).

eine Befragung, die im Wintersemester 1998/99 durchgeführt wurde. Dabei wurden solche Studierende um ihre Qualitätseinschätzungen gebeten, die zum Befragungszeitpunkt im achten oder einem höheren Fachsemester an der Universität Potsdam eingeschrieben waren. Die Fragebögen wurden von 640 Studierenden beantwortet (vgl. Tabelle 7.2).<sup>8</sup>

#### 9.3.1.1 Hypothesen

Lineare Strukturgleichungsmodelle können über ihre Funktion zur Analyse von Qualitätsdimensionen (und damit verbundenen Validitätsaspekten) hinaus für Stabilitätsanalysen, zur Untersuchung der Relevanz bzw. Urteilswirksamkeit von Qualitätsaspekten im Zeitverlauf eingesetzt werden (vgl. z.B. Engel und Pötschke 2003). Diesen Ansatz aufgreifend, werden die Ausführungen zur Frage der Variabilität der studentischen Urteile in verschiedenen Studienphasen zu den folgenden Hypothesen verdichtet, die im folgenden kausalanalytischen Modell zur Erklärung der Studienzufriedenheit fortgeschrittener Studierender getestet werden sollen:

1. *Für die Studierenden unterschiedlicher Studienphasen sind die gleichen Komponenten bei der Qualitätseinschätzung urteilsrelevant.* Im Sinne reliabler und (konstrukt-)valider Messungen der Qualitätsdimensionen kann erwartet werden, dass die qualitätsrelevanten Konstrukte über den zeitlichen Verlauf des Studiums hinweg nicht (oder zumindest nicht gänzlich) an

---

<sup>8</sup>Das Potsdamer Modell der Lehrevaluation ist zwar insgesamt im Panel-/Kohortendesign angelegt. Bei der Befragung Studierender des achten (und höheren) Fachsemesters handelt es sich aber um keine Wiederholungsbefragung derjenigen Studierenden, die zuvor in der Übergangsphase zwischen Grund- und Hauptstudium befragt worden waren. Die beiden Stichproben stellen jeweils Einzelstudien zu verschiedenen Studienzeitpunkten dar. Hier ist sicher eine Limitation der Aussagefähigkeit der Auswertungsergebnisse zu sehen, sofern der Längsschnittcharakter der Daten zur Diskussion steht und die entsprechende Anforderung, Urteilsvariationen innerhalb der befragten Personen aufzudecken. Allerdings geben auch die Befragungen unterschiedlicher Studierendekohorten durchaus Einblicke in die Effekte des Zeitverlaufs bei der Qualitätsbeurteilung des Studiums.



Erklärungskraft für das Zustandekommen des studentischen Urteils verlieren.

2. *Studierende unterschiedlicher Studienphasen legen unterschiedliche Schwerpunkte bei der Beurteilung der verschiedenen Aspekte der Studienqualität.* Unterschiedliche Beurteilungen in verschiedenen Studienphasen implizieren unterschiedliche Qualitätsanforderungen, die sich in einer variierenden Gewichtung der verschiedenen Qualitätsaspekte widerspiegeln.

#### 9.3.1.2 Modell und Parameterschätzung

Die Spezifikation des Modells entspricht gänzlich derjenigen, die bereits im Modell zur Erklärung der Studienzufriedenheit der Studierenden im Wechsel zwischen Grund- und Hauptstudium (5. Fachsemester) dargestellt wurde. Im Sinne des längsschnittlichen Charakters des Erhebungsmodells, das über verschiedene Studienphasen studentische Einschätzungen zu verschiedenen Qualitätsaspekten des Studiums erhebt, wurde auch den in der späteren Studienphase (achtes und höhere Fachsemester) befragten Studierenden die oben in die Modellbildung einbezogene Fragenbatterie zur Einschätzung der Studienqualität vorgelegt.<sup>9</sup> Die Spezifikation des vorliegenden Modells ist daher mit der oben ausführlich beschriebenen identisch (vgl. Gleichungen 9.3 - 9.8) und wird deshalb hier nicht erneut dargestellt. Das in Abbildung 9.3 gezeigte Pfaddiagramm zeigt die nach der ML-Methode geschätzten Parameter (standardisierte Version). Auch in die-

---

<sup>9</sup>Die Analyse zielt auf einen Vergleich der Urteilswirksamkeit latenter Konstrukte in unterschiedlichen Gruppen Befragter. Mittels einer vorgeschalteten Multi-Gruppen Analyse wurde daher zunächst sicher gestellt, dass diese als urteilswirksam unterstellten Konstrukte (Studierbarkeit, Struktur, Interdisziplinarität und Wissenschaftlichkeit) in den beiden zu vergleichenden Gruppen in gleicher Weise aufgefasst werden. Es wurde das gleiche Modell simultan für beide Gruppen geschätzt, unter der Annahme, dass die Faktorladungen in beiden Gruppen gleich sind (Gleichheitsrestriktion der  $\lambda_x$  und  $\lambda_y$ -Indikatoren). Das entsprechend geschätzte Modell bestätigte diese Annahme hinreichend (Schätzung nach ML; CFI = .918; RMESA = .028, Konfidenzintervall: .022 - .034).

sem Fall ist der exakte Modell-Fit unbefriedigend. Der signifikante  $\chi^2$ -Wert von 122,6 legt mit 66 Freiheitsgraden ( $p = .000$ ) die Ablehnung des Modells nahe. Allerdings ist auch hier die Größe der Stichprobe ( $n = 640$ ) als mögliche Erklärung für dieses Ergebnis heranzuziehen.<sup>10</sup> Dafür sprechen auch im Fall des vorliegenden Modells die Werte für den approximativen Modell-Fit (CFI = .954; RMSEA = .037, bei einem Konfidenzintervall zwischen .026 und .047). Aufgrund dieser Befunde wird auch das vorliegende Modell zunächst akzeptiert.

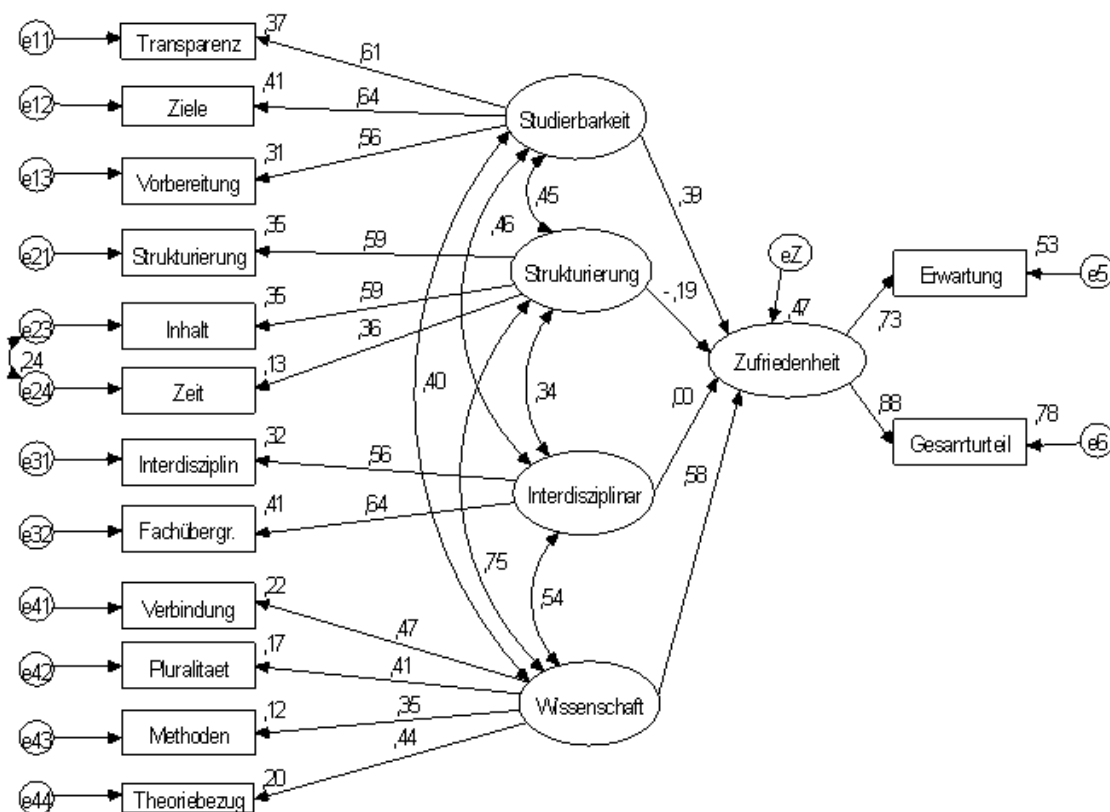


Abbildung 9.3: Qualitätsurteile in späteren Studienphasen; standard. Schätzer

Auch im Fall der Befragung Studierender des achten und höheren Fachsemesters lässt sich anhand der quadrierten multiplen Korrelationen der x-Indikatorvariablen und der Faktorladungen (Korrelationen zwischen den Indikatorvariablen und den latenten exogenen Variablen) schließen, dass die latenten Variablen durch die ihnen zugeordneten Indikatoren gut repräsentiert werden.

<sup>10</sup>Zudem ist wiederum das Kriterium  $\chi^2 \leq 2 * df$  erfüllt (s.o.).

Gleiches gilt für die endogene latente Variable („Zufriedenheit“). Die quadrierten multiplen Korrelationen (.53 und .78) zeigen die Reliabilität der Messung des Konstrukts „Studienzufriedenheit“ durch diese Indikatoren an. Die Korrelationen der y-Indikatorvariablen mit der latenten Variable sind mit Werten von .73 und .88 auch hier sehr hoch. Im Falle der Studierenden der höheren Fachsemester wird durch das Modell 53% der Varianz in der Studienzufriedenheit durch den Einfluss der exogenen latenten Variablen (Studierbarkeit, Strukturierung, Interdisziplinarität, Wissenschaftlichkeit) erklärt. Die Tatsache, dass die Qualitätsdimensionen, die bereits für das Urteil der Studierenden in einer früheren Studienphase als urteilsrelevant identifiziert wurden, auch für die Urteile der fortgeschrittenen Befragten relevant sind, spricht für die Annahme der ersten der oben formulierten Hypothesen. Offenkundig sind die Operationalisierungen der verschiedenen Qualitätsaspekte über die Zeit des Studienverlaufs hinweg von Relevanz für die studentischen Urteile. Ihnen kann entsprechend attestiert werden, reliable Messungen der jeweiligen Konstrukte darzustellen. Die Konsistenz von Faktorstrukturen sowohl über die Zeit, als auch bspw. über verschiedener Urteilergruppen wird zudem als Kriterium für das Vorliegen von Konstruktvalidität angesehen (vgl. z.B. Cadwell und Jenkins 1985).

Auffällig ist insbesondere aber auch, dass die vier Qualitätsdimensionen in einem entscheidend unterschiedlichen Verhältnis zur abhängigen Variable „Studienzufriedenheit“ stehen. So korreliert der Aspekt der „Wissenschaftlichkeit des Studiums“ hoch (.58) mit der Studienzufriedenheit, während diese Urteilskomponente im Fall der Studierenden des fünften Fachsemesters ein negativer Prädiktor war (vgl. Abb. 9.2). Diese Funktion kommt im Fall der fortgeschrittenen Studierenden dem Aspekt der Strukturiertheit des Studiums zu, die mit einem Koeffizienten von -.19 negativ mit der Studienzufriedenheit korreliert ist. Die Studierbarkeit des Studiums korreliert dagegen höher positiv (.39), während die Interdisziplinarität in keinerlei korrelativem Zusammenhang mit der Studienzufriedenheit steht.

Studierende der früheren Studienphase stellen offensichtlich eher solche Aspekte der Studienqualität, in das Zentrum ihrer Qualitätseinschätzungen, die

eine gute Organisation und eine dementsprechend leichte Orientierung im Studienalltag beschreiben. Bei den fortgeschrittenen Studierenden verhält es sich dagegen genau umgekehrt. In einer fortgeschrittenen Phase des Studiums ist die Orientierung hinsichtlich organisatorischer, auf die Strukturierung des Studiums bezogener Aspekte des Studiums weitgehend abgeschlossen. Eine gute Strukturierung stellt sich entsprechend nicht (mehr) als zentrales Problem bei der Beschreibung der Studienqualität dar. Vielmehr kann eine zu weitgehende Strukturierung die Studienzufriedenheit beeinträchtigen, insofern sie möglicherweise als Verschulung des Studiums und als Eingriff in die Selbstorganisation des Studiums aufgefasst wird. In der negativen Korrelation der Qualitätsdimension „Strukturierung“ mit der Studienzufriedenheit findet diese Interpretation ihre Entsprechung. Ähnliches gilt für die inhaltliche Ausrichtung des Studiums fortgeschrittener Studierender. Auch diese ist in einer späteren Studienphase abgeschlossen, Entscheidungen zur fachlichen Spezialisierung sind getroffen worden. Die Frage eines orientierenden und fachübergreifenden Angebots unterbreitenden Studienangebots ist daher für die Studierenden dieser Studienphase nicht in dem Umfang relevant, wie sie es für Studierende zu Beginn ihres Studiums ist. Diese sind im Sinne einer späteren Schwerpunktsetzung ihrer Studieninhalte darauf angewiesen, sich frühzeitig über das Angebot anderer Fächer und entsprechende etwaige Kombinationsmöglichkeiten zu informieren. Dementsprechend wird hier die wahrgenommene diesbezügliche Studienqualität (bspw. ausgedrückt als die Betonung fächerübergreifender Lehransätze durch das Fach und das entsprechende Engagement der Lehrenden) in höherem Maße als relevante Urteilskomponente exponiert.

Die Frage der Studierbarkeit des Studiums nimmt bezogen auf die Stärke der Korrelation mit der Studienzufriedenheit einen mittleren Platz ein, der Wert für die Korrelation ist aber deutlich höher, als dies bei den Studierenden der Übergangsphase zwischen Grund- und Hauptstudium zutrifft. Angesichts des nahenden Studienabschlusses sind Aspekte wie die Transparenz von Leistungsanforderungen und die Frage der Prüfungsvorbereitung naturgemäß von äußerster Dringlichkeit und gehen entsprechend in die Bilanz hinsichtlich der Studienzufriedenheit ein.

Schließlich ist im Falle der fortgeschrittenen Studierenden der Qualitätsaspekt der „Wissenschaftlichkeit“ des Studiums als derjenige zu identifizieren, der am stärksten (positiv) mit der Studienzufriedenheit korreliert. Für die Studienanfänger treten die akademischen Ausbildungsziele gegenüber der Notwendigkeit, die Studienorientierung hauptsächlich organisatorisch zu meistern, in den Hintergrund. Dieses Verhältnis kehrt sich im Laufe des Studiums und nach dem Sammeln entsprechender Studiererfahrungen und Fachkenntnisse, offensichtlich um. Inhaltliche Aspekte der Ausbildung, wie eben Fragen der Verbindung von Forschung und Lehre, eine gute fachbezogene Methodenausbildung, etc. gewinnen an Bedeutung für die Einschätzung der Studienqualität. Dies schlägt sich auch in der Art nieder, wie Studierende in unterschiedlichen Studienphasen das Studium bewerten und in den Unterschiedlichkeiten, die in diesem Zusammenhang bei Studierenden verschiedener Studienphasen gefunden werden können. Die Analyseergebnisse zeigen diese Verschiebung der Gewichtung bei der studentischen Qualitätseinschätzung überaus deutlich, so dass die zweite der beiden oben formulierten Hypothesen durch das hier gerechnete Modell bestätigt wird. Es erweisen sich zwar die gleichen Qualitätsaspekte als relevant für das Zustandekommen der studentischen Urteile (*Hypothese 1*), aber deren Bedeutung wird über die verschiedenen Studienphasen hinweg unterschiedlich aufgefasst.

Dieser Befund macht gleichzeitig deutlich, dass die Validität von studentischen Befragungsdaten bedroht ist, sofern die offenkundige Vielgestaltigkeit der Qualitätsauffassungen nicht angemessen bei der Operationalisierung der zu messenden Konstrukte berücksichtigt wird. Bezogen auf die zeitliche Verschiebung bei der Relevanz von Urteilskategorien heißt dies, dass Studierende unterschiedlicher Studienphasen mit Fragen zu unterschiedlichen Qualitätsaspekten konfrontiert werden müssen, um ihre Studienzufriedenheit in angemessen valider Weise abzubilden.

Insofern wird auch der längsschnittlich orientierten Ansatz des Potsdamer Modells der Lehrevaluation bestätigt, der darin besteht, Studierende unterschiedlicher Studienphasen zusätzlich zu einem allgemeinen Fragenrepertoire auch mit unterschiedlichen Aspekten zur Studienqualität zu konfrontieren. Die mit die-

sem Ansatz verbundenen Vorteile bestehen wie erwähnt in einer vergleichsweise höheren Validität der gemessenen Konstrukte. Diese ermöglicht ihrerseits die Formulierung differenzierterer Ansätze zur Nutzung der Auswertungsergebnisse für hochschulpolitische Steuerungsentscheidungen.

#### 9.3.2 Längsschnittdaten in der Lehrevaluation

Neben der im vorangegangenen Abschnitt beschriebenen Analyseperspektive, ermöglichen die Daten des Potsdamer Modells der Lehrevaluation auch Analysen im Paneldesign. Diese sind dadurch charakterisiert, dass studentische Befragungsdaten der gleichen Personen im Zeitverlauf betrachtet werden. Dazu wird im Folgenden eine Studie herangezogen, die Befragungen Studierender zu verschiedenen Zeitpunkten zusammenfügt.

Ziel der nachfolgend dargestellten Analyse ist es, in Anlehnung an die Hypothesen des vorangegangenen Abschnittes zu überprüfen, inwieweit Kategorien, die zu verschiedenen Zeitpunkten des Studiums zur Beurteilung vorgelegt werden, im Zeitverlauf urteilswirksam bleiben. Ein korrelativer Zusammenhang der Messungen zu einer Urteilsdimension an den gleichen Personen zu verschiedenen Zeitpunkten würde insofern für die *Reliabilität* dieser Dimension sprechen, als dieser Zusammenhang ein Indiz dafür wäre, dass die befragten Personen die zur Beurteilung vorgelegte Dimensionen im Zeitverlauf in gleicher Weise interpretieren und zur Grundlage ihrer Qualitätseinschätzung machen.

Diese Frage stellt sich ganz unabhängig von der im vorangegangenen Abschnitt diskutierten, inwieweit sich die Bedeutung bzw. Gewichtung der einzelnen Urteilkategorien für die Studierenden verändert.

##### 9.3.2.1 Untersuchungsansatz und Ergebnisse

Mit den vorliegenden im Paneldesign erhobenen Daten, lässt sich ermitteln, inwieweit studentische Qualitätsurteile, die zu einem nachgelagerten Zeitpunkt ( $t_2$ ) abgegeben werden, aus solchen vorhergesagt werden können, die zeitlich vorgelagert ( $t_1$ ) erhoben wurden.

Die in Tabelle 7.2 als Studie S8 ausgewiesene Panelstudie umfasst die Angaben von Studierenden zu verschiedenen Aspekten der Studienqualität nach der Studieneingangsphase (1./2. Fachsemester) und nach der Übergangsphase zwischen Grund- und Hauptstudium (5./6. Fachsemester). Als relevante Faktoren der studentischen Zufriedenheit in diesen Studienphasen wurden in Voruntersuchungen neben den oben ausführlich beschriebenen, auch die (rückblickende) Beurteilung der Organisation der Studieneingangsphase bzw. des bisherigen Studiums sowie die Bereitschaft der Dozenten, sich mit den Problemen der Studierenden auseinander zu setzen, identifiziert (vgl. Engel 2000; Engel und Krekeler 2001; Engel und Pohlenz 2001).

Diese beiden Variablen sowie die oben bereits diskutierte Erwartungsbilanz wurden in beiden genannten Befragungen in gleicher Weise erhoben. Es lagen also von den im Paneldesign befragten Studierenden zu beiden Messzeitpunkten Daten vor, so dass sie für die nachfolgend dargestellte Stabilitätsanalyse genutzt werden konnten.

Latente Variable	Indikatorvariablen	Kürzel (Strukturgleichungsmodell)
Messzeitpunkt $t_1$ ( $\xi$ )	- Organisation der Studieneingangsphase ( $x_1$ ) - Engagement der Dozenten ( $x_2$ ) - Erwartungsbilanz ( $x_3$ )	- Org - Doz - Erwa
Messzeitpunkt $t_2$ ( $\eta$ )	- Organisation der Studieneingangsphase ( $y_1$ ) - Engagement der Dozenten ( $y_2$ ) - Erwartungsbilanz ( $y_3$ )	- Orga - Doze - Erwar

Tabelle 9.2: Latente Variablen und Indikatoren (Stabilitätsanalyse)

Im folgenden Strukturgleichungsmodell bilden diese drei Variablen jeweils die Indikatorvariablen für die globale Beurteilung der Studienqualität zum ersten ( $t_1$ ) und zum zweiten ( $t_2$ ) Messzeitpunkt. Tabelle 9.2 zeigt die einbezogenen exogenen und endogenen latenten Variablen, deren Indikatoren und die im Pfaddiagramm (Abbildung 9.4) benutzten Kürzel. Der erste Messzeitpunkt stellt in dem Modell die exogene latente Variable ( $\xi$ ) dar, über die die Stabilität der zu Protokoll gegebenen Einschätzungen bis zum zweiten Messzeitpunkt (endogene latente

Variable  $\eta$ ) geschätzt werden soll. Die Nutzung des Ansatzes linearer Strukturgleichungsmodelle ergibt sich in diesem Zusammenhang daraus, dass dieser die Möglichkeit bietet, meßfehlerbereinigte Schätzungen der latenten Variablen vorzunehmen. Die Wiederholungsmessungen der genannten Indikatorvariablen sind durch etwaige Messfehler belastet, die eine Trennung der „wahren“ Stabilität in den Antworten einerseits von unreliablem Antwortverhalten (zufällige Übereinstimmungen bzw. Einstellungsänderungen über die Messzeitpunkte hinweg) andererseits unmöglich machen. Dem Ansatz linearer Strukturgleichungsmodelle liegt dagegen die Idee der Schätzung meßfehlerbereinigter (latenter) Konstrukte durch die ihnen zugeordneten meßfehlerbehafteten manifesten Indikatorvariablen zugrunde (Engel und Reinicke 1994, S. 31 ff.), so dass etwaige Messfehler wie die genannten in den Indikatorvariablen berücksichtigt werden.

### 9.3.2.2 Modellspezifikation

Analog zu der Modellspezifikation, die für die oben berichteten Faktoren- und linearen Strukturgleichungsmodelle vorgenommen wurde, werden auch hier die Strukturgleichungen der Mess- und Strukturmodelle für das Modell zur Untersuchung der Längsschnittdaten in Matrixschreibweise ausgewiesen:<sup>11</sup>

(a) Messmodell der exogenen latenten Variablen ( $\xi$ ):

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(x)} \\ \lambda_{21}^{(x)} \\ \lambda_{31}^{(x)} \end{bmatrix} * [\xi] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \quad (9.9)$$

(b) Messmodell der endogenen latenten Variablen ( $\eta$ ):

---

<sup>11</sup>Auf die Darstellung der verallgemeinernden Schreibweisen der jeweiligen Matrizen wird verzichtet, sie sind mit den für die oben dargestellten Modelle identisch.



$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^{(y)} \\ \lambda_{21}^{(y)} \\ \lambda_{31}^{(y)} \end{bmatrix} * [\eta] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (9.10)$$

(c) Strukturkomponente:

$$\eta = [\gamma_{11}] * [\xi] + [\zeta] \quad (9.11)$$

### 9.3.2.3 Parameterschätzung

Die Schätzung des Modells mittels der ML-Methode erbrachte einen guten Modell-Fit. Mit einem  $\chi^2$ -Wert von 4,18 bei 6 Freiheitsgraden ( $p = .653$ ) kann dem Modell eine gute Anpassung an die empirischen Korrelationen bescheinigt werden. Dieser Befund wird durch die hohen Modell-Fit Indizes bestätigt (CFI = 1.00; RMSEA = .000, mit einem Konfidenzintervall von .000 - .071). Abbildung 9.3 zeigt die geschätzten Parameter (standardisierte Schätzer).

Die Beziehung zwischen den beiden latenten Variablen ist stark positiv korreliert (.77). Die Beurteilung der Studienorganisation, der Bereitschaft der Lehrenden, sich mit den Problemen der Studierenden auseinander zu setzen und die rückblickende Erwartungsbilanz, zu der die Befragten zum ersten Befragungszeitpunkt kommen, hat entsprechend eine starke prognostische Relevanz für die entsprechenden Einschätzungen zum zweiten Befragungszeitpunkt. Darüber hinaus zeigt die Messfehlervarianz der abhängigen Variable (t2) von .59 an, dass ein beträchtlicher Anteil von 41% der Varianz in t2 durch die Beziehung zu den Einschätzungen, die zum Messzeitpunkt t1 abgegeben werden, erklärt werden kann. Die hohen Korrelationen (Werte zwischen .52 und .71) und die quadrierten multiplen Korrelationen der Indikatorvariablen ( $x_{1-3}$  und  $y_{1-3}$ ) verweisen zusätzlich darauf, dass die latenten Variablen gut durch sie gemessen werden.<sup>12</sup>

<sup>12</sup>Zum Vorgehen bei der Durchführung der Stabilitätsanalyse vgl. bspw. Engel und Reinicke (1994).

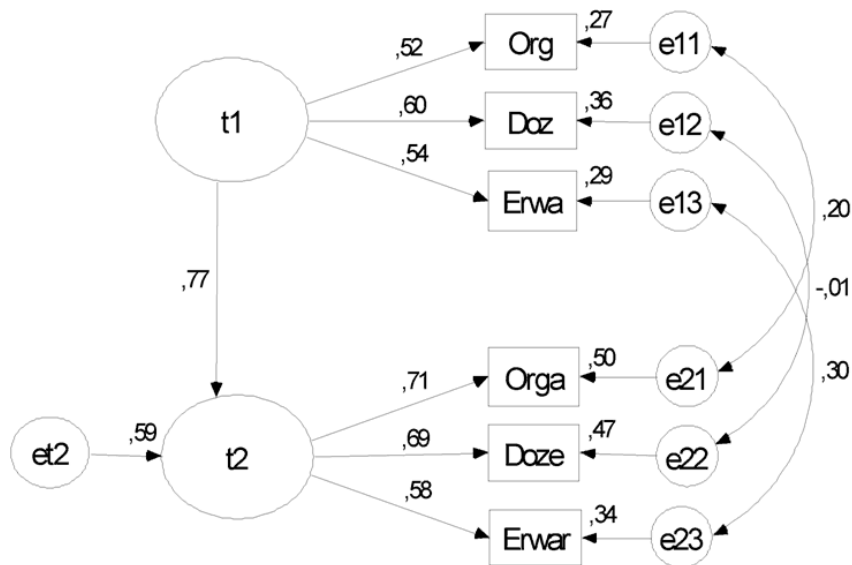


Abbildung 9.4: Studentische Qualitätsurteile im zeitlichen Längsschnitt

Diese Reproduzierbarkeit der Befragungsergebnisse über den Zeitverlauf hinweg spricht für eine hohe Reliabilität der Konstrukte, die im vorliegenden Modell der Lehrevaluation zur Beschreibung der studentischen Qualitätseinschätzungen zu Lehre und Studium herangezogen wurden. Die empirische Überprüfung des Vorliegens reliabler Messungen der zur Qualitätsbeschreibung herangezogenen Konstrukte ist ein wichtiger Bestandteil der Analysen zur Datenqualität. Insbesondere die längsschnittlich orientierten Daten leisten hier einen überaus relevanten Beitrag, weil sie methodologisch anspruchsvolle Stabilitätsanalysen ermöglichen, die ihrerseits belastbare Ergebnisse zur Frage der einschlägigen Testgütekriterien liefern. Diese sind wiederum die Grundlage für eine rationale Beurteilung der Qualität von Lehre und Studium und für die Entwicklung künftiger Verbesserungsstrategien.

### 9.3.3 Ein Ausblick: Längsschnittdaten und Qualitätsentwicklung

Über die methodologische Bedeutung von Längsschnittdaten für die Überprüfung von Testgütekriterien hinaus, haben diese eine relevante Funktion für die Qualitätsentwicklung von Studium und Lehre. Diese ergibt sich aus einer veränderten Qualitätsvorstellung der akademischen Ausbildung, die ihrerseits in einer steigenden Lernerzentrierung und der Bildung von praxis- und berufsrelevanten Handlungskompetenzen besteht. Ziel der Ausbildung ist inzwischen weniger die Akkumulation fester Wissensbestände, sondern die Fähigkeit, Anwendungswissen zu erwerben und die Fähigkeit zu entwickeln, (lebenslanges) Lernen zu lernen. Dementsprechend bezieht sich auch die Beurteilung der Qualität der Ausbildung zunehmend auf den Ausbildungsprozess, der die entsprechenden Kompetenzen vermittelt. Die längsschnittliche Analyse von Studienbiographien eröffnet dementsprechend die Möglichkeit, einerseits auf der Individualebene Studienverläufe nachzuzeichnen und andererseits, die Bedingungen von Lernerfolg bzw. -misserfolg zu ermitteln. Ersteres hat die Implikation, neue Lerntechnologien und -paradigmen durch eine methodisch anspruchsvolle, auf den zeitlichen Verlauf bezogene Evaluation zu begleiten. Diese Lernparadigmen beziehen sich auf die genannte Lernerzentrierung, die Förderung der Selbstreflexion über den eigenen Lernfortschritt sowie den bereits genannten Erwerb von Handlungskompetenzen.

Der Beitrag von Längsschnittdaten für die Qualitätsentwicklung kann dementsprechend darin bestehen, Lernerfolg im Sinne einer Veränderung (bspw. hinsichtlich des Wissenszuwachses oder der Lernprogression) zwischen zwei Zeitpunkten zu dokumentieren und zur Grundlage für eine Beurteilung sowohl der individuellen Lerndispositionen, als auch der Wirkungen der Lehre zu machen.

# Kapitel 10

## Rationalität studentischer Qualitätsurteile

Ein weiterer wichtiger Validitätsaspekt neben der Frage nach der Konstruktvalidität der zugrunde gelegten Qualitätsdimensionen ist das Maß, zu dem studentischen Urteilen bescheinigt werden kann, dass sie sachbezogen die zu beurteilenden Eigenschaften von Lehre und Studium widerspiegeln. Wie die Zusammenfassung der diesbezüglichen Kontroverse gezeigt hat (vgl. Kapitel 6), herrscht in diesem Zusammenhang keineswegs Einigkeit über die Qualität der landläufig erhobenen Daten. Vielmehr stehen diese in dem Verdacht, durch subjektiv gefärbte sowie sachfremde Urteilskomponenten verzerrt, mithin in ihrer Validität bedroht zu sein. Als prototypisches Beispiel für subjektiv gefärbte Urteilkategorien können bspw. das individuelle studentische Interesse oder die Studienmotivation angeführt werden. Diese beeinflussen die studentische Eindrucksbildung mit, ohne auf die zu beurteilenden, rational als sach- bzw. qualitätsbezogen zu qualifizierenden Eigenschaften von Lehre und Studium zurückführbar zu sein. Als sachfremde Urteilskomponenten sind weiterhin bspw. persönliche Eigenschaften und Hintergrundmerkmale der Studierenden, wie deren Alter und Geschlecht oder der elterliche Bildungshintergrund zu kennzeichnen. Werden die studentischen Qualitätseinschätzungen durch derlei Faktoren beeinflusst, so lässt dies den Schluss zu, dass die Urteile in ihrer Validität - die sich in diesem Zusammenhang ja gerade in der Sachbezogenheit der Urteile ausdrückt - bedroht sind und für eine

---

Verwendung für Steuerungszwecke durch die Hochschulen nur sehr bedingt nutzbar erscheinen.

Als Beispiel für die praktische Bedeutung dieses Validitätsaspekts lässt sich das vom Centrum für Hochschulentwicklung (CHE) durchgeführte Hochschul-Ranking anführen, welches in den letzten Jahren stetig an öffentlicher Aufmerksamkeit gewonnen hat. In diesem werden zur Ermittlung der studentischen Studienzufriedenheit verschiedene Einzelindikatoren erhoben und ein Globalurteil zur „Gesamtzufriedenheit“ mit der Studiensituation im studierten Fach gerankt. Es stellt sich die Frage, inwieweit dieses Globalurteil, welches in der öffentlichen Diskussion als Leistungsindikator der Hochschulen wahrgenommen wird und den Hochschulen über die Rangreihung mehr oder weniger vorteilhafte Ranglistenplätze zuweist, durch die genannten subjektiven bzw. sachfremden Einflussfaktoren in seiner Validität und damit Aussagefähigkeit eingeschränkt ist.

Da die Daten des CHE-Rankings für Sekundäranalysen nicht zur Verfügung standen, werden im Rahmen dieser Arbeit Analyseergebnisse zusammenfassend dargestellt, die mit den im „Potsdamer Modell der Lehrevaluation“ erhobenen Daten durchgeführt und an anderer Stelle berichtet wurden (Engel 2001; Engel und Pohlenz 2001; Pohlenz und Tinsner 2004; Pohlenz und Hagenmüller 2005; Pohlenz 2002; Pohlenz 2003). In diesen wurden in vergleichbarer Weise studentische Qualitätseinschätzungen zu verschiedenen Aspekten des Studiums erhoben sowie die Frage nach der globalen Studienzufriedenheit gestellt. Mittels Regressionsgleichungen wurden die erklärungskräftigen Einflussfaktoren des studentischen Globalurteils zur Studienzufriedenheit identifiziert. Diese können - nachdem sie auf statistischem Wege ermittelt wurden - inhaltlich hinsichtlich ihrer Sachbezogenheit bzw. des Ausmaßes, in dem sie einen subjektiv verzerrenden Einfluss auf das Gesamturteil nehmen, diskutiert werden.

Die herangezogenen Daten unterscheiden sich von denjenigen Erhebungen, auf die sich die in Kapitel 6 beschriebene „Student Ratings Debate“ bezieht (vgl. zusammenfassend z.B. Wolf u. a. 2001; Spiel und Gössler 2000). Während in dieser vorwiegend Qualitätseinschätzungen auf der Ebene von einzelnen Lehrveranstaltungen zur Diskussion stehen bzw. standen und die abgegebenen Urteile

entsprechend solche zur konkreten Lehrpräsentation sind, schließen die Daten des Potsdamer Evaluationsmodells (ebenso wie die für das CHE-Ranking erhobenen Daten) die Ebene der Studiengänge ein. Gefragt wird nach der studentischen Einschätzung der allgemeinen Studiensituation, der Organisationsqualität, u.ä. (vgl. Kapitel 7). Derartige Befragungen zur allgemeinen Studienqualität erfahren im Zuge der derzeitigen Diskussion um die Leistungen der Hochschulen und deren Beurteilung eine steigende Bedeutung, weshalb in der vorliegenden Arbeit auch die genannten, bereits berichteten Analyseergebnisse zur Rationalität studentischer Urteile erneut zusammengefasst werden sollen.

## 10.1 Analysen zur Sachbezogenheit studentischer Urteile

In der Rationalität bzw. Sachbezogenheit der studentischen Einschätzungen wird eine zentrale Voraussetzung für ihre Nutzung für Qualitätsbeurteilungen von Lehre und Studium gesehen. Dabei wird unter Rationalität verstanden, dass die Urteile „*rational in identifizierbaren Eigenschaften des zu beurteilenden Gegenstandes, beispielsweise des Studienfachs*, begründet sind“ (Engel und Pohlenz 2001, S. 137, Kursivsetzung im Original). Validitätsbedrohungen, als Resultat von Abweichungen von dieser Forderung ergeben sich potenziell daraus, dass sich in den studentischen Urteilen sachfremde Urteilskomponenten widerspiegeln. Diese gehen in das Urteil ein, hängen aber mit dem Beurteilungsgegenstand inhaltlich nicht zusammen. Derartige Urteilskomponenten oder sachfremde Variablen werden als Biasvariablen bezeichnet. „Sie wirken sich systematisch auf das Urteil aus, ohne selbst aus den perzipierten Merkmalen des (beurteilten; *P.P.*) Gegenstandes ableitbar zu sein“ (Engel und Pohlenz 2001, S. 138). Als typische Biasvariablen werden sowohl persönliche Hintergrundmerkmale, als auch subjektiv gefärbte Urteilskomponenten angeführt (s.o.) und zum Gegenstand von Untersuchungen gemacht. Als sachbezogene Urteilsaspekte können dagegen solche bezeichnet werden, die die angezielten Qualitätskonzeptionen auch tatsächlich abbilden, wie bspw. die oben beschriebenen Indikatoren der latenten Qualitätskonstruktionen

(„Vermittlung fachübergreifender Qualifikationen“ zur Beschreibung der „Interdisziplinarität“ des Studienganges).

Insbesondere an den genannten subjektiv gefärbten Urteilkategorien zeigt sich die potenziell resultierende Validitätsbedrohung deutlich: das Interesse, das dem Studium entgegengebracht wird, ist nicht allein ein Merkmal des Studiums selbst. Vielmehr spiegeln sich in ihm Eigenschaften der beurteilenden Person wider. Etwas, was für eine Person interessant sein kann, ist möglicherweise für eine andere Person gänzlich uninteressant und führt zu einer entsprechenden Beurteilung (Engel und Pohlenz 2001). Angesichts der offenkundigen Unmöglichkeit, alle individuellen Interessen durch ein Studienangebot zu befriedigen (und den diesbezüglichen Erfolg in den entsprechenden Befragungen zur Studienzufriedenheit attestiert zu bekommen), stellt sich für die an validen und damit im Sinne einer künftigen Hochschulentwicklung belastbaren Informationen interessierten Hochschulleitungen die Frage, inwieweit sich derartige Verzerrungen auf die studentischen Urteile auswirken. Diese Frage zu beantworten, war das Ziel der genannten (Vor-)Untersuchungen zur Frage des Einflusses von Biasvariablen auf das studentische Urteil.

## **10.2 Erklärung der Determinanten der Studienzufriedenheit**

Zur Identifikation erklärungskräftiger Determinanten der (positiven bzw. kritischen) Gesamtbeurteilung des Studiums wurden in diesen Analysen (multiple) lineare Regressionsmodelle gerechnet. Diese sind geeignet, die Varianz in einer abhängigen, zu erklärenden Variable auf den Einfluss einer oder mehrerer unabhängiger (erklärender) Variablen zurückzuführen und den jeweiligen Einfluss zu quantifizieren (vgl. z.B. Backhaus u. a. 2000). Nachfolgend werden die wichtigsten Ergebnisse vorgestellt.

### 10.2.1 Einschätzungen auf der Ebene der Studiengänge

Die von Engel (2000), Engel und Krekeler (2001) sowie Engel und Pohlenz (2001) berichteten Analysen beziehen insbesondere die in Kapitel 7 beschriebenen, im Panel-/Kohortendesign erhobenen Daten ein (vgl. Tabelle 7.2, Studien S1 - S6). Dabei werden verschiedene Einflussfaktoren als erklärungskräftig für die studentische Gesamtbeurteilung<sup>1</sup> der Studienqualität identifiziert.<sup>2</sup> Im Falle der Studierenden der Übergangsphase zwischen Grund- und Hauptstudium<sup>3</sup> handelt es sich dabei um die subjektive Erwartungsbilanz der Studierenden, die rückblickende Beurteilung der Organisationsqualität des Studiums sowie das wahrgenommene Engagement der Lehrenden für die Belange der Studierenden (Engel und Pohlenz 2001, S. 140). Diese drei Aspekte (die sich auch in der oben berichteten Stabilitätsanalyse als prädiktiv für die Stabilität der studentischen Urteile erwiesen) erklärten in einem gemeinsamen Regressionsmodell rund 44 % der Varianz in der studentischen Gesamtbeurteilung des Studiums (a.a.O.). Tabelle 10.1 zeigt die Parameter der Regressionsgleichung (zit. nach Engel und Pohlenz 2001, S. 140).<sup>4</sup>

Das Maß, zu dem die Erwartungen der Studierenden als erfüllt bezeichnet wurden, hing seinerseits von den verschiedenen Gründen, die für die Studienfachwahl als ausschlaggebend bezeichnet worden waren, von den (Einzel-)Erwartungen, die an das Studium gerichtet wurden sowie von der Einschätzung, das Studium in der vorgesehenen Regelstudienzeit abschließen zu können, ab (a.a.O.).<sup>5</sup>

---

<sup>1</sup>Analog zur Vorgehensweise bei der Bestimmung des durch Biasvariablen erklärten Varianzanteils bei studentischen Beurteilungen von *Lehrveranstaltungen* wurde auch hier das Kriterium der Gesamtbeurteilung der Studienqualität durch die Studierenden herangezogen (vgl. Rindermann 1997; Spiel und Gössler 2000).

<sup>2</sup>Die studentische Gesamtbeurteilung des Studiums war auf einer 11stufigen Ratingskala mit den Polen „sehr schlecht“ (1) und „sehr gut“ (11) abgefragt worden.

<sup>3</sup>Befragt wurden 519 Studierende; vgl. auch Abschnitt 9.1 und Tabelle 7.2.

<sup>4</sup>Die Effekte der berichteten Prädiktoren gingen in die erwartete Richtung, die Schätzungen waren statistisch signifikant ( $p \leq .05$ ).

<sup>5</sup>Das Studium wurde positiver beurteilt, wenn die Gründe für die Studienfachwahl in einer dem Studium entsprechenden Neigung und Begabung, der Erwartung einer beruflichen Quali-



Variable	b	$\beta$	t
Wurden Ihre Erwartungen (an das Studium) bislang erfüllt? „nein überhaupt nicht“ (1) → „ja, absolut“ (11)	.44	.45	12,45
Wie beurteilen Sie rückblickend die Organisation des Grundstudiums in Ihrem Hauptfach? „sehr schlecht“ (1) → „sehr gut“ (11)	.22	.26	6,83
Wie schätzen Sie die Bereitschaft der Dozentinnen bzw. Dozenten ein, sich mit fachlichen und organisatorischen Problemen der Studierenden auseinander zu setzen? „sehr gering“ (1) → „sehr hoch“ (11)	.14	.17	4,56

Tabelle 10.1: Prädiktoren des Gesamturteils Studierender

Die Faktoren zur Erklärung der Organisationsqualität, die ihrerseits als Prädiktor des Gesamturteils identifiziert worden war, finden sich bei der Globalbeurteilung der Lehrveranstaltungen, in der Strukturierung des Studiums und in der wahrgenommenen Qualität der Organisation der Zwischenprüfungen (Transparenz der Anforderungen, organisatorischer Aufwand).<sup>6</sup> Aus den Ergebnissen dieser

fizierung sowie in einer Ablehnung der Aussage „ein Studium ist das kleinste Übel“ bestanden. Weiter trugen die Erwartung einer durch das Studium geförderten Selbstreflexion positiv und die Erwartung, Fähigkeiten zum Betreiben von Forschung vermittelt zu bekommen negativ zur Erwartungsbilanz bei. Schließlich wird diese positiv beeinflusst, wenn erwartet wird, das Studium fristgerecht abschließen zu können, mit der Begründung, dass das Studium gut organisiert sei. Negativ wird sie beeinflusst, wenn aufgrund einer als schlecht wahrgenommenen Organisationsqualität oder aufgrund einer verfehlten Studienfachwahl davon ausgegangen wird, das Studium nicht innerhalb der Regelstudienzeit abschließen zu können. Diese Prädiktoren zur Erklärung der Erwartungsbilanz tragen gemeinsam allerdings lediglich knapp 12 Prozent zur Varianzaufklärung in der Erwartungsbilanz bei. Diese ist demnach nur sehr bedingt aus den genannten studienbezogenen Aspekten vorhersagbar (Engel und Pohlenz 2001, S. 142).

<sup>6</sup>Die Organisationsqualität wird gut bewertet, wenn die Organisation der *Lehrveranstaltungen* insgesamt als positiv bewertet wird und wenn diese ein interessantes Studienangebot vermitteln. Sie wird dagegen kritisch bewertet, wenn die Lehrveranstaltungen als zu groß (bezüglich der Teilnehmerzahl) eingeschätzt werden. Ferner spielen Organisationsaspekte wie die zeitliche Abstimmung des Lehrangebots, die Transparenz der Leistungs- und Prüfungsanforderungen sowie die Strukturierung und der zu betreibende zeitliche und organisatorische Aufwand bei

Analysen wird geschlussfolgert, dass die studentischen Urteile in Teilen durchaus sachbezogene Einschätzungen der Studienqualität reflektieren, aber eben auch durch Faktoren beeinflusst werden, die nicht unmittelbar in den Eigenschaften des beurteilten Gegenstandes zu finden sind (Engel und Pohlenz 2001). In der Transparenz von Leistungsanforderungen oder der zeitlichen Abstimmung des Lehrangebots sind sicher Urteilkategorien zu sehen, die durchaus rationale und auf den Beurteilungsgegenstand bezogene Qualitätsurteile widerspiegeln. Skeptischer sind dagegen das Studieninteresse und die subjektive Erwartungsbilanz zu sehen, die ebenfalls einen Einfluss auf die studentische Eindrucksbildung hatten. Die diesbezüglichen Einschätzungen spiegeln auch Eigenschaften der beurteilenden Personen wider und sind dementsprechend als potenziell subjektiv verzerrte Biasvariablen anzusehen, die ihrerseits in Validitätsbedrohungen resultieren (a.a.O.).

### 10.2.2 Rückblickende Urteile von Studienabbrechern und Absolventen

Ein ähnliches Bild stellt sich im Falle der Exmatrikulierten und Absolventen dar. Quer zu den Befragungswellen des Potsdamer Modells der Lehrevaluation wurden im Studienjahr 2001/02 erstmals ehemalige Studierende unterschiedlicher Semester (vgl. Tabelle 7.2, Studie S7), die entweder als Studienabbrecher, Hochschulwechsler oder Absolventen die Universität verlassen hatten, zu ihrer rückblickenden Einschätzung der Studienqualität befragt (Pohlenz und Tinsner 2004).

Auch hier wurden die Determinanten der Studienzufriedenheit (Gesamtbeurteilung des Studiums) mittels multipler linearer Regressionsanalysen ermittelt (a.a.O.). Die Tabellen 10.2 und 10.3 zeigen die erklärungskräftigen Prädiktoren zur Erklärung des Gesamturteils der Studienabbrecher bzw. Absolventen

---

der Vorbereitung von Prüfungen eine prädiktive Rolle für die Erklärung der Beurteilung der Organisationsqualität des Studiums.

(zit. nach Pohlenz und Tinsner 2004, S. 109 ff.).<sup>7</sup> Das Qualitätsurteil der Studienabbrecher wird durch deren Studienabbruchgründe, durch Aspekte der Studienbedingungen sowie durch die (offenkundig enttäuschende) Wahl des Studienortes, sofern diese auf einer Empfehlung anderer Studierender beruhte, bestimmt.<sup>8</sup> Gemeinsam erklären diese Einflussgrößen 77 % der Varianz in der Qualitätseinschätzung durch die Studienabbrecher.

Variable	b	β	t
<i>Gründe für den Studienabbruch:</i>			
Studienbedingungen sind schlecht	.49	.25	3,47
Erwartungen an das Studium wurden nicht erfüllt	.55	.31	4,57
Betreuung durch die Lehrenden ist schlecht	.54	.25	3,58
Prüfungsängste	.48	.17	2,59
<i>Studienbedingungen:</i>			
Öffnungszeiten der Bibliothek	.21	.26	3,86
Qualität studentischer Beiträge (Referate, etc.)	.16	.18	2,56
<i>Wahl des Studienortes:</i>			
Empfehlungen anderer Studierender	-.51	-.20	-3,12

Tabelle 10.2: Prädiktoren des Gesamturteils der Studienabbrecher

Die Einschätzungen der Absolventen werden weit gehend durch Aspekte bestimmt, die die Qualität der Studienorganisation widerspiegeln und durch die objektive Probleme des Studienalltags in unterschiedlicher Weise angesprochen werden (Pohlenz und Tinsner 2004). Gleichwohl zeigen sich auch hier Faktoren

<sup>7</sup>Die Effekte der berichteten Prädiktoren gingen in die erwartete Richtung, die Schätzungen waren statistisch signifikant ( $p \leq .05$ ).

<sup>8</sup>Das Studium wird negativer eingeschätzt, wenn die Abbruchgründe in den als schlecht wahrgenommenen Studienbedingungen und Betreuungsleistungen der Lehrenden, insgesamt enttäuschten Erwartungen und in Prüfungsängsten bestand. Zu den erklärungskräftigen Aspekten der Studienbedingungen zählen die Öffnungszeiten der Bibliothek und die studentischen Beiträge zum Gelingen der Lehrveranstaltungen (die Gesamtbeurteilung fällt kritischer aus, wenn diese Aspekte negativ beurteilt werden).

als erklärungskräftig, die nicht unmittelbar auf die Qualität des Studiums zurückgeführt werden können. Das „Job-Angebot in der Region“ ist vermutlich unter diese Kategorie von Determinanten zu subsumieren (vgl. Tabelle 10.3).<sup>9</sup>

Im Sinne einer Nutzung von Evaluationsergebnissen durch Hochschulleitungen einerseits und vor dem Hintergrund von zunehmend auch auf rückblickende Qualitätseinschätzungen von Absolventen aufgesetzte Leistungsbeurteilungen, ist hier sicher festzustellen, dass Hochschulen auf die regionale Arbeitsmarktsituation nur sehr bedingt unmittelbaren Einfluss nehmen können und insofern keinen Handlungsspielraum haben, um die entsprechende studentische Eindrucksbildung zu verbessern. Die - offenkundig urteilswirksame - Beurteilung der diesbezüglichen Situation stellt also eine potenziell verzerrende Komponente zur Beurteilung der Hochschulleistungen in Lehre und Studium dar, die gleichzeitig durch die Hochschulen nicht zu verantworten ist (vgl. auch Engel und Pohlenz 2001). Gleiches gilt für viele Urteilskomponenten, die sich auf die Ausstattung der Hochschulen, deren kulturelles Umfeld, und ähnliche Indikatoren beziehen. Diese werden vielfach erhoben, um die studentische Studienzufriedenheit zu messen. Sofern die entsprechenden Messergebnisse mit einem Urteil über die Leistungsstärke der Hochschule gleichgesetzt werden, muss jeweils diskutiert werden, inwieweit derlei Indikatoren überhaupt im Rahmen der Einflussmöglichkeiten entsprechend beurteilter Hochschulen liegen, mithin ein valides Beurteilungskriterium ihrer Leistungen darstellen können. Da das studentische Gesamturteil zur Studienzufriedenheit zumeist eine Verdichtung der Messergebnisse verschiedener derartiger Indikatoren darstellt, muss entsprechend diskutiert werden, inwieweit die Validität des Globalurteils durch den Einfluss von Biasvariablen insgesamt potenziell bedroht ist.

Im Falle der Studienabbrecher wurde deren Gesamtbilanz zur Studienqualität als Grundlage für die Entscheidung aufgefasst, das Studium abzubrechen

---

<sup>9</sup>Dabei ist zu beachten, dass sich die entsprechende Frage auf das Angebot von typischen „Studentenjobs“, nicht aber auf das für die Absolventen *nach* Verlassen der Hochschule relevante Arbeitsmarktsegment akademischer Tätigkeiten bezog.

Variable	b	$\beta$	t
<i>Allgemeine Studienbedingungen:</i>			
inhaltliche Abstimmung des Studienangebots	.26	.27	4,83
Vermittlung fachübergreifender Qualifikationen	.12	.16	2,77
Beratung/Betreuung im Prüfungsvorfeld	.07	.12	2,07
<i>Rahmenbedingungen des Studiums:</i>			
Jobangebote in der Region	.12	.15	2,78
<i>Aspekte der Lehrveranstaltungen</i>			
Qualität der studentischen Beiträge	.22	.30	4,92
Vorbereitung der Lehrenden	.15	.18	2,96

Tabelle 10.3: Prädiktoren des Gesamturteils der Absolventen

(Pohlenz und Tinsner 2004, S. 117). Insofern stellen die in diesem Zusammenhang erhobenen Daten eine wichtige Informationsquelle für die Hochschulleitungen dar, die daran interessiert sind, auf welche Bestimmungsgründe sich der Studienabbruch zurückführen lässt. Diese Informationen sind sowohl für die interne Entwicklungsplanung als auch für die Leistungspräsentation nach außen von größter Bedeutung. So verlässt sich bspw. auch das bereits erwähnte Hochschulranking des CHE auf eine Gesamtbeurteilung der Studienqualität durch eingeschriebene Studierende als Informationsgrundlage zur Leistungsbeurteilung von Hochschulen, die sich ausdrücklich auch an künftige Studienanfänger richtet, mit dem Ziel, diesen die Wahl des Studienortes zu erleichtern (Berghoff u. a. 2004; Berghoff u. a. 2006). In der Konkurrenz zu anderen Standorten befindliche Hochschulen sollten dementsprechend bestrebt sein, Kenntnisse über die Determinanten der studentischen Qualitätseinschätzungen für etwaige Qualitätsverbesserungen zu nutzen. Gleichzeitig ist aber in der hochschulpolitischen Diskussion auch darauf zu verweisen, dass die studentischen Gesamturteile zur Studienqualität wie oben gezeigt auch auf subjektiv gefärbte und damit potenziell in ihrer Validität bedrohte Urteils Komponenten zurückgehen. Aus vorliegender Sicht ist aufgrund der vorgelegten Befunde zur Datenqualität bei einer Nutzung der Befragungsergeb-

nisse für Leistungsbeurteilungen zu beachten, dass die potenziell verzerrenden Einflussfaktoren des studentischen Qualitätsurteils angemessen identifiziert und aus den entsprechenden Leistungsbeurteilungen ausgeschlossen werden.

Dies ist durch den Einsatz der entsprechenden statistischen Verfahren zwar durchaus möglich, stellt aber einen zusätzlichen methodischen Aufwand für die Praxis der Lehrevaluation dar. Belohnt würde dieser Aufwand durch das Vorliegen einer belastbaren und - sofern sie für vergleichende Leistungsbeurteilungen benutzt wird - gerechten Datenbasis. Verfahren, wie die oben beschriebenen Kausalanalysen und die hier vorgestellten Regressionsanalysen sind geeignet, die unterschiedlichen Einflussgrößen in ihren Wirkungszusammenhängen zu identifizieren. Sie stellen daher eine relevante Grundlage für die inhaltliche Diskussion dar, inwieweit bei der Beurteilung von Hochschulleistungen von Einflüssen potenziell invalider Urteilskomponenten auszugehen ist und inwiefern die Aussagefähigkeit der erhobenen Daten ggf. eingeschränkt ist.

# Kapitel 11

## Diskussion und Ausblick

Ziel der in den vorangegangenen Kapiteln berichteten Analysen war es zu überprüfen, ob Daten, die zum Zweck der Evaluation von Lehre und Studium erhoben werden, die Qualität des beurteilten Gegenstandes (Studienqualität) sachgemäß und verzerrungsfrei darstellen. Die Ergebnisse der diesbezüglichen Berechnungen ergaben ein zweigeteiltes Bild. Die Befunde sind daher abschließend vor dem Hintergrund der Frage nach der Nützlichkeit von Evaluationsergebnissen für die Hochschulentwicklung zu diskutieren. Im Sinne eines Ausblicks ist darüber hinaus die Frage nach möglichen und nötigen Weiterentwicklungen der landläufig eingesetzten Evaluationsroutinen zu stellen.

### 11.1 Potenzielle Beeinträchtigungen von Validität und Reliabilität

Die Analysen zur Datenqualität von Evaluationen, die für Leistungsbeurteilungen der Hochschulen eingesetzt werden, stützten sich sowohl auf studentische Befragungsdaten, als auch auf Daten der amtlichen Hochschulstatistik. Die jeweils spezifischen Quellen möglicher Bedrohungen der Datenqualität, die sich als Einschränkungen von Validität und Reliabilität auswirken, werden nachfolgend diskutiert.

### 11.1.1 **Studentische Qualitätsurteile**

#### 11.1.1.1 **Dimensionierung und Kontextsensitivität der Urteile zur Studienqualität**

In Befragungen Studierender zu ihrer Einschätzung der Studienqualität werden ihnen üblicherweise Fragebögen vorgelegt, mittels derer verschiedene Beurteilungskategorien abgefragt werden. Dies impliziert, dass die Befragenden ein a priori formuliertes theoretisches Modell der Studienqualität entwickelt haben, in dem diese in verschiedenen Urteilskomponenten und -dimensionen abgebildet werden. Im Fall der in den Kapiteln 8 und 9 beschriebenen Reanalysen von Evaluationsdaten handelte es sich dabei um die Beurteilungsdimensionen Studierbarkeit, Strukturierung, Interdisziplinarität und Wissenschaftlichkeit, die als relevante Kennzeichen eines guten Studiums festgelegt worden waren und die über verschiedene Indikatoren abgebildet werden sollten.

Vielfach werden studentische Urteile zur Beschaffenheit des Studiums hinsichtlich derartiger Urteilskomponenten für Leistungsbeurteilungen genutzt. In diesem Fall ist es nötig sicher zu stellen, dass die als gegeben unterstellte Urteilsrelevanz der in Erhebungsinstrumenten vorgegebenen Qualitätsdimensionen sich auch tatsächlich empirisch in diesen studentischen Urteilen wiederfinden lässt. Unter Umständen verbinden Studierende ein anderes Qualitätskonzept mit ihrem Studium, als die Personen, die den betreffenden Fragebogen entwickelt haben. In diesem Fall wäre zu schließen, dass die vorgegebenen Qualitätsdimensionen nicht geeignet sind, die studentischen Qualitätseinschätzungen in valider und reliabler Weise abzubilden. Ein solcher Befund würde seinerseits gegen eine Verwendung der Erhebungsdaten für Rückschlüsse aus den erhobenen studentischen Urteilen auf die Beschaffenheit des beurteilten Gegenstandes - die Studienqualität - sprechen. Die entsprechende Leistungsbeurteilung wäre in diesem Fall ein Artefakt der entsprechenden Datenerhebung.

Die Analysen haben gezeigt, dass zumindest in Teilen davon ausgegangen werden muss, dass die Qualitätsvorstellungen der Befragten von denen der Befragenden abweichen. Dies äußerte sich darin, dass in einem entsprechenden konfir-



matorischen Faktorenmodell nachgewiesen werden konnte, dass die formulierten Qualitätsdimensionen sich nur eingeschränkt im Urteil der Studierenden wiederfinden lassen. Insbesondere die Operationalisierung der Urteilsdimension „Wissenschaftlichkeit des Studiums“ scheint nicht die intendierte Beurteilung eben dieser Kategorie zu provozieren. Vielmehr kovariiert das entsprechende studentische Urteil mit den anderen vorgegebenen sowie mit weiteren, nicht identifizierten Urteilkategorien. Relevant ist dieser Befund, wenn aus den studentischen Urteilen zur „Wissenschaftlichkeit“ des Studiums auf die diesbezügliche Leistungsfähigkeit der Hochschule bzw. des jeweils beurteilten Studienganges geschlossen wird. Insofern die studentischen Urteile nicht in der erwarteten Form - oder Dimensionierung - das diesbezügliche Qualitätsurteil abbilden, sind die Einschätzungen als nicht hinreichend valide zu kennzeichnen und mithin als ungeeignet für die (Leistungs-)Beurteilung der Hochschule oder des Studienganges zu betrachten.

Generell zeigen die Analysen einmal mehr, dass unterschiedliche bzw. unterscheidbare Dimensionen in die studentische Beurteilung der Studienqualität eingehen. Die vorgelegten Auswertungen der Evaluationsdaten legen nahe, dass Evaluationen von Lehre und Studium bestrebt sein müssen, der Komplexität der studentischen Qualitätswahrnehmung gerecht zu werden. Wenn bereits innerhalb von mehrdimensionalen Modellen zur Erklärung der studentischen Qualitätseinschätzungen festgestellt werden kann, dass die einbezogenen Qualitätsdimensionen in unterschiedlich valider Weise die zugrunde gelegten Sachverhalte abbilden, darf bezweifelt werden, dass die Verwendung eines oder weniger Globalindikatoren, der genannten Komplexität der Urteile hinreichend gerecht wird. Gerade diese Reduzierung der Datenerhebung auf wenige, handhabbare Indikatoren wird aber vielfach gefordert, durch deren Einsatz der Zweck der Evaluation, bestehend in einer umseitigen Informierung der beteiligten Akteure, konterkariert würde.

Die festgestellten Beeinträchtigungen der Validität lassen sich aber auch auf die Kontextabhängigkeit der studentischen Qualitätsurteile zurückführen. Die vorgegebenen Qualitätsdimensionen erfahren in den verschiedenen Studienphasen unterschiedliche Hierarchisierungen und Wertungen. So beeinflusst ein als „zu

wissenschaftlich“ (vulgo: zu theoretisch, zu praxisfern) empfundenes Studium die Gesamtbeurteilung der Studierenden im Anfangsstadium negativ, während genau diese Qualitätsdimension im Falle von fortgeschrittenen Studierenden prädiktiv für die Zufriedenheit mit dem Studium ist. Insofern die Qualitätsdimensionen eine unterschiedliche Interpretation und Gewichtung durch Studierende verschiedener Studienphasen erfahren, stellt der Befund einer beeinträchtigten Validität für sich genommen keine Überraschung dar. Relevant wird diese Erkenntnis allerdings vor dem Hintergrund der Frage nach der Nützlichkeit von Evaluationsdaten für hochschulpolitisches Handeln und Steuern. Insofern sich dieses auf Daten und dabei eben auch: studentische Qualitätsurteile stützen will, muss die genannte Kontextabhängigkeit studentischer Urteile und die damit einhergehende Vielgestaltigkeit und teilweise sogar vorliegende Widersprüchlichkeit dieser Einschätzungen berücksichtigt werden. Möglicherweise wird das Procedere der Lehrevaluation durch die genannten Bedingungen und erforderlichen Untersuchungen zur Datenqualität komplexer. In dem Maße jedoch, wie gleichzeitig Zugewinne an inhaltlicher Relevanz sowie methodischer Seriosität der Daten gemacht werden können, scheint dieser Effekt verschmerzbar zu sein.

#### 11.1.1.2 Verzerrungen durch subjektive Qualitätsurteile

Auch die Analysen zur Ermittlung des Einflusses von sachfremden Urteilskategorien auf das studentische Qualitätsurteil vermitteln ein geteiltes Bild. Einerseits werden die Einschätzungen durchaus von Aspekten bestimmt, die rational in den die Qualität beschreibenden Eigenschaften des Studiums begründet sind. Bei diesen handelt es sich bspw. um auf die Organisationsqualität des Studiums bezogene Urteilskategorien. Gleichzeitig wird das studentische Urteil zur Qualität von Lehre und Studium aber auch durch Determinanten beeinflusst, die zum einen subjektiv gefärbt sind, wie individuelles Interesse und Studienmotivation und die zum anderen außerhalb des Handlungsspielraums der Universitäten liegen, wie Infrastrukturaspekte des Studienortes. Beide Variationsquellen des studentischen Urteils beeinträchtigen dessen Validität, wenn die Absicht der diesbezüglichen Messung darin bestand, durch die entsprechenden Daten Aussagen über die Stu-

dienqualität zu erheben. Durch das persönliche Studieninteresse wird vermutlich mehr über die urteilende Person als über den beurteilten Gegenstand ausgesagt. Durch allgemeine, regionale Rahmenbedingungen (das Gesamturteil der Absolventen zur Studienqualität wird bspw. durch ihre Einschätzung der Situation bezüglich vorhandener Studentenjobs in der Region mitbestimmt; vgl. Kap. 10) lässt sich nur sehr bedingt eine inhaltlich relevante Leistungsbeurteilung der Hochschule begründen, sondern bestenfalls eine Beurteilung des regionalen Kontextes, in dem die Hochschule beheimatet ist. Dieser mag sich zwar durchaus auf die Studiensituation auswirken, kann aber gleichzeitig nicht durch die Hochschule beeinflusst werden. Entsprechend sind die Hochschulen nicht für eine etwaige diesbezüglich negative Situation haftbar zu machen. Dies aber geschieht, wenn die studentischen Einstellungen zu den in diesem Sinne sachfremden Urteilkategorien unidentifiziert in die Leistungsbeurteilung eingehen.

Insofern also die genannten sachfremden Variationsquellen auf die studentischen Qualitätsurteile zu Lehre und Studium wirken, muss konstatiert werden, dass diese in ihrer Validität zumindest bedroht sind. Gleichzeitig ist es aber mittels der im Rahmen dieser Arbeit beschriebenen Verfahren möglich, die Determinanten der Qualitätseinschätzungen im Einzelnen zu identifizieren und das Ausmaß ihres etwaigen verzerrenden Einflusses zu quantifizieren.

Eine Ermittlung und Beschreibung des Einflusses, den *sachbezogene* Urteilkategorien auf die studentische Qualitätseinschätzung haben, kann dagegen die Bemühungen um die Qualitätsentwicklung von Lehre und Studium in wirkungsvoller Weise unterstützen. So lassen sich aus Erkenntnissen zur Frage nach den Determinanten des studentischen Urteils durchaus relevante Schlussfolgerungen ziehen. Dies trifft bspw. zu, wenn sich durch Regressionsmodelle wie die in Kap. 10 beschriebenen, der Einfluss der wahrgenommenen „Strukturiertheit“ des Studienangebots auf das Gesamturteil über Lehre und Studium ermitteln lässt. Unter der Bedingung, dass Klarheit über die Dimensionierung der entsprechenden Urteilsbestandteile herrscht, können handlungsleitende Informationen für künftige Verbesserungsstrategien zu Lehre und Studium gewonnen werden.

Die Frage nach der Validität und Reliabilität von Daten der Lehrevalua-

tion stellt sich aus vorliegender Sicht entsprechend nicht als Kriterium für oder gegen die Durchführung der entsprechenden Datenerhebungen dar. Vielmehr ist es das Anliegen dieser Arbeit, auf mögliche Beeinträchtigungen dieser Testgütekriterien hinzuweisen sowie zu versuchen, die entsprechenden Fehlerquellen zu identifizieren und in ihrem jeweiligen Einfluss zu beschreiben. Das Wissen über die Möglichkeit potenziell durch Validitätsbedrohungen beeinflusster Datenerhebungen trägt zu einer realistischeren Einschätzung der Aussagefähigkeit dieser Daten bei. Einschränkungen dieser Aussagekraft aufgrund von Validitätsbedrohungen sind bei der Interpretation und insbesondere bei der Nutzung für politische Steuerungsentscheidungen zu berücksichtigen. Werden in diese Entscheidungen Daten einbezogen, die in ihrer Validität beeinträchtigt sind, besteht die Gefahr von Fehlinterpretationen und Fehlentscheidungen.

#### 11.1.2 Hochschulstatistische Kennzahlen

Besonders groß ist diese Gefahr bei den in Kapitel 5 beschriebenen hochschulstatistischen Leistungskennzahlen, etwa zur Studienabbruchquote. Deren Potenzial zur adäquaten Leistungsbeschreibung in Forschung und Lehre wird - vermutlich wegen der von ihnen gesprochenen „klaren Sprache der Zahlen“ - landläufig kaum in Zweifel gezogen. Daraus wird in noch größerem Maße, als dies für studentische Befragungsdaten zutrifft, die Forderung abgeleitet, sie für eine Leistungsbeurteilung von Hochschulen einzusetzen und sogar Allokationsentscheidungen auf die entsprechenden Messungen aufzusetzen.

Im Lichte der beschriebenen Sekundärauswertungen von hochschulstatistischen Daten erweist sich die Hoffnung, mit Leistungsindikatoren, bspw. zur Studienabbruchquote ein leicht zu handhabendes Instrument zur Quantifizierung von Hochschulleistungen und damit der Hochschulqualität zu besitzen, als allzu optimistisch. Vielmehr wohnen auch den objektiven Kennzahlen zur Beschreibung von Hochschulleistungen, wie eben der Studienabbruchquote, potenzielle Beeinträchtigungen der Validität inne.

Das diesbezügliche Begriffsverständnis von Validität leitet sich zwar nicht wie im Falle der Erhebung subjektiver Qualitätseinschätzungen durch Studie-

rende direkt aus den Annahmen der klassischen Testtheorie ab, mit den methodologischen Implikationen, die sich daraus für eine diesbezügliche Überprüfung von Daten ergeben (vgl. Kapitel 6). Vielmehr liegt in diesem Zusammenhang eine inhaltlich begründete Anwendung der Bedeutung des Konzepts der Validität zugrunde. Diese bezieht sich darauf, dass im Falle des Vorliegens verzerrter Leistungskennzahlen, durch diese nicht das gemessen wird, was zu messen beabsichtigt war. Eine solche Verwendung des Begriffs Validität hinsichtlich der genannten Leistungskennzahlen entspricht dem oben beschriebenen, aus methodischer Sicht engeren Begriffsverständnis hinsichtlich seiner inhaltlichen Implikationen.

Die diesbezüglichen Auswertungen haben gezeigt, dass objektive Kennzahlen aufgrund methodischer Probleme ebenso mit Validitätsbedrohungen belastet sein können, wie studentische Befragungsdaten. Diese äußerten sich darin, dass die Schwundquote, durch eine fehlende Berücksichtigung bestimmter Sachverhalte - wie hier dem Wechsel zwischen den Fächern einer Universität - den Gegenstand, den zu messen sie beabsichtigt, nicht in der intendierten Weise abbildet. Durch die Schwundquote soll angegeben werden, in welchem Ausmaß Studienfächer, Studiengänge oder Universitäten Studierende verlieren, ohne dass diese den angestrebten Abschluss abgelegt haben. Entsprechende Quoten werden ihrerseits vielfach als hinlänglicher Beleg für die Ineffizienz der Hochschulen interpretiert. Dadurch, dass durch die berichtete Quote - wie in Kapitel 5 gezeigt - aber eine potenziell verzerrte Information vermittelt wird, muss die Validität dieser Quote im beschriebenen Sinne als gefährdet bzw. eingeschränkt bezeichnet werden.

Dies hat insofern gravierende Konsequenzen, als die Schwundquote in der öffentlichen Diskussion vielfach mit der Studienabbruchquote gleichgesetzt wird und sich entsprechend das öffentliche Bild ineffizienter Hochschulen verfestigt, mit den negativen Folgen, die dies für ihre Reputation mit sich bringt. Im Sinne des Anspruchs, der an die Hochschulstatistik zu stellen ist und der darin besteht, die öffentliche Diskussion mit belastbaren Informationen und Kennzahlen zu versorgen, ist sicher zu stellen, dass entsprechend verzerrende Faktoren ermittelt und aus den statistischen Berechnungen von Quoten ausgeschlossen werden.

Schon der relativ einfach zu ermittelnde Wechsel zwischen zwei Fächern

einer Universität erwies sich als geeignet, Überschätzungen der Schwundquote aufzudecken. Inwieweit das als Studienabbruch deklarierte Verlassen einer Hochschule tatsächlich eher als der Wechsel in einen anderen Studiengang einer anderen Hochschule zu kennzeichnen ist, kann mit den derzeitigen Instrumenten der Hochschulstatistik nicht hinreichend belastbar quantifiziert werden, weil keine bundesweit durchgeführte Studienverlaufsstatistik durchgeführt wird. Aber auch das Phänomen des Fachwechsels innerhalb einer Universität reicht offensichtlich aus, um Probleme der auf Bestandsstatistiken beruhenden Schwund-, Abbruch- und Absolventenquoten zu entlarven.

Die dargestellten Sekundäranalysen hochschulstatistischer Daten haben gezeigt, dass unter Einsatz geeigneter, insbesondere studienverlaufsstatistischer Instrumentarien, Quellen für potenzielle Verzerrungen identifiziert werden können. Auch hier ist zu betonen, dass die Anwendung dieser Instrumentarien zwar in einem steigenden Aufwand für die entsprechenden Berechnungen resultiert. Werden aber von der Erstellung hochschulstatistischer Datenmaterials auch Beiträge zu Qualitätssicherung und Qualitätsentwicklung von Lehre und Studium erwartet, so scheint es unerlässlich zu sein, diesen Aufwand zu treiben. Anderenfalls bewegt sich eine Beurteilung der Hochschulleistungen, wie hier hinsichtlich der erfolgreichen Studienabschlüsse bzw. der Relation abgebrochener Ausbildungswege im Reich der Spekulation. Nutzbringend für Entwicklungs- und Verbesserungsstrategien sind Quoten zu Studierendenzahlen erst, wenn die beschriebenen und nachgewiesenen Verzerrungen ausgeschlossen werden können.

## 11.2 Evaluation und Hochschulentwicklung

Vor dem Hintergrund der Analyseergebnisse zur Dimensionierung des studentischen Qualitätsurteils kann auch die Diskussion um die Relevanz verschiedener Evaluationsparadigmen erneut aufgegriffen werden. In Kapitel 3 wurde die Kontroverse zwischen dem „positivistischem“ Paradigma, welches Evaluation als Instrument zur Messung von objektiven Realitäten und Zustandsbeschreibungen verortet und der „naturalistischen“ Evaluation, die das Vorhandensein einer

einigen objektiven Realität zu Gunsten verschiedener sozial konstruierter Wirklichkeiten und Beurteilungen ablehnt, geschildert.<sup>1</sup> Die Position, die versucht, zwischen beiden zu vermitteln wird durch die vorgelegten Ergebnisse gestärkt. Wie die Analysen zeigen, wirken sich verschiedene Qualitätsdimensionen in verschiedenen Kontexten (unterschiedliche Studienphasen) in einer je spezifischen Weise auf das Qualitätsurteil aus. In dem Maße, wie von außen (bspw. seitens der die Qualitätseinschätzungen erforschenden Evaluatoren) nicht a priori festgelegt werden kann, in welcher Weise die zur Beurteilung anstehenden Qualitätskonzepte und -konstrukte verstanden und zur Grundlage von Beurteilungen gemacht werden, ist es nicht möglich, von einem Mess- oder Befragungsergebnis bspw. zur „Strukturiertheit des Studiums“ darauf zu schließen, welche organisationsbezogenen Aspekte des Studiums relevant für die erfragte Beurteilung sind. Entsprechend gering ist der Nutzen der Information einzuschätzen, die sich aus dem erhobenen Urteil zu den verschiedenen Aspekten des Studiums, wie eben bspw. der „Strukturiertheit“ ableiten lässt.

Sachgerechte Interpretationen, die ihrerseits das politische Handeln der Hochschulleitungen anleiten können, werden erst durch die Einbettung der erhobenen Einzelurteile über verschiedene Aspekte der Studienqualität in einen breiteren Kontext möglich. In einem solchen würde die Subjektivität von Qualitätseinschätzungen berücksichtigt, und Evaluation würde nicht lediglich als das Messen von Indikatoren verstanden, sondern vielmehr als Untersuchung der Beurteilungsperspektiven der Beteiligten und als eine an diese Ermittlung anschließende und die möglicherweise konfligierenden Sichtweisen einschließende Beratung.

Eine solche Herangehensweise würde dem durch das naturalistische Evaluationsparadigma formulierten Anspruch vermutlich in stärkerem Maße entgegen kommen. Gleichwohl erfordern die entsprechenden Interpretationen Messungen und Datenerhebungen und den (systematischen) Einsatz von Verfahren zur Analyse des Beziehungsgeflechts zwischen den verschiedenen (subjektiven) Urteilen

---

<sup>1</sup>Vgl. zusammenfassend Lee (2000).

zu den Qualitätsdimensionen. Derartige Informationen sind in Evaluationsdesigns unter Verzicht auf jede Form festgelegten Forschungsplans (Datenerhebungen, etc.), wie häufig von Vertretern des naturalistischen Paradigmas gefordert (Lee 2000; Guba 1990; Guba und Lincoln 1985) kaum herzustellen.

Datenerhebungen, die versuchen, die studentischen Qualitätsurteile mittels entsprechender Erhebungen und Befragungen im Aggregat zu beschreiben, könnten in diesem Sinne als der „Versuch einer Objektivierung von Subjektivität“ bezeichnet werden. Evaluation in diesem Sinne ließe sich als Ansatz zur Integration der beiden genannten Ebenen - objektive Realität und subjektive Wirklichkeitskonstruktion - auffassen. „We have multiple purposes for evaluation: to guide decision-making, to enlighten various stakeholder-groups, to improve programs, to provide information relative to design of new programs (...). We need theory to help us make some sense out of that complexity“ (Lee 2000).

Lehrevaluation, die in einem entsprechend anspruchsvollen theoriegeleiteten und auf die verschiedenen Forschungsparadigmen gestützten Design durchgeführt wird, ist durchaus in der Lage, zur Hochschulentwicklung beizutragen. Sie kann dies, indem sie relevante Informationen zur Verfügung stellt und damit politisches Handeln ermöglicht, dass auf einschlägigen Daten und damit einer hinreichend rationalen Entscheidungsgrundlage stattfindet.

## 11.3 Weiterentwicklung von Qualitätssicherungsverfahren

Zur Qualitätssicherung wird von den Hochschulen vielfach gefordert, dass sie ein professionelles Qualitätsmanagement einsetzen. Evaluation wird in diesem als ein Bestandteil gesehen, der sich hauptsächlich auf die Erhebung und den Bericht von Daten beschränkt. Nach Stockmann (2006) hat Evaluation einen vom Qualitätsmanagement unterschiedenen Eigengehalt. Dieser besteht in ihrer Forschungsorientierung einerseits und der Beratungsfunktion andererseits. Diese Unterscheidung ist insofern relevant, als das Ziel einer qualifizierten (und mit dem entsprechenden Selbstverständnis und Selbstbewusstsein auftretenden wis-



senschaftlichen Disziplin) Evaluation sein muss, diese Aspekte im Sinne der ihr entgegen gebrachten Akzeptanz weiter zu entwickeln.

Forschung wird verstanden als die systematische Anwendung qualifizierter wissenschaftlicher Methoden sowie bewährter Theorien (vgl. auch die Diskussion in Kapitel 3). Alltagsevaluationen, die sich auf die Durchführung von Umfrageforschung beschränken, scheinen nur in sehr begrenztem Maße anschlussfähig und für die Hochschulentwicklung nutzbar zu sein. Vielmehr tragen sie zu einer weiteren Verschlechterung der Akzeptanz der eingesetzten Instrumente bei, was sich (auf der Ebene der praktischen Evaluationstätigkeit) bspw. in einer zurückgehenden Teilnahmebereitschaft an studentischen Befragungen oder der Skepsis gegenüber Befragungsergebnissen in den „evaluierten“ Lehrbereichen äußert. Gegenüber ausufernden, so genannten Evaluationen, die in einem alltagssprachlichen Begriffsverständnis verschiedene, mehr oder weniger zufälliger Weise interessant erscheinende Bereiche des Hochschulalltags unter die Lupe nehmen, sind aus vorliegender Sicht Ansätze zu bevorzugen, die theoriegeleitet eine wissenschaftliche Perspektive auf den Forschungsgegenstand Lehre und Studium entwickeln (Schmidt 2005).

Ein solches Vorgehen leistet einen wichtigen Beitrag zur Professionalisierung von Evaluation und löst insbesondere im Bereich der Evaluation von Lehre und Studium einen wichtigen Anspruch ein. Dieser besteht darin, dass bei Beurteilungen im System Wissenschaft sowohl wissenschaftsadäquate Beurteilungskriterien, als auch Methoden angewendet werden, die diesem System selbst entstammen.

Die Anforderung einer theoriegeleiteten Evaluation stellt sich umso mehr, als zukünftig eine stärkere (Rück-)Besinnung auf die Ergebnisse und Wirkungen von Lehre und Studium erwartet werden kann. Die entsprechende Forschungsperspektive war mit dem Hinweis auf fehlende Möglichkeiten experimentallogische Ansätze hinreichend seriös verfolgen zu können und zu Gunsten von prozessorientierten Fragestellungen in ihrer Bedeutung zurückgestellt worden (vgl. die Ausführungen in Kapitel 3). In dem Maße, wie der Legitimationsdruck hinsichtlich der eingesetzten Ressourcen auf die Hochschulen wächst, ist zu erwarten, dass die Nachfrage nach Evaluationen der Wirkung von Hochschulleistungen nicht oh-

ne weiteres mit dem Hinweis auf methodische Probleme abgespeist werden kann.

Eine zentrale Anforderung an die Evaluation von Lehre und Studium besteht mithin darin, Instrumentarien zu entwickeln, die die entsprechende Auswertungsperspektive zu erschließen erlauben. Dies ist durch eine theoriegeleitete Herangehensweise möglich. Diese erlaubt es, formulierte Hypothesen zur Studienqualität bzw. der Wirksamkeit von Aspekten des Hochschulhandelns bspw. auf Lernerfolge der Studierenden zu testen. Nötig sind dazu differenzierte Analyseverfahren einerseits, die eine kausalanalytische Auswertung erhobener Daten erlauben. Diese wurden in Kapitel 9 u.a. hinsichtlich der Stabilitätsanalyse zu studentischen Qualitätseinschätzungen vorgestellt. Sie ermöglichen, kausale Effekte in fortgeschrittener Weise zu ermitteln. Andererseits ist eine qualifizierte Datenbasis erforderlich. Hier kommt die ebenfalls in der vorliegenden Arbeit dargestellte Längsschnittperspektive der Datenerhebung zum Tragen.

Durch eine kausalanalytische Herangehensweise an die Frage der Wirkung bspw. der Organisationsqualität der Hochschule auf zeitlich nachgelagerte Lernerfolge von Studierenden können durchaus wirkungsorientierte Analysen durchgeführt werden, ohne dass die geforderten experimentallogischen Forschungsdesigns in aller methodischen Strenge realisiert werden. Die entsprechenden Instrumentarien wurden in ihren Grundlegungen in der vorliegenden Arbeit vorgestellt und können in der genannten Perspektive weiterentwickelt werden. Insbesondere stellt sich im Zusammenhang mit der Weiterentwicklung der Instrumentarien aber auch die forschungspragmatische Frage nach dem nötigen Aufwand, der für den erwarteten zusätzlichen Erkenntnisgewinn zu treiben ist. Dieser ist im Falle einer längsschnittlichen Datenerhebung in der Tat hoch. Gleichzeitig kann dabei auf Instrumente zurückgegriffen werden, die sich bspw. der neuen Medien bedienen und Längsschnittdaten mittels online gestützter Verfahren erheben.<sup>2</sup>

---

<sup>2</sup>Erste Schritte in diese Richtung konnten bereits gegangen werden. So wurde im Rahmen eines mit Mitteln des BMBF geförderten Projekts ein Instrument zur internetbasierten Erhebung von Längsschnittdaten entwickelt. Dieses wird für zukünftige entsprechende Befragungen Studierender zu ihren Qualitätseinschätzungen sowie zur Erhebung einschlägiger objektiver Daten

Wirkungsorientierte Untersuchungen sind nicht nur geeignet, das von außen an die Hochschulen herangetragene Informationsbedürfnis zur Effizienz der Hochschulleistungen zu befriedigen. Vielmehr stellen sie auch für die interne Strategieentwicklung eine relevante, weil in höchstem Maße aussagekräftige Handlungs- und Entscheidungsgrundlage dar.

## 11.4 Evaluation oder Evaluitis?

Die möglichen Folgen einer Evaluationskultur, die die beschriebenen Anforderungen nicht erfüllt, sondern vielmehr als ein in seiner Durchführungsfrequenz ausuferndes Instrument der Kontrolle von außen wahrgenommen wird, sind vielfach diskutiert worden (vgl. z.B. Frey 2006). Diese bestehen in opportunistischen Anpassungseffekten, die sich in der Lehre bspw. dadurch auswirken können, dass die Notenvergabe an eine erwartete bessere Evaluation durch Studierende geknüpft und „nach oben korrigiert“ wird (Greenwald und Gillmore 1997).

Im Bereich der Forschung wird darüber hinaus befürchtet, dass Mittelmäßigkeit gefördert wird, indem auf die Umsetzung innovativer Ideen zugunsten von solchen Forschungsvorhaben verzichtet wird, von denen a priori angenommen werden kann, dass sie zu einem auf die einschlägigen Leistungskriterien bezogenen Erfolg führen. Von der Investition in Risikokapital wird dagegen Abstand genommen. Dieser Effekt lässt sich in einem gewissen Maß auch für die Lehre erwarten. Er tritt ein, wenn nur noch das Abhandeln von vorgeschriebenen Inhalten, die in möglichst kurzer Zeit beruflich verwertbares Anwendungswissen vermitteln sollen als Ausbildungserfolg in den entsprechenden Evaluationsverfahren anerkannt wird und alles, was darüber hinaus geht, als Wettbewerbsnachteil gesehen wird, weil es die (Zeit-)Effizienz des Studiums beeinträchtigt. Als besonders gravierende Folge eines solchen Verständnisses von Qualitätssicherung in der Lehre ist wahrscheinlich der mögliche Prestigeverlust des Berufsbildes Hochschullehrer zu sehen. Dieser stellt sich ein, durch eine an den erwarteten Evaluationsergebnissen orien-

---

zum Lernerfolg eingesetzt (vgl. <http://www.up-panel.de>).

tierte Standardisierung der Ausbildungstätigkeit. Die Freiheit der Gestaltung der Lehre ist wesentliches Merkmal des Berufsbildes, welches die hohe Autonomie und Selbstständigkeit von Hochschullehrern ausdrückt. Entsprechende Einschränkungen, die durch die (wahrgenommenen und tatsächlichen) Forderungen des Evaluationsregimes ausgelöst werden, resultieren in dem genannten Prestigeverlust. Damit einher gehend ist ein Verlust an Attraktivität des Hochschullehrerberufs, so dass möglicherweise Entscheidungen gegen eine entsprechende Karriere getroffen werden. Hochschullehrer, die bereits negative Erfahrungen mit einer falsch verstandenen Evaluation gemacht haben, könnten zudem den inneren Rückzug antreten und die Motivation verlieren, innovative Lehrvorhaben jenseits der eingetretenen Pfade zu verwirklichen. Gerade diese sind es aber, die ein qualitativ hochwertiges Studienangebot definieren.

Evaluation steht vor der Aufgabe, Routinen zu entwickeln, die der Vielgestaltigkeit der Qualität der Lehre sowie der Realität des Hochschulalltages gerecht werden. Nur so kann sie den Anspruch darauf erheben, als Instrument der Hochschulentwicklung ernst genommen zu werden. Insofern dies den Einsatz komplexer und verschiedene Ansätze versöhnende Verfahren und damit einen stärkeren Mitteleinsatz erfordert, ist das Hochschulwesen insgesamt gefragt, inwieweit die vielfach verkündete Formel von der Entwicklung von Exzellenz durch qualitätssichernde Verfahren ernst gemeint und mit der Bereitschaft verknüpft ist, die entsprechend notwendigen Investitionen in die Zukunft zu tätigen.

# Literaturverzeichnis

- [Abrami 1980] ABRAMI, Philip C.: Do Teacher Standards for Assigning Grades Affect Student Evaluations of Instruction? In: *Journal of Educational Psychology* 72 (1980), Nr. 1, S. 107–118
- [Abrami 2001] ABRAMI, Philip C.: Improving Judgments About Teaching Effectiveness: How to Lie Without Statistics. In: THEALL, Michael (Hrsg.) ; ABRAMI, Philip C. (Hrsg.) ; METS, Lisa (Hrsg.): *The Student Ratings Debate*. San Francisco, 2001, S. 97–102
- [Abrami u. a. 1990] ABRAMI, Philip C. ; D'APOLLONIA, Sylvia ; COHEN, Peter A.: Validity of Student Ratings of Instruction: What We Know and What We Do Not. In: *Journal of Educational Psychology* 82 (1990), Nr. 2, S. 219–231
- [Abrami u. a. 1982] ABRAMI, Philip C. ; PERRY, Raymond P. ; LEVENTHAL, Les: The Relationship Between Student Personality Characteristics, Teacher Ratings, and Student Achievement. In: *Journal of Educational Psychology* 74 (1982), Nr. 1, S. 111–125
- [Anderson und Bennett 2002] ANDERSON, Lesley (Hrsg.) ; BENNETT, Nigel (Hrsg.): *Developing Educational Leadership: Using Evidence for Policy and Practice*. London u.a. : Sage, 2002
- [Argyris 1992] ARGYRIS, Chris: *On Organizational Learning*. Cambridge : Blackwell, 1992
- [Argyris und Schön 1999] ARGYRIS, Chris ; SCHÖN, Donald A.: *Die lernende Organisation. Grundlagen, Methoden, Praxis*. Stuttgart : Klett-Cotta, 1999

- [Arnold 1997]    ARNOLD, Harald: Welche Informationen zum Studienabbruch kann die Bildungsstatistik des Landes Baden-Württemberg liefern? In: *Beiträge zur Hochschulforschung* 20 (1997), Nr. 4, S. 337–348
- [Arnold und Faber 2000]    ARNOLD, Rolf ; FABER, Konrad: *Qualitätsmanagement - aber wie? Qualitätssysteme und ihre Relevanz für die Schule: Einführung und Überblick*. Leipzig : Kallmeyer, 2000
- [Arnold 2006]    ARNOLD, Stefan: Gestaltung der Reakkreditierung, Clusterakkreditierung am Beispiel der ZEvA. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Qualitätsentwicklung an Hochschulen. Erfahrungen aus 10 Jahren Evaluation*. Bd. 8/2006. Bonn, 2006, S. 176–184
- [Backhaus u. a. 2000]    BACKHAUS, Klaus ; ERICHSON, Bernd ; PLINKE, Wulff ; WEIBER, Rolf: *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 9. Aufl. Berlin u.a. : Springer, 2000
- [Bayer 2004]    BAYER, Christian R.: *Abhandlungen zu Bildungsforschung und Bildungsrecht*. Bd. 14: *Hochschul-Ranking: Vorschlag eines ganzheitlichen Ranking-Verfahrens*. Berlin : Duncker u. Humblot, 2004
- [Berghoff u. a. 2003]    BERGHOFF, Sonja ; FEDERKEIL, Gero ; GIEBISCH, Petra ; HACHMEISTER, Cort-Dennis ; MÜLLER-BÖLING, Detlef: Das CHE-Hochschul Ranking. Vorgehensweisen und Indikatoren / Centrum für Hochschulentwicklung. April 2003; Arbeitspapier Nr. 46 2003. – Forschungsbericht
- [Berghoff u. a. 2004]    BERGHOFF, Sonja ; FEDERKEIL, Gero ; GIEBISCH, Petra ; HACHMEISTER, Cort-Dennis ; MÜLLER-BÖLING, Detlef: Das CHE-Hochschul Ranking / Centrum für Hochschulentwicklung. April 2004; Arbeitspapier Nr. 54 2004. – Forschungsbericht
- [Berghoff u. a. 2006]    BERGHOFF, Sonja ; FEDERKEIL, Gero ; GIEBISCH, Petra ; HACHMEISTER, Cort-Dennis ; MÜLLER-BÖLING, Detlef ; RÖLLE, Daniel: CHE -Hochschul Ranking. Vorgehensweise und Indikatoren / Centrum für

- Hochschulentwicklung. 03.05.2006; Arbeitspapier Nr. 75 2006. – Forschungsbericht
- [Beywl 1988] BEYWL, Wolfgang: *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt/Main u.a. : Peter Lang, 1988 (Europäische Hochschulschriften)
- [Bühner 2004] BÜHNER, Markus: *Einführung in die Test- und Fragebogenkonstruktion*. München : Pearson, 2004
- [BMBF ] BMBF: Hochschulreform. In: *Pressemitteilung des BMBF* 12.12.2005
- [BMBF 2003] BMBF: Bulmahn will Studienbedingungen verbessern. Hohe Abbrecherquote muss dringend sinken. In: *Pressemitteilung des BMBF* (2003), 06.03.2003 [www.bmbf.de/presse01/811.html](http://www.bmbf.de/presse01/811.html)
- [Bortz und Döring 2002] BORTZ, Jürgen ; DÖRING, Nicola: *Forschungsmethoden und Evaluation*. Berlin u.a. : Springer, 2002
- [Brennan und Shah 2000] BRENNAN, John ; SHAH, Tarla: *Managing Quality in Higher Education - an International Perspective on Institutional Assessment and Change*. Buckingham, 2000
- [Börger 2006] BÖRGER, Christian: Zielvereinbarungen zwischen Hochschule und Staat. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*. Bonn, 2006, S. 115–125
- [Brommes u. a. 2004] BROMMES, Michael ; GUTER, Jochen ; WOLFF-JONTOFSOHN, Ulrike: Miteinander - Erfahrungen mit Betzavta: Evaluationsprozess und Ergebnisse. In: UHL, Katrin (Hrsg.) ; ULRICH, Susanne (Hrsg.) ; WENZEL, Florian M. (Hrsg.): *Evaluation politischer Bildung. Ist Wirkung messbar?* Gütersloh : Bertelsmann Stiftung, 2004, S. 91–120

- [Bundestag 2005] BUNDESTAG: 13. Bericht des Ausschusses für die Hochschulstatistik für den Zeitraum 1. Juni 2000 bis 31. Mai 2004 / Deutscher Bundestag. URL <http://dip.bundestag.de/btd/15/054/1505400.pdf>; Stand: 10.02.2006, 2005. – Forschungsbericht
- [Bushe 1995] BUSHE, Gervase R.: Advances in Appreciative Inquiry as an Organizational Development Intervention. In: *Organizational Development Journal* 13 (1995), Nr. 3, S. 14–22
- [Bushe und Kassam 2005] BUSHE, Gervase R. ; KASSAM, Aniq F.: When is Appreciative Inquiry Transformational? A Meta-Case Analysis. In: *Journal of Applied Behavioral Science* 41 (2005), Nr. 2, S. 161–181
- [Byrne 2001] BYRNE, Barbara M.: *Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programming*. New Jersey/London : Lawrence Earlbaum Ass. Publishers, 2001 (Multivariate Applications Book Series)
- [Cadwell und Jenkins 1985] CADWELL, Joel ; JENKINS, Jeffrey: Effects of the Semantic Similarity of Items on Student Ratings of Instruction. In: *Journal of Educational Psychology* 77 (1985), Nr. 4, S. 383–393
- [Campbell und Fiske 1959] CAMPBELL, D.T. ; FISKE, D.W.: Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. In: *Psychological Bulletin* 56 (1959), S. 81–105
- [Cannon 2001] CANNON, Robert: Broadening the Context for Teaching Evaluation. In: KNAPPER, Christopher (Hrsg.) ; CRANTON, Patricia (Hrsg.): *Fresh Approaches to the Evaluation of Teaching* Bd. 88. San Francisco : Jossey-Bass, 2001, S. 87–97
- [Carstensen 2004] CARSTENSEN, Doris: Lernen in Veränderungsprozessen. Organisationales Lernen und defensive Routinen an Universitäten. In: *Die Hochschule. Journal für Wissenschaft und Bildung*. 13 (2004), Nr. 1, S. 49–62



- [Carstensen und Reissert 1995] CARSTENSEN, Doris (Hrsg.) ; REISSERT, Reiner (Hrsg.): *HIS Kurzinformation*. Bd. A16/95: *Interne und externe Evaluation - Modell und Praxis. Eine Zwischenbilanz aus der Sicht von HIS*. Hannover, 1995
- [Chelimsky 1985] CHELIMSKY, Eleanor: Old Patterns and New Directions in Program Evaluation. In: DIES. (Hrsg.): *Program Evaluation: Patterns and Directions* Bd. 6. Washington D.C., 1985, S. 1–35
- [Christie und Alkin 2003] CHRISTIE, Christina A. ; ALKIN, Marvin C.: The User-Oriented Evaluator's Role in Formulating a Program Theory: Using a Theory-Driven Approach. In: *American Journal of Evaluation* 24 (2003), Nr. 3, S. 373–385
- [Cooperrider und Srivastva 1987] COOPERRIDER, D. L. ; SRIVASTVA, S.: Appreciative Inquiry in Organizational Life. In: WOODMAN, R. (Hrsg.) ; PASMORE, W. (Hrsg.): *Research in Organizational Change and Development* Bd. 1. Greenwich : JAI Press, 1987, S. 129–169
- [Cranton und Smith 1990] CRANTON, Patricia ; SMITH, Ronald A.: Reconsidering the Unit of Analysis. In: *Journal of Educational Psychology* 82 (1990), Nr. 2, S. 207–212
- [Daniel 1994] DANIEL, Hans-Dieter: Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung. In: *Empirische Pädagogik* 8 (1994), Nr. 2, S. 109–129
- [Daniel 2006] DANIEL, Hans-Dieter: Pro Peer Review: 5 Thesen. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*. Bonn, 2006 (Beiträge zur Hochschulpolitik, Bd. 1/2006), S. 185–192
- [d'Apollonia und Abrami 1997] D'APOLLONIA, Sylvia ; ABRAMI, Philip C.: Navigating Student Ratings of Instruction. In: *American Psychologist* 52 (1997), Nr. 11, S. 1198–1208
- [DeGEval 2002] DEGEVAL: *Standards für Evaluation*. Köln, 2002

- [Desforges 2002] DESFORGES, Charles: Evidence-Informed Policy and Practice in Teaching and Learning. In: ANDERSON, Lesley (Hrsg.) ; BENNETT, Nigel (Hrsg.): *Developing Educational Leadership: using Evidence for Policy and Practice*. London : Sage, 2002, S. 3–10
- [DuCette und Kenney 1982] DUCETTE, Joseph ; KENNEY, Jane: Do Grading Standards Affect Student Evaluations of Teaching? Some New Evidence on an Old Question. In: *Journal of Educational Psychology* 74 (1982), Nr. 3, S. 308–314
- [Ehlert und Drees 2004] EHLERT, Holger ; DREES, Stefan: Paradigmenwechsel in der Bildungspolitik: Evaluation und Akkreditierung als Instrumente der Qualitätssicherung. In: EHLERT, Holger (Hrsg.) ; WELBERS, Ulrich (Hrsg.): *Qualitätssicherung und Studienreform. Strategie- und Programmentwicklung für Fachbereiche und Hochschulen im Rahmen von Zielvereinbarungen am Beispiel der Heinrich-Heine- Universität Düsseldorf*. Düsseldorf : Grupello, 2004, S. 80–103
- [Enders und Schimank 2001] ENDERS, Jürgen ; SCHIMANK, Uwe: Faule Professoren und vergreiste Nachwuchswissenschaftler. In: STÖLTING, Erhard (Hrsg.) ; SCHIMANK, Uwe (Hrsg.): *Die Krise der Universitäten*. Wiesbaden : 159-178, 2001
- [Engel 2000] ENGEL, Uwe: Determinanten studentischer Qualitätsurteile. Über Variationsursachen, Verzerrungen und Verantwortlichkeiten. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Leitbild der Hochschule - Qualität der Lehre*. Bonn, 2000 (Beiträge zur Hochschulpolitik, Bd. 2/2000), S. 127–140
- [Engel 2001] ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. zur Qualitätsbewertung von Studium und Lehre*. Frankfurt/Main : Campus, 2001
- [Engel und Krekeler 2001] ENGEL, Uwe ; KREKELER, Gaby: Studienqualität. Über studentische Bewertungen und Rankings von Studienfächern einer Universität. In: ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. Frankfurt/Main : Campus, 2001, S. 121–176

- [Engel und Pohlenz 2001] ENGEL, Uwe ; POHLENZ, Philipp: Lehre und Studium im Spiegel studentischer Bewertungen: Über das Potsdamer Modell der Lehrevaluation. In: SPIEL, Christiane (Hrsg.): *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster u.a. : Waxmann, 2001, S. 131–150
- [Engel und Pötschke 2003] ENGEL, Uwe ; PÖTSCHKE, Manuela: *Auto und sonst nichts? Zum Verhältnis von Umweltschutz und Verkehrsmittelwahl*. Münster : LiT Verlag, 2003
- [Engel und Reinicke 1994] ENGEL, Uwe ; REINICKE, Jost: *Panelanalyse. Grundlagen, Techniken, Beispiele*. Berlin u.a. : De Gruyter, 1994
- [Engels 2004] ENGELS, Maria: Eine Annäherung an die Universität aus organisationstheoretischer Sicht. In: *Die Hochschule. Journal für Wissenschaft und Bildung*. 13 (2004), Nr. 1, S. 12–29
- [ENQA 2000] ENQA: Evaluation of European Higher Education: A Status Report / European Network for Quality Assurance. 2000. – Forschungsbericht
- [Esser u. a. 1977] ESSER, H. ; KLENOVITS, K. ; ZEHNPFENNIG, H.: *Wissenschaftstheorie*. Bd. Bd. 1. Stuttgart, 1977
- [Esser 1997] ESSER, Hartmut: Zweifel an der Evaluation der Lehre. In: *WiSt* (1997), Nr. 1, S. 45–49
- [Falter 1977] FALTER, Jürgen W.: Zur Validierung theoretischer Konstrukte - Wissenschaftstheoretische Aspekte des Validierungskonzepts. In: *Soziologie* 6 (1977), Nr. 2, S. 349–369
- [Federkeil 2005] FEDERKEIL, Gero: Leistungsparameter für Lehre und Studium. In: *Qualität messen - Qualität managen. Leistungsparameter in der Hochschulentwicklung*. Bonn, 2005 (Beiträge zur Hochschulpolitik, Bd. 6/2005), S. 51–54
- [Feinstein 2002] FEINSTEIN, Osvaldo N.: Use of Evaluations and the Evaluation of their Use. In: *Evaluation* 8 (2002), Nr. 4, S. 433–439

- [Feldman 1977] FELDMAN, Kenneth A.: Consistency and Variability among College Students in Rating Their Teachers and Courses: A review and Analysis. In: *Research in Higher Education* 6 (1977), S. 223–274
- [Feldman 1978] FELDMAN, Kenneth A.: Course Characteristics and College Students' Ratings of their Teachers: What We Know and What We Don't. In: *Research in Higher Education* 9 (1978), S. 199–242
- [Fischer-Bluhm 1995] FISCHER-BLUHM, Karin: Gemeinsam geht es besser! Evaluationsprojekte im Verbund norddeutscher Universitäten. In: *Handbuch Hochschullehre*. Bonn : Raabe Fachverlag für Wissenschaftsinformation, 1995
- [Fischer-Bluhm 1998] FISCHER-BLUHM, Karin: Evaluation im Verbund Norddeutscher Universitäten. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation und Qualitätssicherung an den Hochschulen in Deutschland - Stand und Perspektiven* Bd. 6/1998. Bonn, 1998, S. 25–33
- [Fischer-Bluhm 2000] FISCHER-BLUHM, Karin: Qualitätsentwicklung als Antwort auf Individualisierung im Bildungsbereich - am Beispiel der Hochschulen. / Kongress Forum Bildung. 14.7.2000 2000. – Forschungsbericht
- [Fischer-Bluhm u. a. 2003] FISCHER-BLUHM, Karin ; TSCHIRKOV, Carmen ; ZEMENE, Susanne: Verbund Norddeutscher Universitäten. In: *Zeitschrift für Evaluation* (2003), Nr. 2, S. 311–321
- [Frey 2006] FREY, Bruno: Evaluitis - Eine neue Krankheit / Institute for Empirical Research in Economics University of Zurich. 2006 (Working Paper No. 293). – Forschungsbericht
- [Fröhlich 2006] FRÖHLICH, Gerhard: Informed Peer Review - Ausgleich der Fehler und Verzerrungen? In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*. Bonn, 2006 (Beiträge zur Hochschulpolitik, Bd. 1/2006), S. 193–204

- [Futterer 2005] FUTTERER, Torsten: Befragungen in der internen Evaluation. Erhebung eines Meinungsbildes in der Evaluation von Lehre und Studium an Hochschulen / ZEvA Zentrale Evaluations- und Akkreditierungsagentur. 2005. – Forschungsbericht
- [Ganz 1984] GANZ, Carole: Plädoyer für eine Evaluierung des Evaluierungsforschungsprozesses. In: HELLSTERN, Gerd-Michael (Hrsg.) ; WOLLMANN, Hellmut (Hrsg.): *Handbuch zur Evaluationsforschung*. Opladen : Westdeutscher Verlag, 1984
- [Garacelli 2000] GARACELLI, Valerie J.: Methodology: Building Bridges to Knowledge. In: STOCKMANN, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 165–190
- [Ginsburg und Rhett 2003] GINSBURG, Alan ; RHETT, Nancy: Building a Better Body of Evidence: New Opportunities to Strengthen Evaluation Utilization. In: *American Journal of Evaluation* 24 (2003), Nr. 4, S. 489–498
- [Green und McAllister 1998] GREEN, Beth L. ; MCALLISTER, Carol: Theory-Based, Participatory Evaluation: A Powerful Tool for Evaluating Family Support Programs. In: *Bulletin of the National Center for Zero to Three* 2/1998 (1998), S. 30–36
- [Greene 1989] GREENE, Jennifer C.: Naturalistic Interviewing. In: MERTENS, Donna M. (Hrsg.): *Creative Ideas for Teaching Evaluation. Activities, Assignments and Resources*. Boston, 1989, S. 123–129
- [Greene 1989a] GREENE, Jennifer C.: Naturalistic Data Collection: Case Study Discussion. In: MERTENS, Donna M. (Hrsg.): *Creative Ideas for Teaching Evaluation. Activities, Assignments and Resources*. Boston : Kluwer, 1989a, S. 135–140
- [Greenwald 1997] GREENWALD, Anthony C.: Validity Concerns and Usefulness

- of Student Ratings of Instruction. In: *American Psychologist* 52 (1997), Nr. 11, S. 1182–1186
- [Greenwald und Gillmore 1997] GREENWALD, Anthony G. ; GILLMORE, Gerald M.: Grading Leniency is a Removable Contaminant of Student Ratings. In: *American Psychologist* 52 (1997), Nr. 11, S. 1209–1217
- [Greimel und Geyer o.J.] GREIMEL, Bettina ; GEYER, Alois: Welche Faktoren beeinflussen die Evaluation von Lehrkräften? / Wirtschaftsuniversität Wien. o.J. – Forschungsbericht
- [Griesbach u. a. 2000] GRIESBACH, Heinz ; LEWIN, Karl ; HEUBLEIN, Ulrich ; SOMMER, Dieter: *HIS-Kurzinformationen*. Bd. 5/2000: *Studienabbruch. Typologie und Möglichkeiten der Abbruchquotenbestimmung*. Hannover, 2000
- [Grob 2003] GROB, George F.: A Truly Useful Bat is One Found in the Hands of a Slugger. In: *American Journal of Evaluation* 24 (2003), Nr. 4, S. 499–505
- [Guba 1990] GUBA, Egon G.: The Alternative Paradigm Dialog. In: GUBA, Egon G. (Hrsg.): *The Paradigm Dialog*. Newbury Park : Sage, 1990, S. 17–30
- [Guba und Lincoln 1981] GUBA, Egon G. ; LINCOLN, Yvonna S.: *Effective Evaluation. Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*. San Francisco : Jossey-Bass, 1981
- [Guba und Lincoln 1985] GUBA, Egon G. ; LINCOLN, Yvonna S.: *Naturalistic Inquiry*. Beverly Hills : Sage, 1985
- [Gutjahr-Löser 2000] GUTJAHR-LÖSER, Peter: Kontrolle ist gut - Vertrauen ist besser. Warum Universitäten nicht wie Unternehmen organisiert sind. In: *Forschung und Lehre* 3/2000 (2000), S. 122–125
- [Hage 1996] HAGE, Natalija e.: *Lehrevaluation und studentische Veranstaltungskritik. Projekte, Instrumente und Grundlagen*. Bonn, 1996

- [Hager 2000] HAGER, Willi: Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In: HAGER, Willi (Hrsg.) ; PATRY, Jean-Luc (Hrsg.) ; BREZINGER, Hermann (Hrsg.): *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien*. Bern u.a. : Huber, 2000, S. 180–201
- [Hahn 1998] HAHN, H.-J.: *Education and Society in Germany*. Oxford : Berg, 1998
- [Hahne 2001] HAHNE, Anton: EFQM. In: HANFT, Anke (Hrsg.): *Grundbegriffe des Hochschulmanagements*. Neuwied/Kriftel : Luchterhand, 2001, S. 97–104
- [Hansen 2005] HANSEN, Hanne F.: Choosing Evaluation Models: A Discussion on Evaluation Design. In: *Evaluation* 11 (2005), Nr. 4, S. 447–462
- [Hauck 1993] HAUCK, Gerhard: *Geschichte der soziologischen Theorie. Eine ideologiekritische Einführung*. Reinbek : Rowohlt, 1993
- [Hellstern und Wollmann 1984] HELLSTERN, Gerd-Michael ; WOLLMANN, Hellmut: Evaluierung und Evaluierungsforschung - ein Entwicklungsbericht. In: DIES. (Hrsg.): *Handbuch der Evaluierungsforschung*. Opladen : Westdeutscher Verlag, 1984, S. 17–91
- [Hener 2004] HENER, York: Qualitätsmanagement in Hochschulen. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation - ein Bestandteil des Qualitätsmanagements an Hochschulen*. Bonn, 2004 (Beiträge zur Hochschulpolitik, Bd. 9/2004), S. 111–128
- [Hennen 2002] HENNEN, Manfred (Hrsg.): *Mainzer Beiträge zur Hochschulentwicklung*. Bd. 4: *Evaluation - Erfahrungen und Perspektiven*. Mainz, 2002
- [Heublein und Sommer o. J.] HEUBLEIN, Ulrich ; SOMMER, Dieter ; DAAD, Deutscher Akademischer A. (Hrsg.): *Analyse des Studienverlaufs von ausländischen Studierenden. Manual*. o. J.

- [Heublein u. a. 2003] HEUBLEIN, Ulrich ; SPANGENBERG, Heike ; SOMMER, Dieter: *Hochschulplanung*. Bd. 163: *Ursachen des Studienabbruchs. Analyse 2002*. Hannover, 2003
- [Hirseland u. a. 2004] HIRSELAND, Andreas ; DIMBATH, Oliver ; HAYEK, Julia v. ; SCHNEIDER, Werner: Evaluation des Programms Ächtung (+) Toleranz ein Praxisbericht. In: UHL, Katrin (Hrsg.) ; ULRICH, Susanne (Hrsg.) ; WENZEL, Florian M. (Hrsg.): *Evaluation politischer Bildung. Ist Wirkung messbar?* Gütersloh : Bertelsmann Stiftung, 2004, S. 57–82
- [Hochschulrektorenkonferenz 2000a] HOCHSCHULREKTORENKONFERENZ: *Dokumentze und Informationen*. Bd. 2/2000: *Wegweiser 2000 durch die Qualitätssicherung von Lehre und Studium*. Bonn, 2000a
- [Hochschulrektorenkonferenz 2000b] HOCHSCHULREKTORENKONFERENZ: *Evaluation der Lehre - Sachstandsbericht mit Handreichungen*. Bonn, 2000b
- [Hochschulrektorenkonferenz 2002] HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Ein Jahr vor der Konferenz der europäischen Bildungsminister in Berlin: Qualitätssicherung als Kernaufgabe der Hochschulen im Bologna-Prozess*. 2002 (Pressemitteilung, 7.11.2002)
- [Hochschulrektorenkonferenz 2004] HOCHSCHULREKTORENKONFERENZ: *Beiträge zur Hochschulpolitik*. Bd. 8/2004: *Bologna-Reader. Texte und Hilfestellungen zur Umsetzung der Ziele des Bologna-Prozesses an deutschen Hochschulen*. Bonn, 2004
- [Hocke 2003] HOCKE, Norbert: Aufwachsen in öffentlicher Verantwortung. In: KÖHLER, Gerd (Hrsg.) ; QUAISER, Gunter (Hrsg.): *Bildung - Ware oder öffentliches Gut? Über die Finanzierung von Bildung und Wissenschaft (Dokumentation der 22. GEW-Sommerschule)*. Frankfurt/Main, 2003, S. 39–45
- [Hofmann und Stiksrud 1994] HOFMANN, Hubert ; STIKSRUD, Arne: Zufriedenheit mit einem Psychologie-Lehrbetrieb. Aspekte der Evaluation von Evaluatoren. In: *Empirische Pädagogik* 8 (1994), Nr. 2, S. 169–198



- [Hofmann 1988] HOFMANN, Jürgen M.: Studienmotivation und Veranstaltungsbeurteilung. Eine Korrelationsstudie am Beispiel der Beurteilung psychologischer Lehrveranstaltungen. In: *Psychologie, Erziehung, Unterricht* 35 (1988), S. 119–126
- [Hofmann 2006] HOFMANN, Stefanie: Für Prozessqualität in Lehre und Studium: die Prozessakkreditierung. URL <http://evanet.his.de/positionen/positionen2006/hofmann.pdf>, 2006. – Forschungsbericht
- [Hofmann 2006a] HOFMANN, Stefanie: Grundzüge der Prozessakkreditierung. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Qualitätsentwicklung an Hochschulen. Erfahrungen aus 10 Jahren Evaluation*. Bonn, 2006a (Beiträge zur Hochschulpolitik, Bd. 8/2006), S. 164–170
- [Hopko 2003] HOPKO, Derek R.: Confirmatory Factor Analysis of the Math-Anxiety Rating Scale-Revised. In: *Educational and Psychological Measurement* 63 (2003), Nr. 2, S. 336–351
- [Hornbostel 2001] HORNBOSTEL, Stefan: Der Studienführer des CHE - ein multidimensionales Ranking. In: ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. Frankfurt/Main : Campus, 2001, S. 83–120
- [Hornbostel 2001a] HORNBOSTEL, Stefan: Hochschulranking: Beliebigkeit oder konsistente Beurteilungen? Rankings, Expertengruppen und Indikatoren im Vergleich. In: MÜLLER-BÖLING, Detlef (Hrsg.) ; HORNBOSTEL, Stefan (Hrsg.) ; BERGHOFF, Sonja (Hrsg.): *Hochschulranking. Aussagefähigkeit, Methoden, Probleme*. Gütersloh : Bertelsmann Stiftung, 2001a, S. 7–42
- [Hornbostel und Daniel 1994] HORNBOSTEL, Stefan ; DANIEL, Hans-Dieter: Das SPIEGEL-Ranking. Mediensensation oder ein Beitrag zur hochschulvergleichenden Lehrevaluation? In: MOHLER, Peter P. (Hrsg.): *Universität und Lehre: Ihre Evaluation als Herausforderung an die empirische Sozialforschung*. Münster u.a. : Waxmann, 1994, S. 29–44

- [Howard u. a. 1985] HOWARD, George S. ; CONWAY, Christine G. ; MAXWELL, Scott E.: Construct Validity of Measures of College Teaching Effectiveness. In: *Journal of Educational Psychology* 77 (1985), Nr. 2, S. 187–196
- [Hradil 2001] HRADIL, Stefan: *Soziale Ungleichheit in Deutschland*. 8. Auflage. Opladen : Leske und Budrich, 2001
- [Hu und Bentler 1995] HU, Li-Tze ; BENTLER, Peter M.: Evaluating Model Fit. In: HOYLE, Rick H. (Hrsg.): *Structural Equation Modeling: Concepts, Issues and Applications*. Thousand Oaks, London, Delhi, 1995, S. 76–99
- [Humboldt 1809] HUMBOLDT, Wilhelm v.: *Der königsberger und litauische Schulplan*. Berlin, 1809 (W. v. Humboldts gesammelte Schriften, Bd. XIII)
- [Jackson u. a. 1999] JACKSON, Dennis L. ; TEAL, Cayla R. ; RAINES, Susan J. ; NANSEL, Tonja R. ; FORCE, Ronald C. ; BURDSAL, Charles A.: The Dimensions of Student's Perceptions of Teaching Effectiveness. In: *Educational and Psychological Measurement* 59 (1999), Nr. 4, S. 580–596
- [Jahn 2001] JAHN, Heidrun: Neue Studiengänge und Abschlüsse: Qualitätsbezogene Zieldefinitionen und Zielkonflikte. In: OLBERTZ, Jan-Hendrik (Hrsg.) ; PASTERNAK, Peer (Hrsg.) ; KRECKEL, Reinhard (Hrsg.): *Qualität - Schlüsselfrage der Hochschulreform*. Weinheim : Beltz, 2001, S. 197–226
- [Jahn 2002] JAHN, Heidrun: Neue Studeingänge und Akkreditierung. In: WINTER, Thomas Reil u. M. (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 54–63
- [Jaspers und Rossmann 1961] JASPERS, Karl ; ROSSMANN, Kurt: *Die Idee der Universität*. Berlin u.a. : Springer, 1961
- [Jäger 2006] JÄGER, Michael: Leistungsbezogene Budgetierung an deutschen Universitäten. In: *Wissenschaftsmanagement* (2006), Nr. 3/2006, S. 30–36
- [Joußen 1995] JOUSSEN, W.: Kampagnen als Instrument zur Steuerung des Umweltverhaltens. In: JOUSSEN, W. (Hrsg.) ; HESSLER, A. G. (Hrsg.): *Umwelt und*

- Gesellschaft. Eine Einführung in die sozialwissenschaftliche Umweltforschung.*  
Berlin : Akademie-Verlag, 1995, S. 130–139
- [Jöreskog 1973] JÖRESKOG, Karl G.: A General Method for Estimating a Linear Structural Equation System. In: GOLDBERGER, A. (Hrsg.) ; DUNCAN, O.D. (Hrsg.): *Structural Equation Models in Social Sciences.* New York, London : Seminar Press, 1973, S. 85–112
- [Kline 1998] KLINE, Rex B.: *Principles and Practice of Structural Equation Modeling.* New York, London : Guilford Press, 1998 (Methodology in the Social Sciences)
- [Knaap 2004] KNAAP, Peter van d.: Theory-based Evaluation and Learning: Possibilities and Challenges. In: *Evaluation* 10 (2004), Nr. 1, S. 16–34
- [Knapper 2001] KNAPPER, Christopher: Broadening Our Approach to Teaching Evaluation. In: KNAPPER, Christopher (Hrsg.) ; CRANTON, Patricia (Hrsg.): *Fresh Approaches to the Evaluation of Teaching* Bd. 88. San Francisco : Jossey-Bass, 2001, S. 3–9
- [König 2002] KÖNIG, Joachim: Wichtig und richtig! Aber: Ist nicht weniger mehr? Anmerkungen zu den Standards für Selbstevaluation. In: *Zeitschrift für Evaluation* 1 (2002), Nr. 2, S. 83–88
- [König 2004] KÖNIG, Joachim: *Lernen durch Beteiligung - theoretische und konzeptionelle Überlegungen zur Qualitätsentwicklung an Hochschulen.* Krems/Österreich, 2004 (Vortrag auf der Konferenz: Qualitätsentwicklung in der postgradualen Weiterbildung: Hochschuldidaktik und Qualitätsentwicklung (7.5.2004).)
- [Kohler 2004] KOHLER, Jürgen: Quality Culture als Leitungsaufgabe. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation - ein Bestandteil des Qualitätsmanagement an Hochschulen.* Bonn, 2004 (Beiträge zur Hochschulpolitik, Bd. 9/2004), S. 57–78

- [Kreckel 1992] KRECKEL, Reinhard: *Politische Soziologie der Sozialen Ungleichheit*. Frankfurt/Main : Campus, 1992
- [Kreckel 2002] KRECKEL, Reinhard: Externe und interne Impulse zur Erneuerung der Qualitätssicherung an Hochschulen. Einige einführende Überlegungen. In: WINTER, Martin (Hrsg.) ; REIL, Thomas (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 16–20
- [Kremer 1990] KREMER, John F.: Construct Validity of Multiple Measures in Teaching, Research, and Service and Reliability of Peer Ratings. In: *Journal of Educational Psychology* 82 (1990), Nr. 2, S. 213–218
- [Kromrey 1993] KROMREY, Helmut: Studentische Vorlesungskritik. Empirische Daten und Konsequenzen für die Lehre. In: *Soziologie* (1993), Nr. 1, S. 39–56
- [Kromrey 1995] KROMREY, Helmut: Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeit ihrer Realisierung. In: *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 15 (1995), Nr. 4, S. 313–336
- [Kromrey 1998] KROMREY, Helmut: *Empirische Sozialforschung*. 8. Auflage. Opladen : Leske und Budrich (UTB), 1998
- [Kromrey 2000] KROMREY, Helmut: Qualität und Evaluation im System Hochschule. In: STOCKMANN, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 234–258
- [Kromrey 2001] KROMREY, Helmut: Studierendenbefragungen als Evaluation der Lehre? Anforderungen an Methodik und Design. In: ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. Frankfurt/Main : Campus, 2001, S. 11–48
- [Kromrey 2001a] KROMREY, Helmut: Evaluation von Lehre und Studium - Anforderungen an Methodik und Design. In: SPIEL, Christiane (Hrsg.): *Eva-*

- luation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck.*  
Münster u. a. : Waxmann, 2001a, S. 21–60
- [Kromrey 2001b] KROMREY, Helmut: Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: *Sozialwissenschaften und Berufspraxis*. 24 (2001b), Nr. 2, S. 105–132
- [Kromrey o.J.] KROMREY, Helmut: Evaluation der Lehre durch Umfrageforschung? Methodische Fallstücke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. o.J.
- [Kubinger ] KUBINGER, Klaus D.: Replik auf Jürgen Rost, Was ist aus dem Rasch-Modell geworden? Und für die Psychologische Diagnostik hat es doch eine revolutionäre Bedeutung. URL <http://www.univie.ac.at/Psychologie/diagnostik/files/Rostrep.pdf>; zuletzt 13.1.2006. – Forschungsbericht
- [Kubinger 1995] KUBINGER, Klaus D.: *Einführung in die Psychologische Diagnostik*. Weinheim : Beltz Psychologie Verlags Union, 1995
- [Kubinger und Jäger 2003] KUBINGER, Klaus D. (Hrsg.) ; JÄGER, Reinhold S. (Hrsg.): *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim u.a. : Beltz, 2003
- [Kulik 2001] KULIK, James: Student Ratings: Validity, Utility, and Controversy. In: THEALL, Michael (Hrsg.) ; ABRAMI, Philip C. (Hrsg.) ; METS, Lisa A. (Hrsg.): *The Student Ratings Debate: Are They Valid? How Can We Best Use Them?* Bd. 109. San Francisco : Jossey-Bass, 2001, S. 9–26
- [Kultusministerkonferenz ] KULTUSMINISTERKONFERENZ: Qualitätssicherung in der Lehre. URL [http://www.kmk.org/doc/besch1/BS\\_050922\\_Qualitaetssicherung\\_Lehre.pdf](http://www.kmk.org/doc/besch1/BS_050922_Qualitaetssicherung_Lehre.pdf); zuletzt 15.8.2007
- [Kultusministerkonferenz 2000] KULTUSMINISTERKONFERENZ: Realisierung der Ziele der Bologna-Erklärung. Sachstanddarstellung. / KMK. 10.11.2000 2000. – Forschungsbericht

- [Kunczik 1994] KUNCZIK, M.: *Public Relations. Konzepte und Theorien*. Köln : Böhlau, 1994
- [Lange 1983] LANGE, Elmar: Zur Entwicklung und Methodik der Evaluationsforschung in der Bundesrepublik Deutschland. In: *Zeitschrift für Soziologie* 12 (1983), Nr. 3, S. 253–270
- [Lange 2002] LANGE, Josef: Hochschule der Zukunft. In: *Deutsche Universitätszeitung (DUZ)* 21 (2002), S. 25–34
- [Langer ] LANGER, Wolfgang: *Konfirmatorische Faktorenanalyse*. – URL <http://www.soziologie.uni-halle.de/langer/lisrel/skripten/lisrelmodelle.pdf>, zuletzt 15.8.2007
- [Lee 2000] LEE, Barbara: Theories of Evaluation. In: STOCKMANN, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 127–164
- [Leszczensky und Orr 2004] LESZCZENSKY, Michael ; ORR, Dominic: *HIS-Kurzinformationen*. Bd. A2/2004: *Staatliche Hochschulfinanzierung durch indikatorgestützte Mittelverteilung. Dokumentation und Analyse der Verfahren in 11 Bundesländern*. Hannover, 2004
- [Leviton 2003] LEVITON, Laura C.: Evaluation Use: Advances, Challenges and Applications. In: *American Journal of Evaluation* 24 (2003), Nr. 4, S. 525–535
- [Lewis 2005] LEWIS, Richard: Twin Development in quality assurance in higher education - from input to output measures: from an external to an internal focus. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Qualität messen - Qualität managen. Leistungsparameter in der Hochschulentwicklung*. Bonn, 2005 (Beiträge zur Hochschulpolitik, Bd. 6/2005), S. 25–43
- [Lienert und Raatz 1994] LIENERT, G.A. ; RAATZ, U.: *Testaufbau und Testpraxis*. Weinheim : Beltz Psychologie Verlags Union, 1994
- [Lord und Novick 1968] LORD, F.M. ; NOVICK, M.R.: *Statistical Theories of Mental Test Scores*. Reading, Mass., 1968

- [Lüthje 1997] LÜTHJE, Jürgen: Qualität in Lehre und Studium. Evaluation im Verbund Norddeutscher Universitäten. In: *Forschung und Lehre* (1997), Nr. 3, S. 126–127
- [Lüthje 2003] LÜTHJE, Jürgen: Können Universitäten lernen? In: LÜTHJE, Jürgen (Hrsg.) ; NICKEL, Sigrun (Hrsg.): *Universitätsentwicklung. Strategien, Erfahrungen, Reflexionen*. Frankfurt/Main u.a. : Peter Lang, 2003, S. 43–67
- [Marsh 1982] MARSH, Herbert W.: Validity of Students' Evaluations of College Teaching: A Multitrait-Multimethod Analysis. In: *Journal of Educational Psychology* 74 (1982), Nr. 2, S. 264–279
- [Marsh 1984] MARSH, Herbert W.: Students' Evaluation of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility. In: *Journal of Educational Psychology* 76 (1984), Nr. 5, S. 707–754
- [Marsh 1987] MARSH, Herbert W.: Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research. In: *International Journal of Educational Research* 11 (1987), S. 253–388
- [Marsh und Roche 1997] MARSH, Herbert W. ; ROCHE, Lawrence A.: Making Students' Evaluations of Teaching Effectiveness Effective. The Critical Issues of Validity, Bias, and Utility. In: *American Psychologist* 52 (1997), Nr. 11, S. 1187–1197
- [Meinhold 2000] MEINHOLD, Marianne: Von der Evaluation zur Qualitätssicherung. 2000, S. 77–89
- [Mertens 1998] MERTENS, Donna M.: *Research Methods in Education and Psychology. Integrating Diversity with Quantitative and Qualitative Approaches*. Thousand Oaks : Sage, 1998
- [Mertens 2000] MERTENS, Donna M.: Institutionalizing Evaluation in the United States of America. In: STOCKMANN, Reinhard (Hrsg.): *Evaluationsfor-*

- schung. Grundlagen und ausgewählte Forschungsfelder.* Opladen : Leske und Budrich, 2000, S. 41–57
- [Mittag 2006] MITTAG, Sandra: *Qualitätssicherung an Hochschulen.* Münster u.a. : Waxmann, 2006
- [Mittag u. a. 2003] MITTAG, Sandra ; BORNMANN, Lutz ; DANIEL, Hans-Dieter: Mehrstufige Verfahren für die Evaluation von Studium und Lehre - Eine Zwischenbilanz europäischer Erfahrungen. In: SCHWARZ, Stefanie (Hrsg.) ; TEICHLER, Ulrich (Hrsg.): *Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung.* Frankfurt/Main u.a. : Campus, 2003, S. 183–206
- [Mittag und Hager 2000] MITTAG, Waldemar ; HAGER, Willi: Ein Rahmenkonzept zur Evaluation psychologischer Maßnahmen. In: HAGER, Willi (Hrsg.) ; PATRY, Jean-Luc (Hrsg.) ; BREZING, Hermann (Hrsg.): *Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien.* Bern u.a. : Huber, 2000, S. 102–128
- [Müller 1990] MÜLLER, Ernst: *Gelegentliche Gedanken über Universitäten.* Leipzig : Reclam, 1990
- [Müller 2001] MÜLLER, Walter: Zum Verhältnis von Bildung und Beruf in Deutschland. Entkopplung oder zunehmende Strukturierung? In: BERGER, Dirk (Hrsg.): *Die Erwerbsgesellschaft. Neue Ungleichheiten und Unsicherheiten.* Opladen : Leske und Budrich, 2001, S. 29–64
- [Müller 2004] MÜLLER, Wilfried: Hochschulentwicklung durch Zielvereinbarungen. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation - ein Bestandteil des Qualitätsmanagements an Hochschulen.* Bonn, 2004 (Beiträge zur Hochschulpolitik, Bd. 9/2004), S. 129–139
- [Müller-Böling 2000] MÜLLER-BÖLING, Detlef: *Die entfesselte Hochschule.* Gütersloh : Bertelsmann Stiftung, 2000
- [Müller-Böling 2001] MÜLLER-BÖLING, Detlef: Qualitätsmanagement. In:



- HANFT, Anke (Hrsg.): *Grundbegriffe des Hochschulmanagements*. Neuwied : Luchterhand, 2001 (Hochschulwesen Wissenschaft und Praxis), S. 388–394
- [Müller-Böling 2006] MÜLLER-BÖLING, Detlef: Hochschule und Profil - zwischen Humboldt und Markt? In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualität als Prinzip der Hochschulsteuerung*. Bonn, 2006 (Beiträge zur Hochschulpolitik, Bd. 1/2006), S. 15–24
- [Müller-Böling u. a. 2001] MÜLLER-BÖLING, Detlef (Hrsg.) ; HORNBOSTEL, Stefan (Hrsg.) ; BERGHOFF, Sonja (Hrsg.): *Hochschulranking. Aussagefähigkeit, Methoden, Probleme*. Gütersloh : Bertelsmann Stiftung, 2001
- [Moosbrugger und Fischbach 1999] MOOSBRUGGER, Helfried ; FISCHBACH, Andrea: How to improve the reliability and validity of a German version of Eysenck's Personality Profiler / Institut für Psychologie der Johann Wolfgang Goethe-Universität. 1999 (8). – Forschungsbericht
- [Naderer 2001] NADERER, Heike: Akkreditierung. In: HANFT, Anke (Hrsg.): *Grundbegriffe des Hochschulmanagements*. Neuwied : Luchterhand, 2001, S. 1–5
- [Nagel 2003] NAGEL, Bernhard: Bildungsfinanzierung in Deutschland. Analyse und Gestaltungsvorschläge. In: KÖHLER, Gerd (Hrsg.) ; QUAISER, Gunter (Hrsg.): *Bildung - Ware oder öffentliches Gut? Über die Finanzierung von Bildung und Wissenschaft (Dokumentation der 22. GEW-Sommerschule)*. Frankfurt/Main, 2003, S. 43–48
- [Nickel 2003] NICKEL, Sigrun: Neupositionierung zwischen Staat und Markt. Zum Zusammenhang zwischen New Public Management, Hochschulentwicklung und Rechtsform. In: LÜTHJE, Jürgen (Hrsg.) ; NICKEL, Sigrun (Hrsg.): *Universitätsentwicklung. Strategien, Erfahrungen, Reflexionen*. Frankfurt/Main u.a. : Peter Lang, 2003, S. 219–232
- [Nickel 2007] NICKEL, Sigrun: *Partizipatives Management von Universitäten. Zielvereinbarungen, Leitungsstrukturen, Staatliche Steuerung*. München : Rai-

- ner Hampp Verlag, 2007 (Universität und Gesellschaft - Schriftenreihe zur Universitätsentwicklung)
- [Nutley 2003] NUTLEY, Sandra: Bridging the Policy/Research Divide. Reflections and Lessons from the UK. Research Unit for Research Utilisation (2003)
- [Nutley u. a. 2004] NUTLEY, Sandra ; DAVIES, Huw ; WALTER, Isabel: Learning from Knowledge Management. Research Unit for Research Utilisation (2004)
- [Nutley u. a. 2003] NUTLEY, Sandra ; WALTER, Isabel ; DAVIES, Huw T. O.: From Knowing to Doing. A Framework for Understanding the Evidence-Into-Practice Agenda. In: *Evaluation* 9 (2003), Nr. 2, S. 125–148
- [Ory 2001] ORY, John C.: Faculty Thoughts and Concerns About Student Ratings. In: LEWIS, Karron G. (Hrsg.): *Techniques and Strategies for Interpreting Student Evaluations* Bd. 87. San Francisco : Jossey-Bass, 2001, S. 3–15
- [Pasternack 2004] PASTERNAK, Peer: Qualitätsorientierung an Hochschulen. Verfahren und Instrumente / HoF-Wittenberg an der Martin-Luther-Universität Halle-Wittenberg, Arbeitsberichte 5/2004. 2004. – Forschungsbericht
- [Pasternack 2005] PASTERNAK, Peer: QM an Hochschulen - Wie geht das? In: *VHV-Mitteilungen* (2005), Nr. 3, S. 8–12
- [Patton 1997] PATTON, Michael Q.: Utilization-focused Evaluation. 2nd edition. Newbury Park : Sage, 1997
- [Pawson 2002] PAWSON, Ray: Evidence-based Policy: The Promise of Realist Synthesis. In: *Evaluation* 8 (2002), Nr. 3, S. 340–358
- [Pawson 2003] PAWSON, Ray: Nothing as Practical as a Good Theory. In: *Evaluation* 9 (2003), Nr. 4, S. 471–490
- [Pellert 2002] PELLERT, Ada: Hochschule und Qualität. In: WINTER, Thomas (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 21–31

- [Pohlenz 2002] POHLENZ, Philipp: *Ergebnisberichte zur Evaluation*. Bd. 5: *Exmatrikulierte der Universität Potsdam im Studienjahr 2001/02. Befragung zu studentischen Qualitätsurteilen über Lehre und Studium*. Potsdam : Universität Potsdam, 2002
- [Pohlenz 2003] POHLENZ, Philipp: *Ergebnisberichte zur Evaluation*. Bd. 6: *Exmatrikulierte der Universität Potsdam im Studienjahr 2001/02. Befragung zu studentischen Qualitätsurteilen über Lehre und Studium. Teil II*. Potsdam : Universität Potsdam, 2003
- [Pohlenz und Hagenmüller 2005] POHLENZ, Philipp ; HAGENMÜLLER, Jan-Peter: eValuation - Neue Medien in der Lehrevaluation. In: LATTEMANN, Christoph (Hrsg.) ; KÖHLER, Thomas (Hrsg.): *Multimediale Bildungstechnologien 1: Anwendungen und Implementationen*. Frankfurt/Main u.a. : Peter Lang, 2005, S. 27–44
- [Pohlenz und Tinsner 2004] POHLENZ, Philipp ; TINSNER, Karen: *Reihe Beiträge zur Lehrevaluation*. Bd. 1: *Determinanten des Studienabbruchs. Eine empirische Untersuchung zu Gründen und Verantwortlichkeiten*. Potsdam : Universitätsverlag, 2004
- [Pohlenz u. a. 2007] POHLENZ, Philipp ; TINSNER, Karen ; SEYFRIED, Markus: *Studienabbruch. Ursachen, Probleme, Begründungen*. Saarbrücken : Vdm-Verlag, 2007
- [Popper 1993] POPPER, Karl R.: *Objektive Erkenntnis. Ein evolutionärer Entwurf*. 4. Auflage (zuerst 1972). Hamburg : Hoffmann und Campe, 1993
- [Popper 1994] POPPER, Karl R.: *Logik der Forschung*. 10. verb. u. vermehrte Aufl. (zuerst 1934). Tübingen : Mohr, 1994
- [Preskill 2004] PRESKILL, Hallie: The Transformational Power of Evaluation. Passion Purpose, and Practice. In: ALKIN, M. (Hrsg.): *Evaluation Roots. Tracing Theorists' Views and Influences*. Thousand Oaks : Sage, 2004, S. 343–355

- [Preskill und Torres 1999] PRESKILL, Hallie ; TORRES, Rosalie T.: Building Capacity for Organizational Learning Through Evaluative Inquiry. In: *Evaluation* 5 (1999), Nr. 1, S. 42–60
- [Prim und Tilman 1997] PRIM, Rolf ; TILMAN, Heribert: *Grundlagen einer kritisch-rationalen Sozialwissenschaft. Studienbuch zur Wissenschaftstheorie Karl Poppers*. Wiesbaden : Quelle und Meyer, 1997
- [Prooijen und Kloot 2001] PROOIJEN, Jan-Willem v. ; KLOOT, Willem A. van d.: Confirmatory Analysis of Exploratively Obtained Factor Structures. In: *Educational and Psychological Measurement* 61 (2001), Nr. 5, S. 777–792
- [Race und Powell 2000] RACE, Kathrin E. ; POWELL, Kristin R.: Assessing Student Perceptions of Classroom Methods and Activities in the Context of an Outcomes-Based Evaluation. In: *Evaluation Review* 24 (2000), Nr. 6, S. 635–646
- [Reil 2002] REIL, Thomas: Weiterentwicklung des Akkreditierungswesens. In: REIL, Thomas (Hrsg.) ; WINTER, Martin (Hrsg.): *Qualitätssicherung an Hochschulen. Theorie und Praxis*. Bd. Band 1. Bielefeld : Bertelsmann, 2002, S. 64–71
- [Reissert 1991] REISSERT, Reiner: Fachstudiendauer - Ist das Problem schon fixiert, welche Handlungsmöglichkeiten gibt es? In: WEBLER, Wolff-Dietrich (Hrsg.) ; OTTO, Hans-Uwe (Hrsg.): *Der Ort der Lehre in der Hochschule*. Weinheim : Deutscher Studienverlag, 1991, S. 29–60
- [Reissert 1994] REISSERT, Reiner (Hrsg.): *HIS Kurzinformation*. Bd. A8/94: *Evaluation der Lehre - interne Selbstevaluation und externe Begutachtung durch Peers*. Hannover, 1994
- [Reissert und Konnerth 2001] REISSERT, Reiner ; KONNERTH, Tanja: Evaluation von Studium und Lehre - Ein wirksames Instrument zur Qualitätsverbesserung? In: ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. Zur Qualitätsverbesserung von Studium und Lehre*. Frankfurt/Main : Campus, 2001, S. 177–194

- [Reuke 2002] REUKE, Herrmann: Zwei Seiten einer Medaille: Evaluation und Akkreditierung. In: WINTER, Thomas Reil u. m. (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 103–109
- [Rindermann 1995] RINDERMANN, Heiner: *Verbesserung der Lehre durch Veranstaltungsevaluation?* 28.9.1995 1995
- [Rindermann 1995a] RINDERMANN, Heiner: *Psychologie*. Bd. 6: *Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen*. Landau : Empirische Pädagogik, 1995a
- [Rindermann 1997] RINDERMANN, Heiner: Die studentische Beurteilung von Lehrveranstaltungen. Forschungsstand und Implikationen für den Einsatz von Lehrevaluationen. Weinheim und Basel : Beltz, 1997
- [Rindermann 2001] RINDERMANN, Heiner: Die studentische Beurteilung von Lehrveranstaltungen - Forschungsstand und Implikationen. In: SPIEL, Christiane (Hrsg.): *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster u. a. : Waxmann, 2001, S. 61–88
- [Rindermann und Amelang 1994] RINDERMANN, Heiner ; AMELANG, Manfred: Entwicklung und Erprobung eines Fragebogens zur studentischen Veranstaltungskritik. In: *Empirische Pädagogik* 8 (1994), Nr. 2, S. 131–151
- [Rossi und Freeman 1993] ROSSI, Peter H. ; FREEMAN, Howard E.: *Evaluation. A Systematic Approach*. 5th. London : Sage, 1993
- [Rost 1999] ROST, Jürgen: Was ist aus dem Rasch-Modell geworden? In: *Psychologische Rundschau* 50 (1999), S. 140–156
- [Sandfuchs und Stewart 2006] SANDFUCHS, Gabriele ; STEWART, Gerdi: *Lehrberichte an Bayerischen Universitäten*. München, 2006
- [Scarborough u. a. 1999] SCARBOROUGH, H. ; SWAN, J. ; PRESTON, J.: Knowledge Management: A Literature Review. / Institute of Personnel Development. 1999. – Forschungsbericht

- [Schade 2002] SCHADE, Angelika: Die Praxis der Akkreditierung von Studiengängen. In: WINTER, Thomas Reil u. M. (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 48–53
- [Schelsky 1963] SCHELSKY, Helmut: *Einsamkeit und Freiheit. Idee und Gestalt der deutschen Universität und ihrer Reformen*. Reinbek bei Hamburg : Rowohlt, 1963
- [Schenker-Wicki 2002] SCHENKER-WICKI, Andrea: Evaluation und Akkreditierung im Schweizer Universitätswesen. In: WINTER, Thomas (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 90–98
- [Schimank 2004] SCHIMANK, Uwe: New Governance in Hochschulen und die Auswirkungen auf die Forschung. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation - ein Bestandteil des Qualitätsmanagements an Hochschulen*. Bonn, 2004 (Beiträge zur Hochschulpolitik, Bd. 9/2004), S. 79–88
- [Schimank 2006] SCHIMANK, Uwe: Contra Leistungsindikatoren. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*. Bonn, 2006 (Beiträge zur Hochschulpolitik, Bd. 1/2006), S. 215–219
- [Schmellerleh-Engel und Keith 1998] SCHMELLERLEH-ENGEL, Karin ; KEITH, Nina: Einführung in die Analyse von linearen Strukturgleichungsmodellen: Arbeitsmaterialien und Übungsbeispiele / Institut für Psychologie der Johann-Wolfgang-Goethe Universität. 1998 (2/1998). – Forschungsbericht
- [Schmellerleh-Engel und Keith 1999] SCHMELLERLEH-ENGEL, Karin ; KEITH, Nina: Konfirmatorische Faktorenanalyse mit LISREL: Ein Übungsbeispiel mit allen Ablaufschritten / Institut für Psychologie der Johann-Wolfgang-Goethe Universität. 1999 (3/1999). – Forschungsbericht

- [Schmellerleh-Engel u. a. 2003] SCHMELLERLEH-ENGEL, Karin ; MOOSBRUGGER, Helfried ; MÜLLER, Hans: Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of Fit Measures. In: *Methods of Psychological Research Online* 8 (2003), Nr. 2, S. 23–74
- [Schmellerleh-Engel und Schweizer 2003] SCHMELLERLEH-ENGEL, Karin ; SCHWEIZER, Karl: Diskriminante Validität. In: KUBINGER, Klaus D. (Hrsg.) ; JÄGER, Reinhold S. (Hrsg.): *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim : Beltz, 2003, S. 103–110
- [Schmidt 2005] SCHMIDT, Uwe: Zwischen Messen und Verstehen. Anmerkungen zum Theoriedefizit in der deutschen Hochschulevaluation. <http://evanet.his.de> 24.5.2005, 2005. – Forschungsbericht
- [Schnell u. a. 1999] SCHNELL, Rainer ; HILL, Paul B. ; ESSER, Elke: *Methoden der empirischen Sozialforschung*. 6., völlig überarbeitete und erweiterte Auflage. München, Wien : Oldenbourg, 1999
- [Schröder 2003] SCHRÖDER, Thomas: *Leistungsorientierte Ressourcensteuerung und Anreizstrukturen im deutschen Hochschulsystem. Ein nationaler Vergleich*. Berlin : Duncker und Humblot, 2003
- [Schröder-Gronostay 2003] SCHRÖDER-GRONOSTAY, Manuela: Bedingungsfaktoren des Studienabbruchs. Ergebnisse einer retrospektiven schriftlichen Exmatrikuliertenbefragung an der Universität Kassel. In: SCHWARZ, Stefanie (Hrsg.) ; TEICHLER, Ulrich (Hrsg.): *Universitäten auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung*. Frankfurt/Main u.a. : Campus, 2003, S. 157–168
- [Schreier 2002] SCHREIER, Gerhard: Fachbezogene Evaluationsverfahren - Zwischenbilanz und Ausblick. In: WINTER, Martin (Hrsg.): *Qualitätssicherung an Hochschulen: Theorie und Praxis*. Bielefeld : Bertelsmann Verlag, 2002, S. 32–38

- [Scriven 1980] SCRIVEN, Michael: *The Logic of Evaluation*. California : Edg.-Press, 1980
- [Seidel 2001] SEIDEL, Hinrich: Sicherung und Weiterentwicklung der Qualität in Lehre und Studium als Elemente von Wettbewerb und Profilbildung. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Wettbewerb-Profilbildung-Evaluation. Qualitätssicherung von Lehre und Studium in Gegenwart und Zukunft*. Bd. 6/2001. Bonn, 2001, S. 23–37
- [Silins u. a. 2002] SILINS, Halia C. ; MULFORD, Wiliam R. ; ZARINS, Silja: Organizational Learning and School Change. In: *Educational Administration Quarterly* 38 (2002), Nr. 5, S. 613–642
- [Smith 1994] SMITH, M. F.: Evaluation: Review of the Past, Preview of the Future. In: *Evaluation Practice* 15 (1994), Nr. 3, S. 215–227
- [Spiel 2001a] SPIEL, Christiane: Der differentielle Einfluss von Bias-Variablen auf studentische Lehrveranstaltungsbewertungen. In: ENGEL, Uwe (Hrsg.): *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre*. Frankfurt/Main : Campus, 2001, S. 61–82
- [Spiel 2001b] SPIEL, Christiane: *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster u.a. : Waxmann, 2001
- [Spiel und Gössler 2000] SPIEL, Christiane ; GÖSSLER, Martin: Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende. In: *Zeitschrift für Pädagogische Psychologie* 14 (2000), Nr. 1, S. 38–47
- [Stange 2003] STANGE, Eva-Maria: Öffentliche Armut - Privatisierung der Bildungskosten? In: KÖHLER, Gerd (Hrsg.) ; QUAISER, Gunter (Hrsg.): *Bildung - Ware oder öffentliches Gut? Über die Finanzierung von Bildung und Wissenschaft (Dokumentation der 22. GEW-Sommerschule)*. Frankfurt/Main, 2003, S. 36–38
- [Starbatty 2004] STARBATTY, Joachim: Sire geben Sie Gestaltungsfreiheit!



- Mehr Effizienz und Wettbewerb an den Universitäten setzt wirkliche Autonomie voraus. In: *Forschung und Lehre* 7/2004 (2004), S. 366–368
- [Steyer und Eid 1993] STEYER, Rolf ; EID, Michael: *Messen und Testen*. Berlin u.a. : Springer, 1993
- [Stifter 2002] STIFTER, Eva P.: *Qualitätssicherung und Rechenschaftslegung an Universitäten. Evaluierung universitärer Leistungen aus rechts- und sozialwissenschaftlicher Sicht*. Wien : Böhlau, 2002
- [Stölting 2002] STÖLTING, Erhard: Wissenschaft als Sport. In: *Die Hochschule. Journal für Wissenschaft und Bildung*. 11 (2002), Nr. 2, S. 58–78
- [Stockmann 2000] STOCKMANN, Reinhard: Evaluation in Deutschland. In: DERS. (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 11–40
- [Stockmann 2002] STOCKMANN, Reinhard: Qualitätsmanagement und Evaluation. Konkurrierende oder sich ergänzende Konzepte? In: *Zeitschrift für Evaluation* 1 (2002), Nr. 2, S. 209–243
- [Stockmann 2006] STOCKMANN, Reinhard: *Sozialwissenschaftliche Evaluationsforschung*. Bd. Band 5: *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Münster u.a. : Waxmann, 2006
- [Suchman 1967] SUCHMAN, Edward A.: *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York : Russell Sage Foundation, 1967
- [Tegethoff 2005] TEGETHOFF, Hans G.: Von der Evaluation zum Qualitätsmanagement - Ein Weg zur Exzellenz in der Lehre. In: *Das Hochschulwesen* (2005), Nr. 2, S. 55–62
- [Teichler 2005] TEICHLER, Ulrich: Was ist Qualität? In: *Das Hochschulwesen* (2005), Nr. 4, S. 130–136

- [Tenbruck 1986] TENBRUCK, Friedrich H.: Das Werk Max Webers: Methodologie und Sozialwissenschaften. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38 (1986), Nr. 1, S. 13–31
- [Tent und Stelzl 1993] TENT, Lothar ; STELZL, Ingeborg: *Pädagogisch-Psychologische Diagnostik*. Göttingen u.a. : Hogrefe, 1993
- [Thoenig 2000] THOENIG, Jean-Claude: Evaluation as Usable Knowledge for Public Management Reforms. In: *Evaluation* 6 (2000), Nr. 2, S. 217–229
- [Uhl u. a. 2004] UHL, Katrin (Hrsg.) ; ULRICH, Susanne (Hrsg.) ; WENZEL, Florian M. (Hrsg.): *Evaluation politischer Bildung. Ist Wirkung messbar?* Gütersloh : Bertelsmann Stiftung, 2004
- [Ulrich und Wenzel 2003] ULRICH, Susanne ; WENZEL, Florian M.: *Partizipative Evaluation. Ein Konzept für die politische Bildung*. Gütersloh : Bertelsmann Stiftung, 2003
- [Ulrich und Wenzel 2004] ULRICH, Susanne ; WENZEL, Florian M.: Partizipative Evaluation. In: UHL, Karin (Hrsg.) ; ULRICH, Susanne (Hrsg.) ; WENZEL, Florian M. (Hrsg.): *Evaluation politischer Bildung. Ist Wirkung messbar?* Gütersloh : Bertelsmann Stiftung, 2004, S. 27–48
- [Vedung 2000] VEDUNG, Evert: Evaluation Research and Fundamental Research. In: STOCKMANN, Reihard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 103–126
- [Voßkuhle 2004] VOSSKUHLE, Andreas: Regulierte Selbstregulierung und Organisationsentwicklung. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Evaluation - ein Bestandteil des Qualitätsmanagements an Hochschulen* Bd. 9/2004. Bonn, 2004, S. 7–32
- [Vroeijenstijn 1994] VROEIJENSTIJN, Ton: Qualitätsverbesserung oder Bedrohung der Autonomie? Evaluation der Lehre und des Studiums in den Niederlanden. In: *Forschung und Lehre* (1994), Nr. 7, S. 263–266

- [Weber 1988] WEBER, Max: *Gesammelte Aufsätze zur Wissenschaftslehre*. 7. Aufl. (zuerst 1922). Tübingen : Mohr, 1988
- [Webler 1991] WEBLER, Wolf-Dietrich: Zur Steigerung des Prestiges von Lehrleistungen an deutschen Hochschulen. Das Bielefelder Memorandum. In: WEBLER, Wolf-Dietrich (Hrsg.) ; OTTO, Hans-Uwe (Hrsg.): *Der Ort der Lehre in der Hochschule*. Weinheim, 1991, S. 75–93
- [Widmer 2000] WIDMER, Thomas: Qualität der Evaluation - wenn Wissenschaft zur praktischen Kunst wird. In: STOCKMANN, Reinhard (Hrsg.): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. Opladen : Leske und Budrich, 2000, S. 77–102
- [Wilkesmann 2001] WILKESMANN, Uwe: Organisationales Lernen. In: HANFT, Anke (Hrsg.): *Grundbegriffe des Hochschulmanagements*. Neuwied : Luchterhand, 2001, S. 331–336
- [Willems u. a. 1994] WILLEMS, J. ; GIJSELAERS, W. ; BIE, D. d.: *Qualitätssorge in der Lehre. Leitfaden für die studentische Lehrevaluation*. Neuwied, Kriftel, Berlin : Luchterhand, 1994 (Das Hochschulwesen)
- [Willke 1998] WILLKE, Helmut: Organisierte Wissensarbeit. In: *Zeitschrift für Soziologie* (1998), Nr. 3/1998, S. 161–177
- [Willke 1998a] WILLKE, Helmut: *Systemtheorie III: Steuerungstheorie*. 2. Auflage. Stuttgart, 1998a
- [Wissenschaftsrat 1996] WISSENSCHAFTSRAT: *Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation*. Bonn, 1996
- [Wolf u. a. 2001] WOLF, Patrick ; SPIEL, Christiane ; PELLERT, Ada: Entwicklung eines Fragebogens zur globalen Lehrveranstaltungsevaluation - ein Balanceakt zwischen theoretischem Anspruch, Praktikabilität und Akzeptanz. In: SPIEL, Christiane (Hrsg.): *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster : Waxmann, 2001, S. 89–110

- [Wottawa 2001] WOTTAWA, Heinrich: Qualitätsmanagement durch Zielvereinbarung. In: SPIEL, Christiane (Hrsg.): *Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck*. Münster u.a. : Waxmann, 2001, S. 151–164
- [Wottawa und Thierau 1998] WOTTAWA, Heinrich ; THIERAU, Heike: *Lehrbuch Evaluation*. 2., vollst. überarb. Aufl. Bern u.a. : Huber, 1998
- [Zechlin 2006] ZECHLIN, Lothar: Interne Zielvereinbarung - Von der Zielgenerierung zum Monitoring der Leistungserbringung: Ein Proaxisbericht. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung*. Bonn, 2006 (Beiträge zur Hochschulpolitik, Bd. 1/2006), S. 131–148
- [ZEVA 2003] ZEVA: *Handbuch 2003: Qualitätssicherung in Lehre und Studium; Erst- und Folgeevaluation sowie Akkreditierung*. Hannover, Zentrale Evaluations- und Akkreditierungsagentur, 2003
- [Ziegele 1997] ZIEGELE, Frank: Grundlage der Analyse von Studienabbrüchen. Erfassung, Bewertung und Maßnahmen. In: *Beiträge zur Hochschulforschung* 20 (1997), Nr. 4, S. 435–454
- [Ziegele 2001] ZIEGELE, Frank: Indikatorgestützte Mittelvergabe. In: HANFT, Anke (Hrsg.): *Grundbegriffe des Hochschulmanagements*. Neuwied, Kriftel : Luchterhand, 2001, S. 195–201
- [Ziegele 2006] ZIEGELE, Frank: Zielvereinbarungen als Kern des Neuen Steuerungsmodells. In: HOCHSCHULREKTORENKONFERENZ (Hrsg.): *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung* Bd. 1/2006. Bonn, 2006, S. 77–105



Hochschulen wandeln sich zunehmend zu Dienstleistungsunternehmen, die sich durch den Nachweis von Qualität und Exzellenz gegen ihre Wettbewerber durchsetzen müssen. Zum Vergleich ihrer Leistungen werden verschiedene Evaluationsverfahren herangezogen. Diese stehen jedoch vielfach in der Kritik, bezüglich ihrer Eignung, Leistungen der Hochschulen adäquat abzubilden. Verfahren der Evaluation von Lehre und Studium wird vorgeworfen, dass ihre Ergebnisse bspw. durch die Fehlinterpretation hochschulstatistischer Daten und durch die subjektive Färbung studentischer Qualitätsurteile verzerrt sind. Im Zentrum der vorliegenden Arbeit steht dementsprechend die Untersuchung von potenziellen Bedrohungen der Aussagefähigkeit von Evaluationsdaten als Steuerungsinstrument für das Management von Hochschulen.