

Aus dem Zentrum für Methodenwissenschaften und Gesundheitsforschung  
der Philipps-Universität Marburg

Geschäftsführende Direktorin: Frau Prof. Dr. Erika Baum  
Institut für Medizinische Psychologie der Philipps-Universität Marburg  
Leiter: Herr Prof. Dr. Dr. Heinz-Dieter Basler

**Marburger Fragebogen zur Evaluation des Lehrangebots  
in der Medizin**

Eine Untersuchung zur Reliabilität und Dimensionalität des Marburger Fragebogens zur  
Evaluation des Lehrangebots am Fachbereich Medizin

**Inaugural-Dissertation**

zur Erlangung des Doktorgrades  
der gesamten Medizin

Dem Fachbereich Medizin  
der Philipps-Universität Marburg  
vorgelegt

von  
Katja Krebs  
aus Radebeul

Marburg 2006

Angenommen vom Fachbereich Humanmedizin der Philipps-Universität Marburg  
am 11.05.2006.

Gedruckt mit Genehmigung des Fachbereichs

**Dekan:** Prof. Dr. B. Maisch

**Referent:** Prof. Dr. Dr. H.-D. Basler

**Korreferent:** Prof. Dr. J. A. Werner

**Inhalt**

<b>1.</b>	<b>Einleitung</b>	<b>1</b>
1.1.	Problemstellung	1
1.2.	Gang der Arbeit	2
<b>2.</b>	<b>Qualitätsmanagement an Hochschulen</b>	<b>4</b>
<b>2.1.</b>	<b>Qualität</b>	<b>4</b>
2.1.1.	Definitionsansätze zum Begriff „Qualität“	5
2.1.1.1.	Qualität nach DIN-Norm	5
2.1.1.2.	Die Qualitätsdimensionen Donabedians	7
2.1.2.	Qualität in der Hochschule	9
2.1.2.1.	Anforderungen an die Hochschullehre durch Wettbewerb	9
2.1.2.2.	Verantwortung für Qualität in der Hochschullehre	10
2.1.2.3.	Qualitätsmerkmale universitärer Lehre	12
<b>2.2.</b>	<b>Theorie des Qualitätsmanagements</b>	<b>14</b>
2.2.1.	Die Differenzierung von Qualitätsmanagement, Qualitätsmanagementsystem und Qualitätssicherung	14
2.2.2.	Die Umsetzung von Qualitätspolitik	16
2.2.3.	Der Kreislaufcharakter des Qualitätsmanagements	18
<b>2.3.</b>	<b>Umsetzung von Qualitätsmanagement an Hochschulen</b>	<b>20</b>
2.3.1.	Rahmenbedingungen für Qualitätsmanagement an Hochschulen	21
2.3.2.	Qualitätsverbesserung in der Lehre durch PDCA	22
<b>2.4.</b>	<b>Fazit</b>	<b>25</b>
<b>3.</b>	<b>Evaluation als Beitrag zur Qualitätsverbesserung</b>	<b>27</b>
<b>3.1.</b>	<b>Der Evaluationsbegriff</b>	<b>27</b>
<b>3.2.</b>	<b>Arten, Funktion und Ziele von Evaluationen</b>	<b>28</b>

---

<b>3.3.</b>	<b>Evaluationen in der Hochschule</b>	32
3.3.1.	Arten und Funktionen der Hochschulevaluation	32
3.3.2.	Ziele der Hochschulevaluation	33
<b>3.4.</b>	<b>Studentische Lehrevaluation</b>	35
3.4.1.	Datenerhebungs- und Auswertungsprobleme	36
3.4.2.	Gültigkeit der Evaluationsergebnisse	37
<b>3.5.</b>	<b>Die Entwicklung systematischer Verfahren zur Qualitätsverbesserung an Hochschulen</b>	38
3.5.1.	Das niederländische Evaluationskonzept	39
3.5.2.	Netzwerke, Verbände und Agenturen zur Hochschulevaluation in Deutschland	41
<b>3.6.</b>	<b>Fazit</b>	43
<b>4.</b>	<b>Das Marburger Evaluationsmodell</b>	45
<b>4.1.</b>	<b>Der Marburger Fragebogen zur internen Evaluation und leistungsorientierten Mittelvergabe in der Lehre</b>	45
<b>4.2.</b>	<b>Entwicklung der empirischen Fragestellungen</b>	45
<b>5.</b>	<b>Methodik der empirischen Untersuchung</b>	49
<b>5.1.</b>	<b>Stichprobengewinnung</b>	49
<b>5.2.</b>	<b>Konzeption des Fragebogens</b>	49
<b>5.3.</b>	<b>Methoden zur Ermittlung der Gütekriterien</b>	51
5.3.1.	Itemanalyse	51
5.3.2.	Reliabilitätsanalyse	52

---

<b>5.4.</b>	<b>Methoden zur Überprüfung der Fragestellungen</b>	52
5.4.1.	Analyse der geschlossenen Fragen	52
5.4.1.1.	Aufarbeitung des Datensatzes	52
5.4.1.2.	Faktorenanalyse	53
5.4.1.3.	Varianzanalyse	54
5.4.1.4.	U-Test nach Mann und Whitney	55
5.4.2.	Analyse der offenen Fragen	55
5.4.2.1.	Entwicklung des Kategoriensystems	56
5.4.2.2.	Reliabilität des Kategoriensystems	56
5.4.2.3.	Frequenzanalyse	58
<b>6.</b>	<b>Ergebnisse</b>	59
<b>6.1.</b>	<b>Stichprobenbeschreibung</b>	59
<b>6.2.</b>	<b>Ergebnisse zu den Gütekriterien der 12-Item-Skala</b>	59
<b>6.3.</b>	<b>Ergebnisse der Faktorenanalyse</b>	60
6.3.1.	Faktorenanalytische Ergebnisse der 12-Item-Skala	61
6.3.1.1.	Faktorenstruktur der Veranstaltungstypen	61
6.3.1.2.	Faktorenstruktur der Skalen zur Struktur- und Prozessqualität	63
6.3.2.	Faktorenanalytische Ergebnisse der 9-Item-Skala	64
<b>6.4.</b>	<b>Ergebnisse zu den Gütekriterien der 9-Item-Skala</b>	64
<b>6.5.</b>	<b>Ergebnisse der Varianzanalyse</b>	65
6.5.1.	Varianzanalyse für die Veranstaltungstypen	65
6.5.2.	Varianzanalyse für die Studienabschnitte	66
<b>6.6.</b>	<b>Ergebnisse des U-Testes nach Mann und Whitney</b>	67
<b>6.7.</b>	<b>Ergebnisse der Analyse der offenen Fragen</b>	68
6.7.1.	Ergebnisse der Kategorienbildung	68
6.7.2.	Ergebnisse zu den Gütekriterien der Kategorienliste	69

---

6.7.2.1.	Ergebnisse der Intra-Rater-Reliabilität	69
6.7.2.2.	Ergebnisse der Inter-Rater-Reliabilität	71
6.7.3.	Ergebnisse der Häufigkeitsanalyse	72
6.7.3.1.	Frequenzanalyse der gesamten Stichprobe	72
6.7.3.2.	Frequenzanalyse getrennt für Vorklinik und Klinik	74
<b>7.</b>	<b>Diskussion</b>	<b>77</b>
<b>7.1.</b>	<b>Diskussion der Stichprobe</b>	<b>77</b>
<b>7.2.</b>	<b>Diskussion der Gütekriterien</b>	<b>77</b>
<b>7.3.</b>	<b>Dimensionalität des Fragebogens</b>	<b>78</b>
7.3.1.	Dimensionalität der 12-Item-Skala	78
7.3.2.	„Index der allgemeinen Lehrqualität“	80
<b>7.4.</b>	<b>Diskussion der Varianzanalyse und des U-Tests</b>	<b>81</b>
<b>7.5</b>	<b>Diskussion der Kategorisierung</b>	<b>82</b>
7.5.1.	Reliabilität des Kategoriensystems	83
7.5.2.	Diskussion der Ergebnisse der Frequenzanalysen	83
<b>7.6.</b>	<b>Kritische Auseinandersetzung mit den Ergebnissen</b>	<b>86</b>
<b>8.</b>	<b>Zusammenfassung und Fazit</b>	<b>88</b>
<b>9.</b>	<b>Abstract</b>	<b>95</b>
<b>10.</b>	<b>Literaturverzeichnis</b>	<b>97</b>

---

<b>11.</b>	<b>Anhang</b>	104
Anhang A:	Marburger Fragebogen zur Evaluation des Lehrangebots in der Medizin	105
Anhang B:	Kategoriensystem zur Einordnung der Meinungsäußerungen	109
Anhang C:	Intra-Rater-Reliabilität der Subkategorien	123
Anhang D:	Inter-Rater-Reliabilität der Subkategorien	126
Anhang E:	Frequenzanalyse der gesamten Stichprobe; Subkategorien	129
Anhang F:	Frequenzanalyse der Gruppen; Hauptkategorien	132
Anhang G:	Frequenzanalyse der Gruppen; Subkategorien	133
Anhang H:	Verzeichnis der akademischen Lehrer, Danksagung	136

**Tabellen und Abbildungen**

Tabelle 1:	Itemanalyse der 12- Item-Skala	60
Tabelle 2:	Hauptkomponentenanalyse mit Varimax-Rotation	62
Tabelle 3:	Itemanalyse der 9-Item-Skala	64
Tabelle 4:	Ergebnisse Post-Hoc-Test (Scheffé-Prozedur) - Veranstaltungstypen	66
Tabelle 5:	Ergebnisse Post-Hoc-Test (Scheffé-Prozedur) - Studienabschnitt	67
Tabelle 6:	Zusammenfassung Vorlesung/ Praktikum	67
Tabelle 7:	Intra-Rater-Reliabilität der Hauptkategorien	70
Tabelle 8:	Inter-Rater-Reliabilität der Hauptkategorien	71
Tabelle 9:	Frequenzanalyse der positiven Hauptkategorien	72
Tabelle 10:	Frequenzanalyse der negativen Hauptkategorien	73
Tabelle 11:	Intra-Rater-Reliabilität; Subkategorien	123
Tabelle 12:	Inter-Rater-Reliabilität; Subkategorien	126
Tabelle 13:	Frequenzanalyse; Subkategorien	129
Tabelle 14:	Positive Hauptkategorien - Häufigkeitsanalyse Vorklinik – Klinik	132
Tabelle 15.:	Negative Hauptkategorien - Häufigkeitsanalyse Vorklinik – Klinik	132
Tabelle 16:	Frequenzanalyse der Gruppen; positive Subkategorien	134
Tabelle 17:	Frequenzanalyse der Gruppen; negative Subkategorien	135
Abbildung 1:	Deming-Zyklus	19
Abbildung 2:	PDCA zur ständigen Verbesserung in der Lehre	23
Abbildung 3:	Funktionen von Evaluationen	31
Abbildung 4:	Häufigkeitsanalyse der gesamten Stichprobe	74
Abbildung 5:	Häufigkeitsanalyse der positiven Hauptkategorien	75
Abbildung 6:	Häufigkeitsanalyse der negativen Hauptkategorien	76



## **1. Einleitung**

### **1.1. Problemstellung**

Die Diskussion über die Qualität der universitären Lehre, ihre systematische Entwicklung und Sicherung gewinnt seit Beginn der 90er Jahre in Deutschland stetig an Intensität. Dies geschieht vor dem Hintergrund eines wachsenden Legitimationsdruckes gegenüber Politik und Öffentlichkeit als Folge der zunehmenden Autonomie der Hochschulen als auch vor dem Hintergrund eines zunehmenden nationalen und internationalen Wettbewerbs um Studierende, Forscher, Lehrende und finanzielle Mittel (HRK, 2003). Es bedarf, um die hochschuleigene Position in diesem Wettbewerb zu behaupten, der Entwicklung und der Ein- und Erhaltung von Standards in der Lehre, die zu international anerkannten Abschlüssen führen. Da auf lange Sicht der Erfolg eines Unternehmens im Wettbewerb aus der überlegenen Qualität seiner Produkte bzw. Dienstleistungen resultiert (Kamiske & Brauer, 1995), ist die Qualität der Lehre für Hochschulen zu einem wesentlichen Wettbewerbsfaktor auf dem Ausbildungsmarkt geworden. Die dabei Anwendung findenden Methoden zur Qualitätskontrolle und Qualitätsverbesserung sind angelehnt an die Konzepte des aus der Wirtschaftswissenschaft stammenden Qualitätsmanagements. Das im Hochschulbereich in der Regel eingesetzte Prüfverfahren zur Qualitätskontrolle ist die Evaluation. Die Evaluation der Qualität von Lehrveranstaltungen und die Bekanntmachung der Ergebnisse wurden durch die am 1. Oktober 2003 in Kraft getretene Approbationsordnung für Ärzte (AppOÄ) verbindlich vorgeschrieben (Bundesgesetzblatt, 2002, S. 2407). Die Hochschulen stehen somit vor der Aufgabe, Institutionen und Inhalte der Evaluation zu definieren. Externe Evaluationen werden durch beauftragte Agenturen wie z.B. in Niedersachsen (ZEvA, 2003), vorgenommen, interne Evaluationen erfolgen in der Regel durch die jeweiligen Dekanate der Fakultäten oder Qualitätsmanagementorganisationen innerhalb der Universität.

An der Philipps-Universität Marburg hat sich der Fachbereich Humanmedizin für eine interne Evaluation ausgesprochen. Die Kommission des Studiausschusses zur leistungsbezogenen Mittelvergabe in der Lehre unter der Leitung des Fachvertreters der Medizinischen Psychologie entschied, in Orientierung an den Ausführungen Rindermanns (2001) und in Anlehnung an Donabedians Konzept zur Beurteilung der

Qualität medizinischer Versorgung, auch die Qualität der Lehre unter den Aspekten Prozess-, Struktur- und Ergebnisqualität zu bewerten (Donabedian, 2003). Als Qualitätsmerkmal für das Ergebnis der Lehre sollen die Leistungen der Studierenden in den medizinischen Staatsexamina heran gezogen werden. Struktur- und Prozessqualität sollen durch die Studierenden mittels des Marburger Fragebogens zur Evaluation des Lehrangebotes in der Medizin bewertet werden. Bei der leistungsbezogenen Mittelvergabe sollen die Ergebnisse aller drei Qualitätsmerkmale berücksichtigt werden (Basler & Krebs, 2003). Es wurde ein Punktesystem entwickelt, welches eine Übertragung der gewonnenen Ergebnisse und ein Ranking der einzelnen Fachgebiete in der medizinischen Lehre ermöglichen soll. Dem jeweils erreichten Rang wird ein Geldäquivalent zugeordnet. Damit die Urteile der Studierenden über Struktur- und Prozessqualität in der Lehre in das Punktesystem eingehen kann, muss der Marburger Fragebogen verschiedenen statistische Anforderungen erfüllen, die im Rahmen dieser Arbeit zu überprüfen sind.

In der vorliegenden Arbeit wird der spezifische Beitrag der Lehrevaluation zur Qualitätsverbesserung in der Lehre im Kontext eines Qualitätsmanagements der Hochschule diskutiert. Vor diesem Hintergrund sollen im empirisch-statistischen Anteil der Arbeit der Marburger Fragebogen zur Evaluation des Lehrangebotes in der Medizin beschrieben werden sowie seine Dimensionalität und die psychometrischen Gütekriterien überprüft werden. Ziel ist es, die Entscheidung darüber zu ermöglichen, ob der Marburger Fragebogen die notwendigen Voraussetzungen erfüllt um einen Beitrag zur Qualitätsverbesserung in der Lehre zu leisten sowie eine Ermessensgrundlage für die leistungsorientierte Mittelvergabe in der Lehre zu sein.

## **1.2. Gang der Arbeit**

Die zentralen Themen dieser Arbeit sind die Qualität der Lehre und die Evaluation der Lehre im Hochschulbereich. Evaluationen dienen dabei als Instrument der Qualitätskontrolle und können mittelbar einen Beitrag zur Qualitätsverbesserung leisten. Qualitätskontrolle und Qualitätsverbesserung sind Teilbereiche des aus der Wirtschaft stammenden Qualitätsmanagements. Im zweiten Kapitel werden daher zunächst der Begriff Qualität und die Theorie des Qualitätsmanagements erörtert und auf die Bereiche Hochschule sowie Hochschullehre übertragen. Die Umsetzungsmöglichkeit

eines Qualitätsmanagements an der Hochschule sowie die dafür erforderlichen Rahmenbedingungen werden anschließend diskutiert.

Im dritten Kapitel wird die im Rahmen des Qualitätsmanagements notwendige Qualitätskontrolle herausgegriffen und speziell die im Hochschulbereich in der Regel eingesetzte Evaluation als Prüfverfahren mit seinen Arten, Gegenständen, Ebenen, Funktionen und Zielen dargestellt. Die sich bereits international und national etablierten Evaluationsverfahren beinhalten ebenso wie das im vierten Kapitel vorgestellte Marburger Evaluationsmodell die studentische Lehrevaluation. Daher werden die Möglichkeiten und Grenzen der studentischen Lehrevaluation in diesem Kapitel kritisch betrachtet. Abschließend wird ein Überblick über national und international etablierte Evaluationsverfahren gegeben.

Im vierten Kapitel erfolgt die Vorstellung des Marburger Evaluationsverfahrens. Der Marburger Evaluationsbogen zur Struktur- und Prozessqualität des Lehrangebotes in der Medizin muss, um im Rahmen der leistungsorientierten Mittelvergabe eingesetzt werden zu können, bestimmte Anforderungen erfüllen, die im empirisch-statistischen Anteil der vorliegenden Arbeit untersucht werden. Im vierten Kapitel wird daher zunächst die empirisch-statistische Fragestellung der Arbeit erörtert und die zu untersuchenden Aspekte des Marburger Evaluationsbogens aufgeführt. Die bei der statistischen Untersuchung angewandten Methoden werden im fünften Kapitel vorgestellt und die ausführliche Darstellung der Ergebnisse und deren Diskussion schließen sich mit den darauf folgenden Kapiteln sechs und sieben an.

Abschließend folgt im achten Kapitel die zusammenfassende Schilderung der in dieser Arbeit diskutierten Themen des Qualitätsmanagements im Hochschulbereich, der Qualität in der Lehre, der Lehrevaluation sowie der durch die Untersuchung des Fragebogens gewonnenen Ergebnisse und Schlussfolgerungen.

## 2. Qualitätsmanagement an Hochschulen

Qualität ist in jedem Unternehmen und in jeder Branche ein zentrales Thema. Das von den Unternehmen eingesetzte Qualitätsmanagement zur stetigen Qualitätsverbesserung von Produkten und Dienstleistungen hat sich in den letzten Jahren zu einem immer komplexeren Wissensgebiet entwickelt und findet auch zunehmend in Abwandlung Anwendung im Hochschulbereich mit dem Ziel, Qualität der Lehre und des Studiums zu sichern und zu verbessern.

In diesem Kapitel wird nach Erläuterung des Begriffs „Qualität“ und der Theorie des Qualitätsmanagements der Frage nachgegangen, in wie weit und in welcher Form die Übertragung eines Qualitätsmanagementkonzeptes auf den Hochschulsektor und speziell auf den Bereich der universitären Lehre möglich ist. Dazu wird in den weiteren Ausführungen die Hochschule modellhaft als Unternehmen betrachtet. Es sei zugleich auf die unterschiedliche Zielsetzung von Unternehmen und Hochschulen hingewiesen, die eine völlige Gleichsetzung dieser verbieten. Unternehmen sind orientiert an Unternehmenserhalt und Gewinn. Hochschulen haben vorwiegendes Interesse an Lehre und Forschung. Ihre erbrachten Leistungen sind schwer messbar und es herrschte eine ausgeprägte Organisationskultur der einzelnen Einheiten (Müller-Böling & Schreiterer, 1999).

### 2.1. Qualität

In den letzten Jahren hat sich die Diskussion über Qualität an der Hochschule und über Qualität in der Lehre intensiviert. Zu Beginn ist daher der Begriff der „Qualität“ zu definieren. Dabei wird der erste Schwerpunkt auf den wirtschaftswissenschaftlichen Definitionsansatz gelegt, da auf diesem Ansatz die Theorien des Qualitätsmanagement beruhen. Als zweites wird der in der Medizin verbreitete Qualitätsbegriff Donebedians erläutert, der Eingang in die Konzeption des Marburger Evaluationsbogens und das Konzept der Evaluation am Fachbereich Medizin der Philipps-Universität Marburg findet.

### **2.1.1. Definitionsansätze zum Begriff „Qualität“**

Das Qualitätsverständnis unterlag in den letzten 50 Jahren einem Wandel vom spezifikationsorientierten zum kundenorientierten Qualitätsbegriff. Noch in den 50er Jahren verstanden die Unternehmen unter Qualität fast ausschließlich Produkt- bzw. Sachqualität. Qualitativ hochwertige Ware war fehlerfrei und besaß bestimmte Leistungsmerkmale (Kamiske & Brauer, 1995). Die technischen Anforderungen an ein Produkt wurden von den Unternehmen, Produzenten und Herstellern bestimmt. Wurden die unternehmensintern festgelegten Spezifikationen von einem Produkt erfüllt, so war es von Qualität. Jedoch sind die Anforderungen des Marktes seit den 70er Jahren geprägt durch den Übergang von einem ungesättigten zu einem gesättigten Markt, der sich durch Verdrängungswettbewerb auszeichnet (Hannen & Lindemann, 1996). Es kam zu einer Zunahme von Anbietern vergleichbarer Produkte und Dienstleistungen. Angesichts dieser Entwicklung wurde die Qualität der Produkte bzw. Dienstleistungen zum entscheidenden Wettbewerbsfaktor, der den wirtschaftlichen Erfolg der Unternehmen maßgebend beeinflusst (Jaster, 1997). Der Markt ist zu einem Käufermarkt geworden, in dem nun der Kunde ein sich ständig erweiterndes Qualitätsverständnis definiert (Reinhardt et al., 1996). Um sich im heutigen Verdrängungswettbewerb zu behaupten und sich gegenüber den Wettbewerbern zu profilieren, sind Kundenorientierung und der Kontakt und die Kommunikation mit dem Kunden unabdingbar geworden. Die Ermittlung der Kundenwünsche wird somit eine wichtige Voraussetzung in der Qualitätserzeugung (Pfitzinger, 1998). Die im Folgenden dargestellten Begriffe von Qualität beruhen auf dem neuen Marktprinzip und sind für die Hochschule insofern die relevante Interpretation, da ein zunehmender Wettbewerb auch zwischen den Hochschulen wahrgenommen wird.

#### **2.1.1.1. Qualität nach DIN-Norm**

Die Norm DIN EN ISO 8402 (1995) definiert Qualität als die „Gesamtheit von Merkmalen einer Einheit, bezüglich ihrer Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen“. Eine Einheit wird dabei als materieller oder immaterieller Gegenstand der Betrachtung aufgefasst, also als Produkt oder Dienstleistung. Bei dieser formalen Begriffsklärung des Deutschen Institutes für Normierung (DIN) wird nicht das Produkt bzw. die Dienstleistung allein betrachtet, sondern die Gesamtheit von

Merkmale der dem Kunden angebotenen Leistungen sowie deren Zusammenwirken. Die Beschaffenheit einer Einheit ergibt sich dann aus der Gesamtheit aller so genannten Qualitätsmerkmale und Qualität wird damit zum Gesamteindruck aus Teilqualitäten. Jedes einzelne Merkmal entscheidet über den Qualitätszustand einer Einheit. Die mangelhafte Erfüllung einer Eigenschaft kann dabei nicht durch die Übererfüllung eines anderen kompensiert werden (Reinhardt et al., 1996).

Vereinfacht kann Qualität somit als die Übereinstimmung von Leistungen mit Anforderungen definiert werden. Daraus ergibt sich die Frage, wer Anforderungen stellt. Sie werden gestellt von Konsumenten, Produzenten, Kunden und Anbieter von Dienstleistungen. Die Beurteilung von Qualität leitet sich aus den individuellen Bedürfnissen ab, die auf objektiv feststellbaren und subjektiv zugeschriebenen Produkteigenschaften basieren. Entscheidend hierbei ist, welche Eigenschaften vor dem Hintergrund der Anforderungen wahrgenommen und für qualitätsentscheidend erachtet werden. Der Qualitätsbegriff kann folglich objektiv und subjektiv betrachtet werden. Liegen genau festgelegte Anforderungen vor, die zur Bestimmung der Qualität herangezogen werden können, ist deren Objektivierung möglich. Die Beurteilung der Qualität einer Leistung kann anhand von Kriterien, Standards, Indikatoren oder Normen erfolgen. Der subjektive Qualitätsbegriff beinhaltet die individuellen Präferenzen des Kunden. Je mehr die Leistung des Anbieters mit den Anforderungen des Kunden übereinstimmt, desto höher wird von diesem die Qualität des Produktes oder der Dienstleistung empfunden. Somit ist Qualität im Urteil des Kunden immer subjektiv (Pepels, 1996). Der Begriff Qualität beschreibt gewissermaßen das Verhältnis zwischen dem, was der Kunde von einem Gegenstand oder einer Dienstleistung erwartet und dem, wie diese Erwartung tatsächlich realisiert wurde. Es geht also um das Verhältnis zwischen einem erwarteten Idealzustand und dem tatsächlichen „Ist“-Zustand.

Es zeigt sich, dass es sich bei dem Begriff „Qualität“ um ein mehrdimensionales Phänomen handelt, das sich einer einheitlichen, allumfassenden Definition zu entziehen scheint. Qualität ist mehr als die reine „Beschaffenheit“ einer Sache und geht weit über die technische Erfüllung von Anforderungen und Spezifikationen hinaus. Je nach Blickwinkel und Standpunkt geraten immer wieder andere Qualitätsaspekt und Anforderungen in den Vordergrund. Diese Anforderungen entsprechen den zu erfüllenden Qualitätsmerkmalen, welche anwendungs- und zielgruppenorientiert

definiert werden müssen. Damit kann Qualität nicht als statische Größe verstanden werden. Vielmehr stellt sie eine dynamische Größe dar, da sie mit den sich ändernden Kundenbedürfnissen und Anforderungen an Produkte und Dienstleistungen einem Wandel unterzogen ist und in dem Bestreben um Qualitätsverbesserung stetig neu zu erheben ist.

### 2.1.1.2. Die Qualitätsdimensionen Donabedians

In Erweiterung dieses sehr ergebnisbezogenen Qualitätsbegriffes nach DIN-Norm schlägt Donabedian (1980) in seinen grundlegenden Überlegungen eine Unterteilung des Qualitätsbegriffes in drei Dimensionen vor: Strukturqualität, Prozessqualität und Ergebnisqualität. Dahinter steht der Annahme, dass optimale Ergebnisqualität nur erreicht werden kann, wenn auch die Strukturen und Prozesse, die für die Herstellung eines Produktes oder die Erbringung einer Dienstleistung maßgeblich sind, optimal gestaltet sind.

Die **Strukturqualität** umfasst die strukturellen Charakteristika und Rahmenbedingungen einer Einrichtung, z.B. die Anzahl und die Qualifikation der Mitarbeiter sowie die Qualität und Quantität der weiteren zur Erstellung der Leistung notwendigen Strukturen wie Organisation, finanzielle Ausstattung, Infrastruktur und Gebäude. Mit **Prozessqualität** sind jene Parameter gemeint, die die Durchführung und die Erbringung der Leistung prägen, also Arbeitsabläufe jeder Art. Ein Prozess ist dabei als der systematische Ablauf von Aktivitäten zur Verwirklichung eines oder mehrerer Ziele zu verstehen (Reinhard et al., 1996). Die angebotenen Dienstleistungen einer Einrichtung entstehen in einer Ablaufkette, die sich aus unterschiedlichen Aktivitäten zusammensetzt. Zum Erfolg des gesamten Prozesses tragen alle Teilprozesse bei und müssen daher effektiv und effizient ablaufen. Donabedian unterscheidet dabei zwischen Kern- und Hilfsprozessen. Kernprozesse sind dem Einrichtungszweck dienlich, z.B. in einer Bildungseinrichtung der Unterricht. Hilfsprozesse unterstützen die Kernprozesse, z.B. die Verwaltung. Unter **Ergebnisqualität** sind schließlich all jene Parameter zu verstehen, die auf einen Erfolg oder Misserfolg der Dienstleistung schließen lassen. Sie kann sowohl anhand objektiver Veränderungen als auch anhand subjektiver Kriterien gemessen werden. In Bildungseinrichtungen sind unter Ergebnisqualität der Lernerfolg

der Studierenden, ein erfolgreicher Studienabschluss und die Bewährung der Absolventen im Berufsleben zu verstehen.

Donabedian weist allerdings darauf hin, wie undifferenziert und theoretisch-modellhaft die von ihm vorgenommene Gliederung der Qualität in drei Komponenten ist. Diese drei Qualitätsdimensionen sind in einem funktionalen Zusammenhang zu sehen. So bewirkt eine hohe Strukturqualität eine hohe Prozessqualität, welche wiederum eine gute Ergebnisqualität nach sich zieht. Es ist aber zu beachten, dass die Dimensionen nicht nur in einem linearen Zusammenhang, sondern auch in interdependenter Wechselbeziehung zueinander stehen. Eine hohe Ergebnisqualität kann über eine gute Reputation zu höheren Studierendenzahlen und höheren finanziellen Mittelzuweisungen führen. Diese wiederum können zur Verbesserung der Struktur- und Prozessqualität eingesetzt werden. Donabedian sieht in seinem Modell eher einen Ansatz zur Qualitätsmessung („approach to quality assessment“) als eine allumfassende Definition des Begriffs Qualität und betont, dass seine vorgenommene Einteilung zur Messung der Qualität in die drei Kategorien Struktur, Prozess und Ergebnis offen für Modifikationen ist, so lange diese berechtigt und hilfreich sind und so lange sie richtig verstanden und angewendet werden.

Im Marburger Modell der internen Evaluation findet eine Übertragung der Qualitätsdimensionen Donabedians auf den Hochschulsektor und speziell auf den Bereich der Hochschullehre statt. Der in dieser Arbeit untersuchte Fragebogen zur Evaluation der medizinischen Lehre beinhaltet Items zur Operationalisierung von Struktur- und Prozessqualität. Zur Operationalisierung der Ergebnisqualität dienen die Ergebnisse der schriftlichen Examina des Institutes für Medizinische und Pharmazeutische Prüfungsfragen (IMPP). Zusammengeführt sollen die Resultate der studentischen Lehrevaluation (Daten zur Struktur- und Prozessqualität) und die Prüfungsergebnisse (Daten zur Ergebnisqualität) im Rahmen von Qualitätssicherung und -verbesserung in der medizinischen Lehre eine Bestandsaufnahme sowie eine Diskussionsgrundlage darstellen. Gleichzeitig können die Ergebnisse als Entscheidungsgrundlage für Reformen und leistungsorientierte Mittelvergabe herangezogen werden.



### **2.1.2. Qualität in der Hochschule**

Bei Betrachtung von Qualität in der Hochschule kann entsprechend der Aufgaben der Hochschule in eine Qualität der Lehre und eine Qualität der Forschung differenziert werden. Aufgrund der Thematik dieser Arbeit werden sich die weiteren Ausführungen auf den Bereich der Hochschullehre konzentrieren.

#### **2.1.2.1. Anforderungen an die Hochschullehre durch Wettbewerb**

Die aktuelle Situation der Hochschulen wird im Folgenden vor dem Hintergrund des Wettbewerbs zunächst näher betrachtet, um daraus anschließend die Anforderungen an die Hochschullehre und speziell die Anforderungen an Struktur-, Prozess- und Ergebnisqualität in der Lehre abzuleiten.

Ziel von Studium und Lehre an Universitäten und Fachhochschulen ist gemäß Hochschulrahmengesetz die berufsbefähigende Bildung und Ausbildung bis zu einem ersten berufsqualifizierenden Abschluss. Wird die Bewährung der Absolventen im Berufsleben als Ergebnisqualität der Lehre betrachtet, so müssen auch die Erwartungen von zukünftigen Arbeitgebern an Absolventen bei der Konzeption von Lehrstrukturen und Lehrprozessen Berücksichtigung finden. Um dieser Anforderung und dem Wettbewerb Rechnung zu tragen, wurde für deutsche Hochschulen in den letzten Jahren die Notwendigkeit zur Veränderung der Studieninhalte und Curricula zu einem zentralen Thema (Moewes, 2002). Ein modernes Ausbildungsangebot sollte Nachfrageänderungen des Arbeitsmarktes hinsichtlich der Kompetenzen von Hochschulabsolventen einbeziehen. Damit die Hochschulen sich dessen bewusst sind, was ihre Absolventen in der Berufspraxis erwartet, müssen sie in einem ständigen Dialog mit den Vertretern der Arbeitswelt ihre Ausbildungsangebote für diese entwickeln und überprüfen lassen. Um ein qualitativ hochwertiges Lehrangebot zu verwirklichen, bedarf es entsprechend qualifizierter Mitarbeiter, um die die Hochschulen wiederum im Wettbewerb untereinander und mit der freien Wirtschaft stehen. Denn neben den staatlichen Hochschulen etablieren sich immer mehr private Hochschulen, die sich selbst finanzieren und eine qualitativ hochwertige Ausstattung mit Ressourcen und Personal vorweisen können. Die durch den Staat zur Verfügung gestellten Haushaltsmittel orientieren sich unter anderem an der Anzahl der

Studierenden einer Hochschule. Die Konsequenz daraus wäre der Wettbewerb um eine möglichst hohe Anzahl an Studierenden. Gleichzeitig gereicht es der Hochschule zum Vorteil, wenn neben der Quantität auch eine hohe Qualität der Studierenden erreicht wird, da dies zu guten Absolventen führt und sich damit positiv auf die Reputation der Einrichtung auswirkt, welche wiederum zu höheren Drittmittelzahlungen führen könnte. Werden die Leistungen der Hochschulen transparent gemacht, z.B. durch Veröffentlichung von Lehrberichten und Evaluationsergebnissen, so können neue Studienbewerber die Qualität einer Hochschule abschätzen und diese gezielt aussuchen. Die Folge eines attraktiven Studienangebotes, qualifizierter und motivierter Mitarbeitern und herausragender Reputation ist, dass auch exzellente Studierende an die Hochschule kommen und somit durch die Qualität des Angebotes neben der Quantität auch die Qualität der Nachfrage bestimmt wird (Moewes, 2002). Mit Erhöhung der Anzahl an Studierenden müssen allerdings auch die strukturellen Ressourcen erweitert werden. Weiss (2000) weist darauf hin, dass vor 40 Jahren nur 5%, heute dagegen 35% eines Jahrgangs ein Hochschulstudium aufnehmen, aber Ausbau wie Ausstattung mit dieser Entwicklung nicht Schritt gehalten haben. Daraus ergibt sich die Notwendigkeit der Verbesserung der Strukturqualität in der Lehre, denn angesichts der Überfüllung der Hochschulen muss von einem Verlust an Lehrqualität aufgrund ungünstiger Betreuungsrelation ausgegangen werden (Schilling, 1994). Mängel in Bereich der Strukturqualität wie überfüllte Hörsäle, veraltete Einrichtungen und Personalmangel wirken sich wiederum negativ auf den Lehr – und Lernprozess aus, welches wiederum die Ergebnisqualität nachteilig beeinflussen kann.

Letztendlich wird der Wettbewerb um finanzielle Mittel, Studierende und Mitarbeiter zur Triebfeder aller Qualitätsbemühungen. Es wird somit zur Aufgabe der Hochschulen Leistungen transparent zu machen und für das Lehrangebot hinsichtlich der Aspekte der Qualität der Lehre wie Praxisnähe, Organisation und Betreuung, Strukturiertheit, materielle Ausstattung und Studierbarkeit in Regelstudienzeit Standards zu entwickeln und deren Einhaltung ständig zu überprüfen, um Mängel in der Struktur-, Prozess- und Ergebnisqualität zu erkennen und Verbesserungen zu erwirken.

### **2.1.2.2. Verantwortung für Qualität in der Hochschullehre**

Im Bereich der Hochschullehre sind Lehrende, Studierende, Angehörige der Verwaltung, Fachgesellschaften, Berufsverbände, Bildungs- und Finanzpolitiker sowie Arbeitgeber- und Wirtschaftsvereinigungen an der Qualität der Lehre interessiert und zur Festlegung von Qualitätsmaßstäben bzw. Qualitätsmerkmalen grundsätzlich berechtigt (Arnold, 2003). Jede dieser Gruppen ist für die Gestaltung der Hochschullehre und deren Qualität auf unterschiedliche Weise und unterschiedlicher Ebene verantwortlich, welche den organisatorischen Ebenen bei Erbringung der Lehre entsprechen: Hochschule, Fachbereiche, einzelne Lehrveranstaltung. Im Folgenden wird auf diese Ebenen und die entsprechenden Verantwortlichkeiten näher eingegangen und in Bezug zu den Qualitätsdimensionen Donabedians gesetzt.

Das Setzen von Rahmenbedingungen und die Bereitstellung von materiellen Ressourcen liegen in der Hand der Wissenschaftsministerien, der Hochschulleitung und Verwaltung. Sie sind für die Qualität der Lehre auf der Ebene der Strukturqualität verantwortlich, welche die Gesamtheit der materiellen und personellen Ressourcen, die zur Erbringung der Lehre verfügbar sind sowie gesetzliche und organisatorische Regelungen umfasst. Für die Sicherung und Entwicklung von Prozessqualität in der Lehre sind die Fachbereiche, die Lehrenden und die Studierenden verantwortlich. Aufgabe der Fachbereiche ist die Koordination der Lehrplanung und des Prüfungswesens sowie die Vergabe der Räumlichkeiten. Im Verantwortungsbereich der Hochschullehrenden liegen die inhaltliche und didaktische Gestaltung von Lehrveranstaltungen, Sprechstunden und Prüfungen. Auch die Studierenden gestalten die Prozessqualität in der Lehre entscheidend mit, denn Lehre entwickelt sich aus der Interaktion zwischen Lehrenden und Lernenden (Kromrey, 2001). Dieser Prozess der Kommunikation erfordert auch einen Beitrag von studentischer Seite in Form von Vorbereitung auf und Beteiligung an der Veranstaltung. Prozessqualität beinhaltet somit alle Interaktionen, die im Zuge der Leistungserbringung stattfinden. Dem Studierenden kommt neben der Rolle des Kunden also auch die Rolle des aktiven Mitarbeiters zu. Die Tatsache, dass Prozesse dabei von Anbietern und Kunden gemeinsam gestaltet werden, ist charakteristisch für den Dienstleistungssektor und Bildungsbereich (Heid, 2000). Weniger für die Prozessqualität als für die Ergebnisqualität der Lehre in Form von Fachwissen, Berufsfähigkeit und der Beherrschung wissenschaftlicher Arbeitsmethoden

interessieren sich zukünftige Arbeitgeber, Bildungspolitiker und weite Teile der Öffentlichkeit. Eine Besonderheit und ebenso eine Problematik der Ergebnisqualität im Bildungsbereich ist, dass sie sich erst nach Studienabschluss feststellen lässt, indem betrachtet wird, wie die Absolventen mittels im Studium erworbenen Wissens und Fertigkeiten berufliche oder gesellschaftliche Aufgaben bewältigen.

### **2.1.2.3. Qualitätsmerkmale universitärer Lehre**

Als Qualitätsmerkmale kommen im Allgemeinen die Eigenschaften des betrachteten Gegenstandes in Frage. Lehre ist allerdings kein Gegenstand oder Sachverhalt, vielmehr eine Dienstleistung, deren Produkt sich aus der Interaktion der Lehrenden und Lernenden ergibt. Qualität der Lehre ist somit keine einem Objekt zuschreibbare, sondern eine relationale Eigenschaft (Kromrey, 2000). Soll Lehre ihrem Charakter als Dienstleistung gerecht werden, so unterliegen auch die Qualitätsmerkmale guter Lehre einem dynamischen Prozess.

Aus den in Abschnitt 2.1.2.1. dargestellten Anforderungen an die Hochschullehre lassen sich in Anlehnung an Willems et al. (1994) folgende Kriterien ableiten, nach denen die Qualität der universitären Lehre beurteilt werden kann und die Übertragung der Qualitätsdimensionen Donabedians auf den Bereich der universitären Lehre ermöglichen.

- Lehrausstattung und Curriculum (Strukturqualität)

Bei der Ausstattung der Lehre werden die räumlichen Ausstattungen, die Bibliotheken, die Arbeits- und Lernbedingungen und das studentische Umfeld beurteilt. Die Konzeption des Studiums, d.h. die strukturierte und inhaltlich aufeinander abgestimmte Abfolge von Veranstaltungen ohne terminliche und inhaltliche Überschneidungen oder Lücken ist ein wesentliches Merkmal der Strukturqualität.

- Lehrpräsentation (Prozessqualität)

Die Umsetzung des Curriculums erfolgt durch die Präsentation des Lehrangebotes. Die Qualität der Lehrpräsentation differenziert Multrus (1995) in didaktische, tutoriale und strukturelle Qualität. Qualität in der Lehre setzt didaktische Kompetenz des Lehrpersonals voraus. Kriterien, die für eine „gute“ Didaktik stehen, sind Motivation,

Engagement, rhetorische Kompetenz, strukturierte, verständliche Erklärungen und Darstellungen der Lehrinhalte sowie die Fähigkeit Interesse zu wecken. Unter tutorialer Qualität ist die Gesprächsbereitschaft von Lehrenden und Mitarbeitern, die Beratung und Betreuung der Studierenden durch Lehrende und die Offenheit für Fragen in den Lehrveranstaltungen sowie für Gespräche und Diskussionen außerhalb von Veranstaltungen zu verstehen. Neben diesen Aspekten, welche unter Prozessqualität zusammengefasst werden können, ist die Güte der Lehrpräsentation auch abhängig von der strukturellen Qualität. Darunter sind die verwendeten Mittel, Materialien und Skripte zur Lehrpräsentation als auch die Bekanntmachung von Ablauf, Zielen und Anforderungen der Lehrveranstaltungen und Prüfungen zu verstehen.

- Bezug zum Arbeitsmarkt (Ergebnisqualität)

Die akademische Lehre ist darauf ausgerichtet, Studierende auf die spätere Berufstätigkeit durch Vermittlung der entsprechenden Sachkenntnisse und Fertigkeiten vorzubereiten und somit den erfolgreichen Eintritt in das Berufsfeld zu ermöglichen. Je mehr die Lernziele des Studienganges aus dem Berufsprofil abgeleitet werden, umso mehr kann die Lehre der oben genannten Funktion gerecht werden und eine entsprechend gute Ergebnisqualität erzielen.

Zusammenfassend ist zu sagen, dass Qualität in der Lehre nicht nur eine Frage der Lehrkraft, des Unterrichts und der Entstehung sowie organisatorischen Umsetzung des Lehrplans ist. Ebenso ist es eine Frage der Gestaltung des sozialen Umfeldes und der technischen Ausstattung (Dembski & Lorenz, 1995). Es ist festzustellen, dass die Qualität der Lehre von einer Vielzahl personen- und sachbezogener Faktoren geprägt wird. Personenbezogene Faktoren sind die Interessen, die Motivation und die Vorkenntnisse von Studierenden sowie die didaktische Kompetenz und das Engagement der Lehrenden. Sie haben einen erheblichen Einfluss auf die soziale und kommunikative Beziehung im Lehr- und Lernprozess. Es gilt zu beachten, dass es nicht allein ausreicht gut zu lehren. Die Lernenden müssen auch erfolgreich lernen, d.h. vermittelte Inhalte wahrnehmen, interpretieren, reflektieren, kritisieren und verinnerlichen. Damit ist die Optimierung der Lehre in gleicher Intensität Aufgabe der Lehrenden und der Studierenden. Sachbezogene Faktoren entsprechen weitestgehend den Aspekten der Strukturqualität mit räumlicher Ausstattung sowie Strukturiertheit und Transparenz bezüglich Organisation und Anforderungen eines Studiengangs.

Was als Qualität der Lehre und deren Qualitätskriterien definiert wird, unterliegt einer fortlaufenden Entwicklung, in die der Vergleich mit anderen Lehrinrichtungen ebenso eingeht wie die als positiv empfundene Weiterentwicklung der eigenen Arbeits- und Lernbedingungen. Auch die Qualität in der Lehre ist keine statische, sondern eine dynamische Größe im Prozess der Qualitätsverbesserung. Daher empfiehlt Fischer-Bluhm (2000) bestehende Qualitätsdefinitionen in gewissen zeitlichen Abständen zu hinterfragen und ebenso einmal festgelegte Qualitätskriterien weiter zu entwickeln.

## **2.2. Theorie des Qualitätsmanagement**

Mit dem Wandel des Qualitätsverständnisses vom fehlerfreien hin zu einem allen Anforderungen gerecht werdenden Produkt vollzog sich auch der Übergang von der produktbezogenen Qualitätskontrolle über die prozessorientierte Qualitätssicherung hin zum unternehmensweiten Qualitätsmanagement. Das ganze Unternehmen mit seinen Mitarbeitern, die Beziehungen zwischen Lieferanten und Kunden, der gesamte Lebenszyklus eines Produktes und alle Prozesse sind zu Objekten der Qualitätsbetrachtung geworden (Reinhardt et al., 1996). Während im vorherigen Abschnitt der Qualitätsbegriff definiert und erläutert wurde, setzt sich dieser Abschnitt mit dem Qualitätsmanagement und den damit in Zusammenhang stehenden Begriffen auseinander.

### **2.2.1. Die Differenzierung von Qualitätsmanagement, Qualitätsmanagementsystem und Qualitätssicherung**

Ursprünglich hatte sich im deutschen Sprachraum der Begriff „Qualitätssicherung“ durchgesetzt. Nach der früher gebräuchlichen Norm DIN 55 350, Teil 18 (1987) wurde unter Qualitätssicherung die „Gesamtheit aller Tätigkeiten des Qualitätsmanagements, der Qualitätsplanung, der Qualitätslenkung und der Qualitätsprüfung“ verstanden. Folglich wurde Qualitätsmanagement (QM) als ein Unterbegriff der Qualitätssicherung verstanden und als derjenige Aspekt der Gesamtführungsaufgabe definiert, welcher die „Qualitätspolitik festlegt und diese zur Ausführung“ bringt. Die mit der Aufstellung der ISO Normen einhergehende Benennungsänderung führte dazu, dass sich Anfang 1992 die deutschsprachige Begriffsdefinition der internationalen Entwicklung anschloss, indem sie den Begriff Qualitätssicherung durch den Begriff des Qualitätsmanagements

ersetzte (Kamiske & Brauer, 1995). Die Norm DIN EN ISO 8402 (1995) definiert **Qualitätsmanagement** als „*Gesamtheit aller Tätigkeiten, welche die Qualitätspolitik, Ziele und Verantwortungen festlegen sowie diese durch Mittel wie Qualitätsplanung, Qualitätslenkung, Qualitätssicherung und Qualitätsverbesserung im Rahmen des Qualitätsmanagementsystems verwirklichen*“. Im Gegensatz zur vorherigen Definition beschränkt sich der Aufgabenbereich des Qualitätsmanagement nicht nur auf Tätigkeiten der Führungsebene, sondern liegt in der Verantwortung aller Hierarchieebenen einer Organisation. Ein **Qualitätsmanagementsystem** beschreibt nach DIN EN ISO 8402 die „*zur Verwirklichung des Qualitätsmanagements erforderliche Organisationsstruktur, Verfahren, Prozesse und Mittel*“ und kann als Werkzeug zur Realisierung der Qualitätspolitik verstanden werden. Der Begriff der Qualitätspolitik und dessen Umsetzung werden ausführlich in Abschnitt 2.2.2.

**Qualitätssicherung** ist nur noch ein Teilaspekt des Qualitätsmanagements und umfasst nach DIN EN ISO 8402 (1995) „*alle geplanten und systematischen Tätigkeiten, die innerhalb des Qualitätsmanagementsystems verwirklicht sind, und die wie erforderlich dargelegt werden, um angemessenes Vertrauen zu schaffen, dass eine Einheit die Qualitätsforderung erfüllen wird*“. Darüber hinaus weist die Norm darauf hin, dass es sowohl interne als auch externe Gründe für die Qualitätssicherung gibt. Während die interne Funktion der Qualitätssicherung darin besteht, innerhalb einer Organisation Vertrauen in die das Management zu verschaffen, bezweckt externe Qualitätssicherung Vertrauen beim Kunden oder anderen zu schaffen, dass die Qualitätsanforderungen erfüllt werden. Qualitätssicherung kann in Form von Qualitätsnachweisen z.B. mittels Qualitätsbericht oder Zertifizierung erfolgen.

Zusammengefasst versteht man unter Qualitätsmanagement die aufeinander abgestimmten Tätigkeiten zum Lenken und Leiten einer Organisation, die darauf abzielen, die Qualität der Produkte oder der angebotenen Dienstleistungen zu verbessern. Qualität wird dabei im Konzept des Qualitätsmanagements ganzheitlich verstanden. Die menschlichen Interaktionen sollen wie die technischen Abläufe und gefertigten Produkte unter Qualitätsgesichtspunkten optimiert werden.

### 2.2.2. Die Umsetzung von Qualitätspolitik

Ein wesentlicher Bestandteil der Unternehmenspolitik ist die **Qualitätspolitik**. Nach EN ISO 8402 (1995) beinhaltet sie die „umfassende Absichten und Zielsetzungen einer Organisation zur Qualität“. Das Instrument zur Umsetzung der Qualitätspolitik ist das Qualitätsmanagement mit den Elementen Qualitätsplanung, Qualitätslenkung, Qualitätssicherung und Qualitätsverbesserung (Kamiske & Brauer, 1995). Auf die Elemente Qualitätsplanung, -lenkung und -verbesserung wird in den folgenden Ausführungen nun näher eingegangen.

Die Verantwortung für die Qualitätspolitik eines Unternehmens obliegt der obersten Führungsebene, welche die Qualitätspolitik in Hinblick auf „das Image der Dienstleistungsorganisation und ihren Ruf bezüglich Qualität, die Ziele in Bezug auf die Dienstleistungsqualität, die zur Verfolgung von Qualitätszielen einzuschlagende Vorgehensweise und die Rolle des für die Verwirklichung der Qualitätspolitik verantwortlichen Firmenpersonals“ entwickeln und dokumentieren soll. Die stufenweise Operationalisierung der Qualitätspolitik kann mittels des so genannten Top-down-Ansatzes erfolgen, der die Aufschlüsselung der primären Qualitätspolitik in Qualitätsziele sowie deren Konkretisierung von oberer Ebene zu unterer Ebene bezeichnet (Reinhardt et al., 1996). Auf oberster Führungsebene des Unternehmens werden die Grundsätze der Qualitätspolitik und die Qualitätsziele formuliert. Auf dieser Basis werden auf mittlerer Führungsebene Teilziele ermittelt und Projekte zu deren Umsetzung erarbeitet. Auf der untersten Ebene schließlich leiten sich daraus konkrete Maßnahmen und Einzelziele für die Mitarbeiter ab. Entscheidend für die erfolgreiche Umsetzung der Qualitätspolitik und damit auch des Qualitätsmanagements ist deren Akzeptanz durch alle Mitarbeiter auf allen Hierarchieebenen.

Die **Qualitätsplanung** umfasst alle qualitätsbezogenen Planungsaktivitäten vor Beginn der Leistungserbringung. Ihre Aufgabe ist es, die durch die Qualitätspolitik vorgegebenen Ziele durch Planung konkreter produkt- und prozessbezogener Maßnahmen zu erreichen. Dazu bedarf es zunächst der Identifizierung der Qualitätsmerkmale, wodurch die Anforderungen an ein Produkt bzw. eine Dienstleistung, an den Produktionsprozess sowie die erforderliche Qualitätssicherungsmaßnahmen festgelegt werden können. Dies geschieht unter



Berücksichtigung der Kundenanforderungen, deren Realisierbarkeit und den gegebenen materiellen, personellen und finanziellen Ressourcen des Unternehmens. Nach Bestimmung der produkt- bzw. dienstleistungsbezogenen Qualitätsmerkmale erfolgt die Planung der prozessbezogenen Maßnahmen. Dies schließt die Planung der qualitätsbezogenen Führungstätigkeiten, insbesondere die Anwendung des Qualitätsmanagementsystems sowie Ablauf- und Zeitpläne mit ein. Zusammenfassend kann die Qualitätsplanung als die Festlegung qualitätskonformer Produkt- bzw. Dienstleistungs- und Realisierungsspezifikationen bezeichnet werden. In der Qualitätsplanung werden Vorgaben bezüglich der zu erreichenden Ergebnisqualität und der dazu erforderlichen Struktur- und Prozessqualität erarbeitet. Die in Abschnitt 2.1.1.2. beschriebene Wechselbeziehung der drei Qualitätsdimensionen Struktur-, Prozess- und Ergebnisqualität findet sich in diesen Ausführungen wieder. Dabei beginnt die Verbesserung von Qualität bereits auf der Ebene der Qualitätsplanung. Es sollen alle auf die Qualität Einfluss nehmenden Faktoren in ihrem Zusammenspiel in Richtung optimaler Qualität gestaltet werden. Diese Wechselwirkung von produkt- und prozessbezogenen Maßnahmen ermöglicht die Schaffung eines Rahmens, innerhalb dessen eine kontinuierliche Qualitätsverbesserung der Planungsprozesse möglich wird (Reinhard et al., 1996).

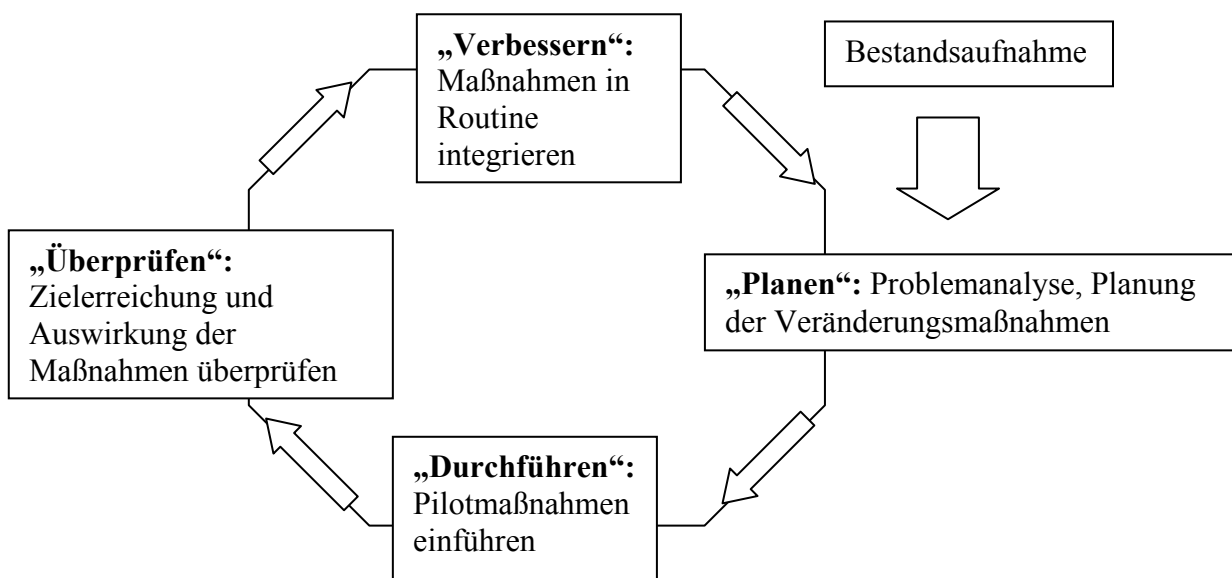
Die **Qualitätslenkung** basiert auf den Forderungen der Qualitätsplanung. Sie umfasst nach ISO 8402 (1995) die „Arbeitstechniken und Tätigkeiten, die zur Erfüllung von Qualitätsforderungen (...) angewendet werden“. Qualitätsforderungen bezeichnen hier jene spezifischen Qualitätsmerkmale, die eine Prüfung des Produktes hinsichtlich seiner Qualität ermöglicht. Die Qualitätsprüfung wird durch Expertengruppen vorgenommen, welche die Daten aus der Qualitätsprüfung mit den Vorgaben der Qualitätsforderung vergleichen. Dabei liegt es im Ermessen der Gruppe, sich konform mit den Ergebnissen der Qualitätsprüfungen zu erklären oder auf Nachbesserungen zu bestehen. Die Maßnahmen der Qualitätslenkung beziehen sich aber nicht nur auf das Produkt im Sinne einer Qualitätsprüfung, sondern auch auf den Herstellungsprozess oder das zur Fertigung eingesetzte Personal wie z.B. Mitarbeiterschulungen (Hannen & Lindemann, 1996). Qualitätslenkung hat somit eine vorbeugende, überwachende und korrigierende Funktion.

Unter **Qualitätsverbesserung** versteht man das Ergreifen von Maßnahmen zur Erhöhung der Effektivität und Effizienz von Prozessen zur Erzielung von Nutzen sowohl für die Kunden als auch die Organisation (Hannen & Lindemann, 1996). Für Unternehmen bedeutet dies, dass alle ergriffenen Maßnahmen auf die kontinuierliche Weiterentwicklung und Verbesserung der Leistungsqualität ausgerichtet sind. Um Qualität im Sinne von Qualitätsmanagement verbessern zu können, ist es einerseits erforderlich, die unterschiedlichen Anforderungen genau zu definieren (Qualitätsplanung) und andererseits die tatsächlich erbrachte Qualität auf jeder Stufe des Prozesses möglichst exakt zu messen (Qualitätsprüfung im Rahmen der Qualitätslenkung). Der Prüfung von Qualität können dabei drei wichtige Funktionen zugesprochen werden: Nachweis des erreichten Qualitätsniveaus, Erkennen von Verbesserungsbedarf und Bewertung der Effektivität von Verbesserungsmaßnahmen. Dazu wiederum muss ein geeignetes Bewertungs- und Prüfverfahren eingeführt werden. Prüfungen stellen einen dynamischen und interaktiven Prozess dar und werden häufig mit einem kybernetischen Regelkreis verglichen. Klassische Regelkreise arbeiten immer mit einer Rückkopplung und orientieren sich am Output. Der Regler vergleicht Ist- und Soll-Zustand und nimmt über die Regelgröße die notwendigen Korrekturen vor. Prüfverfahren nutzen in der Regel dieses allgemeine Grundkonzept. Drei wesentliche Beispiele für Prüfverfahren aus der wirtschafts- und sozialwissenschaftlicher Literatur sind die Revision (englisch: Auditing), das Controlling und die Evaluation (Schenker-Wicki, 1996). Der Problemstellung der Arbeit folgend wird in Kapitel 3 Evaluationen als Prüfverfahren im Hochschulbereich ausführlich erörtert.

### **2.2.3. Der Kreislaufcharakter des Qualitätsmanagements**

Bei Betrachtung der einzelnen Elemente des Qualitätsmanagements wird deutlich, dass auch sie einem Regelkreis entsprechen. Als einzelne Abschnitte dieses Kreislaufs können die Qualitätsplanung, die Qualitätslenkung mit der Qualitätskontrolle und die Qualitätsverbesserung betrachtet werden. Das Ziel der Kontrolle ist die Ermittlung des Ist-Zustandes und dessen Abweichung vom Soll-Zustand, welcher in der Qualitätspolitik und den Qualitätszielen festgelegt ist. Als Konsequenz dieser Bestandsaufnahme folgt die Analyse der Ursache für etwaige Abweichungen und die Erarbeitung von Korrekturen zur Minimierung der Ist-Soll-Abweichung i. S. einer Qualitätsverbesserung.

In Abschnitt 2.1.1.1. wurde Qualität in dem kontinuierlichen Prozess der Qualitätsverbesserung als eine dynamische Größe beschrieben, die mit den sich ändernden Kundenbedürfnissen und Anforderungen an Produkte und Dienstleistungen stetig neu zu erheben ist. Folglich muss auch der zuvor dargestellte Kreislauf von Planung, Lenkung und Kontrolle kontinuierlich durchlaufen werden, um Qualität zu leisten und stetig zu verbessern. Für das Prinzip der ständigen Verbesserung entwickelte Edward Deming (1986) den sog. Deming-Zyklus bzw. Deming-Kreis, der zugleich Anwendungs- und Erklärungsmodell ist. Deming verzichtet auf eine allumfassende Definition von Qualität. Er löst den abstrakten Begriff „Qualität“ in unterschiedliche Profile, Merkmale, Skalen und Konsumentengruppen auf. Damit macht er deutlich, dass die Qualitätsanforderungen an ein Produkt immer wieder neu von den unterschiedlichen Standpunkten aus zu erörtern sind. Demings Philosophie ist auf ständige Verbesserung der Qualität und Optimierung von Produktionsprozessen ausgerichtet, wobei er betont, dass jeder Mitarbeiter des Unternehmens darin einbezogen werden müsse. Der Deming-Zyklus wird auch als **Plan-Do-Check-Act-Zyklus** (PDCA – Zyklus: Planen – Durchführen – Überprüfen – Verbessern, vgl. Abb. 1) bezeichnet. Er systematisiert die ständige Verbesserung und repräsentiert die Aufgaben des QM.



**Abbildung 1: Deming-Zyklus**, entnommen und verändert aus: Oess, A. (1993). *Total quality management: Die ganzheitliche Qualitätsstrategie*. (3.Auflage) Wiesbaden. Gabler Verlag. S.94

Der aus vier Stufen bestehende Kreislauf von Aktivitäten zur Qualitätsverbesserung beginnt mit der Untersuchung der gegenwärtigen Situation. Daten werden gesammelt

und ein Plan zur Verbesserung formuliert. Im Sinne des „**to plan**“ (Qualitätsplanung) legt die Führung die Ziele fest und plant Maßnahmen zur Umsetzung der Ziele mit den betroffenen Kräften und schafft Verantwortung und Handlungsspielräume. Danach folgt die Umsetzung des Plans im Sinne des „**to do**“, zunächst in kleinerem Maßstab. Anschließend wird überprüft, ob die gewünschte Verbesserung eingetreten ist. Die im Rahmen des „**to check**“ (Qualitätsprüfung im Rahmen der Qualitätslenkung) gewonnenen Erkenntnisse führen dann zur Änderung von Strategie und Aktivitäten im Sinne des „**to act**“ (Qualitätsverbesserung). Bei wiederholtem Durchlaufen des Zyklus werden die Ursachen für nicht zufrieden stellende Leistung immer weiter eingegrenzt, der Wissensinhalt nimmt zu und die Erfahrung aus den vorhergehenden Zyklen können angewendet werden. PDCA kann als ein Prozess verstanden werden, durch den kontinuierlich bereits etablierte Standards in Frage gestellt werden, um diese durch bessere zu ersetzen (Oess, 1993). Mit gleicher inhaltlicher Bedeutung wird „ständige Verbesserung“ im angloamerikanischen Sprachraum als Continuous Improvement Process (CIP) und im Japanischen als Kaizen bezeichnet.

### **2.3. Umsetzung von Qualitätsmanagement an Hochschulen**

Auch wenn Hochschulen nicht wie Unternehmen auf Gewinnerzielung hin orientiert sind, müssen sie sich in mancher Hinsicht wie Unternehmen verhalten lernen. In vielen europäischen Ländern wurden die starren verwaltungstechnischen Vorschriften im Hochschulbereich in den letzten Jahren gelockert und Strukturen eines leistungsorientierten Hochschulmanagements eingeführt (Schenker-Wicki, 1996). Den neuen Anforderungen der Profilbildung und der Wettbewerbsfähigkeit ist ohne Erweiterung der institutionellen Handlungsspielräume der Hochschulen in akademischer, finanzieller, personeller und organisatorischer Hinsicht nicht gerecht zu werden. Im Rahmen der Diskussion über die Reform des Hochschulsystems haben Bund, Länder und Hochschulen einen breiten Konsens darüber erzielt, dass die notwendigen Voraussetzung hierfür in Deregulierung, Leistungsorientierung und der Schaffung von Leistungsanreizen bestehen können (Friedrich, 2001). Nachdem die Begriffe Qualität und Qualitätsmanagement in den vorangegangenen Kapiteln erläutert wurden und bereits eine Übertragung des Qualitätsbegriffes auf die Ebene der Hochschule stattgefunden hat, so gilt es nun zu klären, wie QM zur Erfüllung des Aufgabenfeldes der Hochschule und zur Verbesserung der Qualität in der Lehre

angewendet werden kann und welche Rahmenbedingungen für die Umsetzung von QM erforderlich sind.

### 2.3.1. Rahmenbedingungen für QM an Hochschulen

Die Rahmenbedingungen für Qualitätsmanagement an der Hochschule können in externe und interne Rahmenbedingungen differenziert werden. Hochschulexterne Rahmenbedingungen ergeben sich durch gesetzlichen Bestimmungen und Auflagen, durch die zur Verfügung stehenden Haushaltsmittel und der bereits beschriebenen Wettbewerbssituation der Hochschulen (Moewes, 2002). Die Gestaltung der externen Rahmenbedingungen durch den Staat kann anhand der zwei bestehenden Modelle einer externen Hochschulsteuerung veranschaulicht werden. Auf der einen Seite existiert das Modell der staatlichen Aufsicht, in dem das gesamte Hochschulsystem einschließlich der Zulassungsbestimmungen und Prüfungssysteme einer staatlichen Kontrolle unterliegt. Dem gegenüber steht das Modell der staatlichen Aufsicht. Hier gibt der Staat die allgemeinen Rahmenbedingungen vor, deren Umsetzung den Hochschulen selbst obliegt (Schenker-Wicki, 1996) In Deutschland wurde traditionellerweise das Modell der staatlichen Kontrolle praktiziert. Qualitätssicherung erfolgte in erster Linie durch Vorgaben für und Genehmigungen von Prüfungsordnungen durch den Staat im Sinne einer *ex ante-Steuerung*. Aber nur Institutionen, die sich rasch auf neue Entwicklungen in Wissenschaft, Forschung und Arbeitsmarkt einstellen und reagieren können, haben eine Chance, im nationalen und internationalen Wettbewerb zu bestehen. Diese Flexibilität ist im Modell der staatlichen Kontrolle eingeschränkt. Internationale Erfahrungen zeigen, dass sich erfolgreiche Instrumente für die Sicherung und die Verbesserung der Qualität der Lehre und des Studiums zunehmend an deren Ergebnissen orientieren im Sinne einer *ex post-Steuerung*. Um eine *ex post-Steuerung* erfolgreich einzusetzen sind nach Langfried (2001) folgende Voraussetzungen zu erfüllen: Handlungsfreiheit (Autonomie), Handlungsfähigkeit (modernes Hochschulmanagement) sowie die Transparenz des Systems und seiner Leistungen. Mehr Handlungsfähigkeit durch Finanzautonomie und mehr Eigenverantwortung in personellen und organisatorischen Fragen bedeuten mehr Selbständigkeit bei Sach-, Personal- und Haushaltsangelegenheiten. Erreicht werden kann dies durch mehrjährige Globalhaushalte und leistungs- und belastungsorientierte Mittelvergabe für alle Aufgabenbereiche. Der damit verbundene Autonomiegewinn der Hochschulen

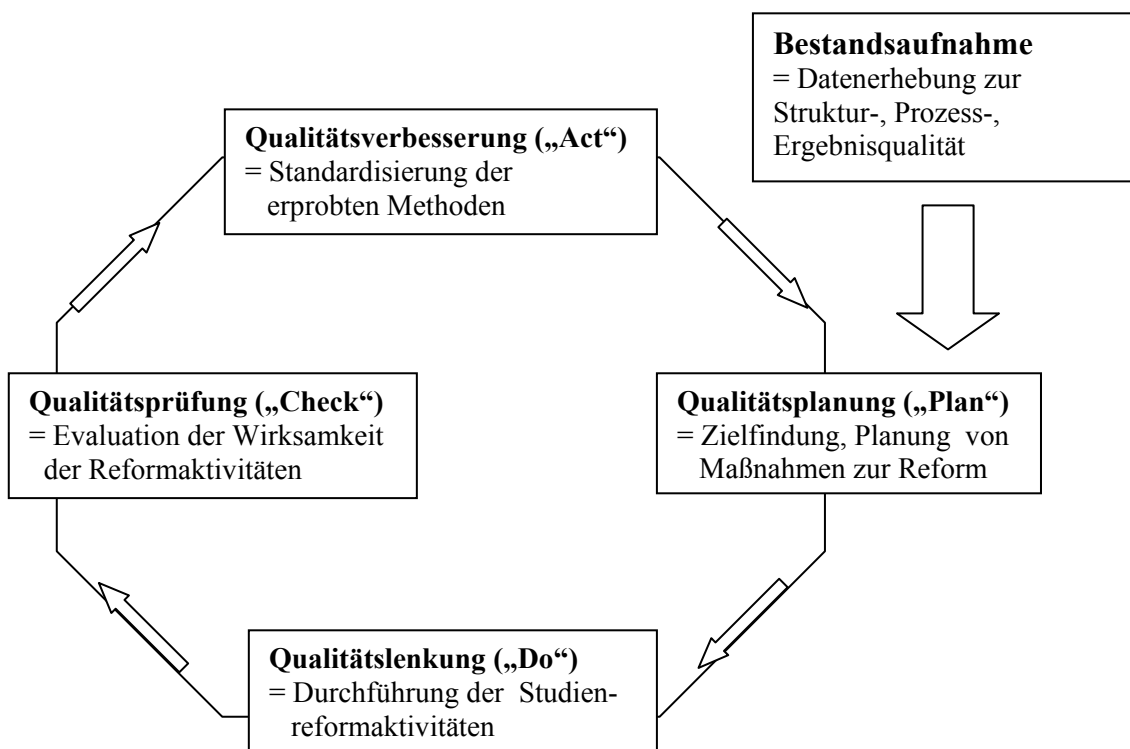
verpflichtet diese zur Transparenz von Kosten und Leistungen und Rechenschaftslegung gegenüber Staat und Öffentlichkeit. Die Rolle des Staates beschränke sich danach auf Globalsteuerung und Rechtsaufsicht.

Aus den externen Vorgaben leiten sich die internen Rahmenbedingungen ab und werden durch die Hochschule selbst entwickelt. Für die Umsetzung der staatlichen Vorgaben, der im Rahmen der Qualitätspolitik selbst gesetzten Ziele, einer kontinuierlichen Qualitätskontrolle, regelmäßiger Leistungsnachweise und der Einbeziehung von Markterfordernissen in die Hochschulpolitik bedarf es eines modernen Hochschulmanagements mit entsprechenden Strukturen und personellen Kompetenzen. Ähnlich der privatwirtschaftlichen Unternehmensführung setzt ein modernes Hochschulmanagement die wesentlichen Führungsinstrumente dabei wie Finanzierung, Qualitätsplanung, Qualitätslenkung und Qualitätssicherung ein (Schenker-Wicki, 1996).

### **2.3.2. Qualitätsverbesserung in der Lehre durch PDCA**

Die Notwendigkeit von intensiven Bemühungen um eine Qualitätsverbesserung in der Lehre hat in den letzten Jahren ebenso zugenommen wie die Bereitschaft der Hochschulen, Fachbereiche und Studiengänge sich dieser Herausforderung zu stellen (Voegelin, 2001). Um die Qualität in der Lehre zu beurteilen und zu verbessern, muss zunächst in gemeinsamer Kommunikation über Anforderungen und Qualitätsmerkmale erarbeitet werden, was Qualität in der Lehre ist. Die Einführung eines Qualitätsmanagement mit den Unterfunktionen Qualitätssicherung und –verbesserung erfordert somit den Aufbau eines Kommunikationssystems, das in der Lage ist, diesen Konsens über Qualität hervorzubringen (Lüthje, 2000). Diese Konsensbildung hat im Hochschulbereich zwischen drei Gruppen stattzufinden. Als erste Gruppe sind die Lehrenden und die Konsensbildung untereinander zu nennen, denn an Lehre als komplexer Prozess nehmen unterschiedliche Lehrende mit individuellen Ziel- und Qualitätsvorstellungen teil. Auch zwischen den Lehrenden und Lernenden als zweite und zwischen der Hochschule und ihrer Umwelt als dritte Gruppe ist i. S. einer Kundenorientierung ein Konsens in Bezug auf die Qualität zu finden. Der Konsens führt als Konsequenz im Rahmen des Qualitätsmanagements zur Entwicklung von Leitbildern und Zielen im Sinne einer Qualitätspolitik der Hochschule, zur Erarbeitung von Strategien zu deren Umsetzung und zur Einführung von Prüfverfahren, welche die

Wirkung der ergriffenen Maßnahmen zur Qualitätsverbesserung betrachten. In der Praxis findet üblicherweise der Kreislaufprozess im Sinne des Deming-Zyklus mit den vier Phasen der Planung, Durchführung, Prüfung und Qualitätsverbesserung Anwendung (Ritz, 2003). Arnold (2003) leitet aus diesen Phasen vier Aufgabengebiete ab: Bestandsaufnahme, Festlegung von Entwicklungszielen, Maßnahmen zur Qualitätssteigerung und Überprüfung der Wirksamkeit von Reformaktivitäten. Eine zusammenfassende Darstellung der im Folgenden erläuterten Schritte findet sich in Abbildung 2.



**Abbildung 2: PDCA zur ständigen Verbesserung in der Lehre**

Qualitätsverbesserung in der Lehre beginnt mit einer Bestandsaufnahme der aktuellen Situation. In der Hochschullehre kann sich diese in Anlehnung an Donabedians Qualitätsdimensionen auf strukturelle Bedingungen, Prozesse und Ergebnisse beziehen. Somit bedarf es der Erhebung von Struktur-, Prozess- und Ergebnisdaten. Zu den allgemeinen Strukturdaten zählen Angaben zur Personalstruktur, Zahl der Studienplätze, Studierenden und Absolventen, Ausstattung, Angaben zu den finanziellen Ressourcen als auch gesetzliche Vorgaben und Verordnungen, welche die externen Rahmenbedingungen aufspannen. Zur Erhebung von Prozessdaten wird vor allem die

Bewertung von Lehrveranstaltungen durch Studierende, die so genannte Lehrveranstaltungsevaluation, herangezogen. Beurteilt werden können neben der Lehrpräsentation die Qualität der Beratung und Betreuung der Studierenden. Schließlich ist es häufig erforderlich auch Ergebnisdaten in die Bestandsaufnahme einzubeziehen. Informationen liefert hierzu wieder die Hochschulstatistik mit Angaben zur Studiendauer, den Prüfungs- und Abschlussnoten. Darüber hinaus können durch Absolventenbefragungen Rückmeldung zu Studienerfolg, erworbenem Wissen und erworbener Kompetenzen erhalten werden. Daraus wird ersichtlich, dass eine allumfassende Bestandsaufnahme mit erheblichem zeitlichen, personellen und finanziellen Aufwand verbunden sein kann (Arnold, 2003).

Die zu Beginn des Zyklus stehende Bestandsaufnahme ist zugleich eine Bedarfsanalyse, welche nach Rossi et al. (1999) eine Form der Evaluation darstellt. Die Thematik der Evaluation der Lehre wird in Kapitel 3 ausführlich dargestellt. Da Bestandsaufnahmen zur Vorbereitung von konkreten Verbesserungsvorhaben dienen, können sie sich auch auf bestimmte Aspekte beschränken. Besteht das Vorhaben Struktur- und Prozessqualität zu verbessern, so können gezielt Daten hierzu erhoben werden. Der Analyse der Ausgangssituation schließt sich die Festlegung von Entwicklungszielen und Planung von Aktivitäten zur Umsetzung dieser Ziele an. Sie entspricht im Modell des Qualitätsmanagements der Festlegung einer Qualitätspolitik auf Führungsebene mit den daraus abgeleiteten Qualitätszielen auf jeder Ebene des Unternehmens sowie der Qualitätsplanung. Im Unterschied zu Unternehmen ist es an Universitäten üblich, dass alle Statusgruppen an der Vereinbarung und Umsetzung von Entwicklungszielen beteiligt sind (Kern, 2000). So gibt es Entscheidungsgremien (z.B. Instituts- und Fachbereichsrat), in denen Repräsentanten aus Reihen der Professoren und Professorinnen, Studierenden und des Verwaltungspersonals sind. Diese Gremien haben in Bezug auf die Lehre klar definierte Aufgaben und Entscheidungsbefugnisse und sind gemeinsam mit dem Dekan für die Qualitätspolitik und Qualitätsziele in Studium und Lehre des entsprechenden Fachbereiches verantwortlich. Somit sind bei Festlegung der Qualitätspolitik und Erarbeitung von Reformaktivitäten alle Mitarbeiter integriert bzw. repräsentiert, während in Unternehmen eine kleine Gruppe von Personen, das Management, die Qualitätspolitik festlegt. Die aus der Bestandsaufnahme mit Stärken- und Schwächenanalyse abgeleiteten Verbesserungsmaßnahmen werden anschließend umgesetzt. Solche Maßnahmen müssen nicht sofort das gesamte System Hochschule



ergreifen, sondern können in Form von Reformprojekten zu einzelnen Themen, z.B. Einführung neuer Betreuungsformen für Praktika, eine Art Testphase durchlaufen. Am Ende dieser Phase steht die Evaluation der Wirksamkeit der eingesetzten Reformmaßnahme. Zusammen mit den Struktur- und Prozessdaten der allgemeinen Bestandsaufnahme vervollständigen die Daten der Evaluation der Reformaktivitäten im Sinne von Ergebnisdaten das Qualitätsbild der Hochschule. Aus diesem können wiederum neue Qualitätsverbesserungsmaßnahmen erarbeitet werden und der Kreislauf kann erneut durchlaufen werden. Zeigen die Ergebnisse der Evaluation die gewünschte Verbesserung, so kann die getestete Methode standardisiert werden. Im oben genannten Beispiel bedeutet dies, dass die neu gewählte Form der Praktikumbetreuung für alle Praktika des Fachbereichs umgesetzt werden kann.

#### **2.4. Fazit**

Die deutschen Hochschulen sehen sich vor dem Hintergrund des Wettbewerbs zwischen nationalen und internationalen Bildungsinstitutionen um Studierende, Mitarbeiter und knappe Finanzmittel sowie dem wachsenden Legitimationsdruck gegenüber Staat und Öffentlichkeit neuen Anforderungen gegenübergestellt. Die Qualität der Lehre, deren Sicherung und Verbesserung im Rahmen eines Qualitätsmanagement rücken dabei in den Mittelpunkt der Diskussion. Um allen Herausforderungen und Ansprüchen gerecht zu werden, ist es notwendig, Qualitätsmanagement an deutschen Hochschulen einzuführen. Es ist ein Managementkonzept, bei dem mittels spezifischer Instrumente und Maßnahmen die internen Strukturen und Prozesse in allen Funktionsbereichen und Ebenen darauf ausgerichtet werden, die Qualität von Dienstleistungen zu optimieren. Ziel ist die stetige Qualitätsverbesserung, um eine herausragende Stellung im Bildungsmarkt zu erhalten und sich somit im Wettbewerb erfolgreich zu behaupten. Es ist nicht allein die Wettbewerbssituation, die die Hochschulen zur Verbesserung der Qualität zwingt, sondern auch die Verpflichtung gegenüber den Studierenden und der Gesellschaft, den von ihr erteilten Bildungsauftrag qualitativ hochwertig zu erfüllen. Um Qualitätsmanagement erfolgreich an Hochschulen einzuführen, bedarf es wird neben der Veränderung von Organisationsstrukturen mit mehr Autonomie für die Hochschule und der Dezentralisierung von Entscheidungen und Verantwortungen. Zudem erforderlich ist ein verbessertes Hochschulmanagement und die Flexibilisierung des Hochschulhaushaltes mit einem Übergang von staatlicher ex ante-Steuerung in eine

ex post-Steuerung. Für Hochschulen gilt es in dem System der ex post-Steuerung ein Verfahren zur Qualitätssicherung einzusetzen, das die notwendige Transparenz von Kosten und Leistungen schaffen kann. Als ein mögliches Instrument der Qualitätssicherung und Qualitätsverbesserung hat sich die Evaluation als Prüfverfahren an Hochschulen etabliert. Im folgenden Kapitel soll daher die Evaluation und ihr Beitrag zur Qualitätsverbesserung in der Hochschule ausführlich dargestellt werden.

### **3. Evaluation als Beitrag zur Qualitätsverbesserung**

Die Einführung eines Qualitätsmanagements an Hochschulen dient der Steigerung des Qualitätsbewusstseins aller Mitwirkenden und trägt zur stetigen Qualitätsverbesserung bei. Um Qualität an Hochschulen dauerhaft zu sichern und zu verbessern kann der im Abschnitt 2.3.2. dargestellte Deming-Kreis aus den vier Handlungen Planung, Durchführung, Prüfung und Qualitätsverbesserung durchlaufen werden. Eine besondere Rolle in diesem Modell kommt der Evaluation als Prüfverfahren zu. Sie ist auf der einen Seite in Form einer Bedarfsanalyse Informationsgrundlage der Qualitätsdiskussion und Qualitätsplanung. Auf der anderen Seite steht sie als Kontrollinstrument im Sinne einer Wirkungsevaluation am Ende eines jeden PDCA-Zyklus. Im folgenden Kapitel wird der Evaluationsbegriff erläutert und die Arten, Funktion und Ziele von Evaluationen im Allgemeinen sowie speziell im Bereich der Hochschullehre werden erörtert. Vor diesem Hintergrund werden die möglichen Probleme bei einer rein studentischen Lehr-evaluation aufgezeigt. Abschließend werden nationale und internationale Konzepte der Lehrevaluation und der Qualitätsverbesserung vorgestellt.

#### **3.1. Der Evaluationsbegriff**

In der allgemein-öffentlichen sowie hochschulpolitischen Diskussion ist „Evaluation“ zu einem vielfältig schillernden Modewort geworden, mit dem je nach Kontext sehr Unterschiedliches verbunden wird (Kromrey, 2001). Gegenstände von Evaluationen, auf die näher in Abschnitt 3.2. eingegangen wird, können Personen, Organisationen, Reformen und Gesetze oder die Evaluation selbst sein (Stockmann, 2000). Dementsprechend findet sich eine Vielfalt an unterschiedlichen Definitionen. Sie reichen von der lateinischen Bedeutung des Wortes „Evaluation“ als „Bewertung, Beurteilung“ bis hin zu komplexen, individuellen Definitionen von „Evaluation“ in den einzelnen Sachgebieten.

Als erste Annäherung wird der Evaluationsbegriff nach Donna Mertens (1998) genutzt, welcher sowohl auf die verschiedenen Evaluationsgegenstände zutrifft als auch bereits auf die besondere Aufgabe und die Ziele der Evaluation eingeht:

*„Evaluation is the systematic investigation of the merit or worth of an object (programm) for the purpose of reducing uncertainty in decision making“.*

Rossi et al. (1988) erweitern diese Definition um die angewandten Methoden aus der Sozialwissenschaft und prägen den Begriff der Evaluationsforschung „als systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluationsforschung bezeichnet den gezielten Einsatz sozialwissenschaftlicher Forschungsmethoden zur Verbesserung der Planung und laufenden Überwachung sowie zur Bestimmung der Effektivität und Effizienz von (...) sozialen Interventionsmaßnahmen“. Aus dieser Definition ergeben sich bereits drei unterschiedliche Schwerpunkte bzw. Analyseperspektiven von Evaluationen bzw. Evaluationsforschung: Programmentwicklungsphase, Implementationsphase (Begleitforschung) und nach Implementation (Wirksamkeitsforschung). Es werden Informationen über die Implementierung und die Auswirkungen von Maßnahmen gesammelt und ausgewertet. Die eigentliche Evaluation besteht dabei nicht in der Formulierung normativer Urteile, sondern in Aussagen darüber, ob das in Erprobung befindliche Modell so implementiert wurde wie geplant, ob die gesetzten Ziele dadurch erreicht wurden, welche Maßnahmen besonders effektiv waren und welche erwünschten oder auch unerwünschten Wirkungen auftraten (Kromrey, 2001).

Zusammenfassend ist festzuhalten, dass Evaluationen die systematische Untersuchung zum Nutzen oder Güte eines Gegenstandes oder Programms unter Anwendung von sozialwissenschaftlichen Forschungsmethoden im Sinne einer Evaluationsforschung darstellen.

### **3.2. Arten, Funktion und Ziele von Evaluationen**

Es finden sich in der Literatur für den Begriff der Evaluation mehrere Ausdifferenzierungen nach Gegenstand der Evaluation, Untersuchungsaspekte der Evaluation, Zeitpunkt, Steuerung und Durchführung sowie deren Umfang und Intensität (Ritz, 2003). Diese Ausdifferenzierungen werden im Folgenden erläutert und dabei die Begriffe Bedarfsanalyse, Prozessevaluation, Wirkungsevaluation, formativ und

summative Evaluation, Selbst- und Fremdevaluation sowie interne und externe Evaluation, Kurz- und Langevaluation eingeführt.

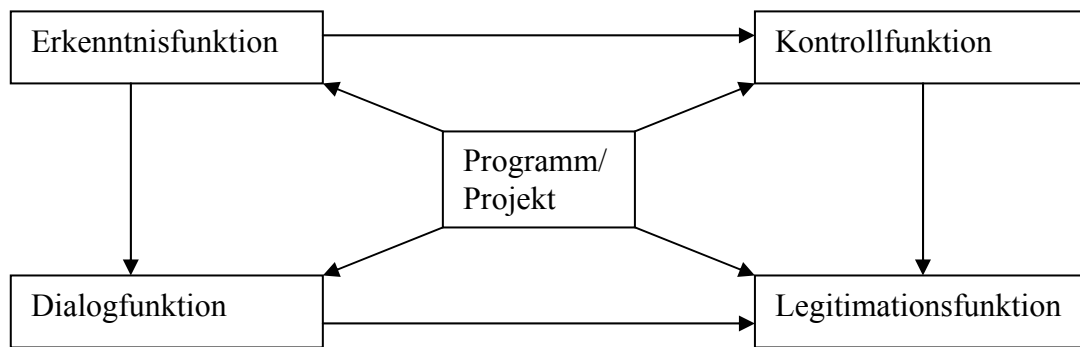
Gegenstand von Evaluationen können z.B. Programme, Produkte, Institutionen und die Evaluation selbst sein. Letzteres wird auch als Metaevaluation bezeichnet. Beginnend mit der Programmevaluation ist zunächst zu definieren, was ein Programm im Allgemeinen darstellt. Ein Programm ist ein Maßnahmenbündel, das gezielt Veränderungen hervorrufen soll (Ritz, 2003), im Hochschulbereich z.B. die Verbesserung der Lehr- und Lernbedingungen. Programmevaluationen beinhalten somit die Untersuchung von Maßnahmen zur Umsetzung von z.B. Qualitätszielen. Produktevaluationen dienen der Bewertung und Verbesserung konkreter Waren oder Dienstleistungen hinsichtlich ihrer Qualität. Institutionsevaluationen widmen sich der Bewertung von Zielen, Verfahren, Organisation, Kultur und davon abhängiger Aufgabenerfüllung einer Institution. Metaevaluationen erfüllen den Zweck, Erkenntnisse über die Qualität von Evaluation zu gewinnen und durch Vergleich mit anderen Evaluationsverfahren eine Verbesserung der Evaluationspraxis zu erreichen.

Bezüglich des Untersuchungsaspektes und Zeitpunktes kann eine Einteilung der Evaluationen in die Phasen der Programmentwicklung, der Implementationsphase und der Wirkungsphase nach Implementation vorgenommen werden. Die sich nun ergebenden Evaluationstypen sind nach Rossi et al. (1999) die Bedarfsanalyse, die Prozessevaluation und die Wirkungsevaluation. Die **Bedarfsanalyse** dient der Beantwortung der Frage nach der Notwendigkeit und Effektivität einer Maßnahme. Die Evaluation hat dabei die Aufgabe, die materiellen, finanziellen, institutionellen, personellen Rahmenbedingungen eines Programms zu untersuchen und damit einen Beitrag zur Konzeption des Programms selbst in der Planungsphase zu leisten (Stockmann, 2000). Gleichzeitig sollen möglichst früh mögliche negative Effekte des Programms oder einer Maßnahme abgeschätzt werden. Da solche Evaluationen vor der Implementation eines Programms durchgeführt werden, werden sie als *ex ante-Evaluationen* bezeichnet. Die **Prozessevaluation** beinhaltet die Analyse der Programmumsetzung in Bezug auf Funktionalität, Prozesse und Leistungserbringung. Sie geschieht *formativ begleitend* (Begleitforschung), d.h. aktiv gestaltend und prozessorientiert, so dass bereits noch während der Umsetzung Schwachstellen erkannt und Verbesserungen bewirkt werden können. Die **Wirkungsevaluation** ist eine *ex post-*

*Evaluation* und dient der Beurteilung der Erreichung der Ziele, der Wirkungen und Nebenwirkungen der implementierten Maßnahme (Wirksamkeitsforschung). Während ex-ante Evaluationen versuchen, die Auswirkungen einer Maßnahme zu antizipieren, dienen ex post-Evaluationen der Beurteilung von Effektivität und Effizienz einer bereits implementierten Maßnahme und haben *summativen Charakter*.

Die Begriffe **Selbst- und Fremdevaluation** beziehen sich auf den Sitz der steuernden Instanz einer Evaluation bzw. der Verantwortlichen, welche über die Kompetenzen verfügen, die Funktionen einer Evaluation zu definieren und über die Wahl des zu evaluierenden Gegenstandes bzw. Programms und die entsprechende Fragestellung entscheiden. Bei der Selbstevaluation befindet sich die steuernde Instanz innerhalb der zu beurteilenden Organisation. Bei Fremdevaluationen dagegen obliegt die Verantwortung einer außerhalb der Organisation befindlichen Institution (Ritz, 2003). Der Ort der Durchführung bezeichnet jene Instanz, die die relevanten Informationen erhebt, erfasst und auswertet. Diese Instanz kann sowohl innerhalb als auch von außerhalb der betreffenden Organisation liegen (Widmer, 2000). Von diesem Tatbestand leiten sich die Begriffe der **internen und externen Evaluation** ab. Je nach Umfang des Evaluationsvorhabens und Intensität der Evaluationsarbeit ist eine weitere Einteilung in **Kurz- und Langevaluationen** möglich. Diese unterscheiden sich in Kosten und Dauer. Dabei sind Kurzevaluationen, z.B. studentische Lehrevaluationen, kostengünstiger und führen schnell zu einem Ergebnis. Im Marburger Evaluationsmodell sind die einzelnen Lehrveranstaltungen Gegenstand der Evaluation. Steuernde und durchführende Instanz ist der Fachbereich Medizin selbst, so dass es sich hier per definitionem um eine interne Evaluation und eine Selbstevaluation handelt.

Aus der Vielzahl von Begriffsbestimmungen lassen sich bereits die allgemeinen Funktionen und gleichzeitig die Ziele von Evaluationen ableiten. Zusammenfassend können der Evaluation vier Funktionen zugeschrieben werden (Stockmann, 2000): Erkenntnisfunktion, Kontrollfunktion, Dialogfunktion und Legitimationsfunktion (vgl. Abb. 3). Die für die Steuerung von Prozessen entscheidungsrelevanten Daten werden durch Evaluationen im Sinne einer Bestandaufnahme (Bedarfsanalyse) erhoben. Sie geben Auskunft darüber, welchen Bedarf die Zielgruppe hat, ob folglich Defizite bestehen und ob veranlasste Maßnahmen gewünschte Veränderungen (Wirksamkeits-evaluation) erbracht haben (Erkenntnisfunktion).



**Abbildung 3: Funktionen von Evaluationen, aus Stockmann (2000)**

Die gewonnenen Erkenntnisse über das Programm bzw. Projekt werden zur Kontrolle und als Ausgangspunkt für Diskussionen verwendet. Bei der formativen Evaluation steht die Kontrolle der Wirkung des Programms bzw. Projektes im Mittelpunkt des Interesses mit dem Ziel Defizite schnell zu erkennen und steuernd einzugreifen (Kontrollfunktion). Die durch Evaluationen bereitgestellten Informationen dienen weiterhin als solide Grundlage für den Dialog zwischen den so genannten „Stakeholders“ (Dialogfunktion). Stakeholders sind Mittelgeber, Durchführungsorganisation, Zielgruppe sowie sonstige Beteiligte und Betroffenen. Im Rahmen dieses Dialoges erfolgt die Analyse der Defizite, um daraus Konsequenzen für die Planung und Entwicklung von Verbesserungsmaßnahmen zu ziehen. Über eine reine Beschreibung hinausgehend sollen Evaluationen also nicht nur bewerten, sondern die Planungsgrundlage für die Optimierung bestehender Prozesse darstellen (Rindermann, 2001), d.h. sie sind die Grundlage für lenkende und innovative Maßnahmen. Die Ergebnisse der Evaluationen können auch genutzt werden, um die Wirksamkeit der implementierten Maßnahmen nachzuweisen und letztendlich die Frage zu beantworten, mit welchem Input welcher Output erreicht wurde. Dadurch können Mittelgeber und die durchführende Organisation aufzeigen, mit welchen finanziellen Mitteln welche Wirkung erreicht wurde. Evaluationen können damit einen Beitrag zur Transparenz und Rechenschaft über verwendete Mittel leisten (Legitimationsfunktion).

Zusammengefasst kann gesagt werden, dass sich die Evaluation in ihren Funktionen weit über das Sammeln von Aussagen, deren Aufbereitung und die Auswertung der gewonnenen Daten hinaus erstrecken. Evaluation bringt den notwendigen Informationsgewinn, zeigt Stärken und Schwächen eines Programms bzw. Projektes

und schafft somit die notwendige Transparenz und die Grundlage für Diskussionen über Qualität. Evaluation ist Teil eines Lernprozesses, der die Analyse komplexer Situationen und die Umsetzung von Veränderungen initialisieren soll. Sie wird damit fester Bestandteil im Prozess der Qualitätsverbesserung.

### **3.3. Evaluationen in der Hochschule**

Evaluation an Hochschulen kann als regelmäßige und systematische Beschreibung und Bewertung der erbrachten Leistungen in Lehre und Forschung betrachtet werden. Aufgrund der Thematik der vorliegenden Arbeit beziehen sich die weiteren Ausführungen auf die Evaluation universitärer Lehre.

#### **3.3.1. Arten und Funktion der Hochschulevaluation**

Der Gegenstand der Evaluation zur Beurteilung universitärer Lehre entspricht den Ebenen der universitären Organisationsstruktur (El Hage, 1996). Folglich zählen dazu die einzelnen Lehrveranstaltungen, die Fakultät / der Fachbereich, die einzelnen Lehrstühle und die Hochschule als Institution. Die erste Ebene der Evaluation im Hochschulbereich ist die Lehrveranstaltungsevaluation. Es werden dabei in der Regel die Urteile der Studierenden über die Qualität der angebotenen Lehrveranstaltung erhoben. Auf der zweiten Ebene werden zusätzlich Informationen über Studienverlauf, Prüfungen, Forschungsleistungen und die Qualität der Verwaltung und Organisation im Fachbereich gesammelt. Die letzte Ebene der Evaluation hat die Verwaltungsleistung und Gesamtstruktur der Hochschule als Bewertungsgegenstand. Evaluationen auf Ebene der Veranstaltungen wie z.B. mittels studentischer Lehrveranstaltungsevaluationen stellen somit nur den Anfang des Dialoges über die Qualität an der Hochschule dar. Es gilt zu beachten, dass mit der Erhebung der Qualität der Lehre in einzelnen Lehrveranstaltungen lediglich ein Teil der Studienbedingungen erfasst wird, zusätzlich bedürfen organisatorische Aspekte, Ausstattung von Arbeitsplätzen und Bibliotheken sowie die Betreuungsangeboten für Studierende der Beachtung. Dies zu erfassen bezeichnet Rindermann (1999) erst als Lehrevaluation.

Die im Hochschulbereich durchgeführten Lehrevaluationen lassen sich je nach Anwendung findendem Verfahren als interne oder externe bzw. Selbst- und



Fremdevaluation, als Kurz- oder als Langevaluationen beschreiben. Eines der häufigsten und auch im Marburger Evaluationsmodell eingesetztes Verfahren ist die interne Evaluation in Form studentischer Lehrveranstaltungsevaluation. Darin sieht Kromrey (2000) allerdings eine unangemessene Vereinfachung durch Reduktion der Problemsicht. Auf der einen Seite wird der komplexe Wirkungszusammenhang von Strukturen und Prozessen eines Studienfachs auf die Lehrpräsentation einer Lehrveranstaltung reduziert. Auf der anderen Seiten kommt es durch diese Art der Evaluation nur zur Befragung einer selbstselektiven Teilmenge von Studierenden, nämlich den Teilnehmern und Teilnehmerinnen der entsprechenden Veranstaltung. Daher wurden in Deutschland in den vergangenen Jahren Evaluationsagenturen wie die Zentrale Evaluationsagentur (ZEVA) in Niedersachsen und Evaluationsverbände wie der Verbund norddeutscher Universitäten gegründet, die ein an das niederländische Evaluationsmodell angelehntes zweistufiges Konzept, bestehend aus interner und externer Evaluation mit anschließendem Follow-up durchführen. Eine nähere Darstellung des Evaluationskonzept und der Entwicklung der Qualitätssicherung und Qualitätsverbesserung an Hochschulen in Deutschland erfolgt gesondert im Abschnitt 3.5.

### **3.3.2. Ziele der Hochschulevaluation**

Übergeordnetes Ziel von Lehrevaluationen an Hochschulen ist es, einen Beitrag zur Qualitätssicherung und zur kontinuierlichen Qualitätsverbesserung in der Lehre zu leisten. Um Verbesserungen zu initiieren, müssen Schwachstellen in einem System transparent werden. Dies geschieht auf einer soliden Informationsbasis, die durch Evaluation gewonnen wird. Die Transparenzschaffung begrenzt sich nicht nur auf die Leistungen in der Lehre und Forschung, sondern auch auf die Verwendung der staatlichen Mittel im Sinne einer Rechenschaftslegung, bei welcher der effiziente Einsatz der Ressourcen sowie die Effektivität der Lehr- und Studienorganisation nachzuweisen sind. Mit der Analyse der Schwächen kristallisieren sich auch die Stärken eines Studienganges heraus, welche ausgebaut und gefördert zur Profilbildung beitragen. Im wachsenden Wettbewerb wird ein eigenständiges Fach- oder Hochschulprofil von zunehmender Bedeutung sein. Die aus Profilbildung und Wettbewerb resultierende und größer werdende Differenziertheit der Hochschulen und ihrer Ausbildungsangebote erfordert vergleichende Informationen zur Qualität von

Studium und Lehre (Block, 2000). Diese Informationen dienen der Orientierung von Studienberechtigten, Eltern, Arbeitgebern, Politikern, aber auch der Orientierung von Hochschulen selbst.

Evaluationen können auch als Grundlage für eine leistungsbezogenen Mittelvergabe dienen (Wissenschaftsrat, 1996). In Deutschland existieren Überlegungen eine solche Verknüpfung der Evaluationsergebnisse mit der Vergabe von Mitteln in einem Fach, die zu einer Erhöhung oder einer Verringerung der zuzuweisenden Mittel führen, vorzunehmen. Die Rolle der Evaluation als hochschulinternes Kontroll- und Steuerungsinstrument, an welches finanzielle Anreize und Sanktionen geknüpft werden, ist jedoch nicht unumstritten. So sprachen sich in einer Studie von Mittag et al. (2003) 40% der Befragten aus den Reihen der Gutachter und Fachangehörigen generell gegen eine solche Verknüpfung aus. 20% möchten dies nur, wenn die Evaluationsergebnisse mit einer Erhöhung der Mittelzuweisung verbunden sind, ein Prozent im Fall einer Verringerung. Die verbleibenden 39% erachteten sowohl negative als auch positive Sanktionen als eine Konsequenz der Evaluation für sinnvoll. Damit sprach sich die Mehrheit der Befragten für eine leistungsbezogenen Mittelvergabe basierend auf Evaluationsergebnissen aus.

Die Ziele von Evaluationsvorhaben an Hochschulen lassen sich wie folgt zusammenfassen

- Informationsgewinn im Sinne einer Bestandsaufnahme mit Stärken-Schwächen-Analyse (Erkenntnisfunktion)
- Entstehung eines Diskurses über Qualität von Lehre und Studium sowie Entwicklung eines Qualitätsbewusstseins (Dialogfunktion) und daraus
- Initialisierung von Maßnahmen zur Veränderung und Verbesserung der Organisationsstruktur und der Lehrqualität und Überprüfung derer Wirksamkeit (Kontrollfunktion)
- Transparenz und Rechenschaftslegung der Hochschule gegenüber der Gesellschaft (Legitimationsfunktion).

Unter Berücksichtigung der Ausführungen in Kapitel 2 zur Qualitätsverbesserung in der Lehre stellt die Evaluation der Hochschullehre den Anfang des beschriebenen Kreislaufprozesses dar. Als Bedarfsanalyse bringt sie lediglich den notwendigen Informationsgewinn, um Schwächen aufzudecken und die Wirksamkeit bisher ergriffener Maßnahmen zu überprüfen. Auf diesen Erkenntnissen basierend sind als Konsequenz Verbesserungsmaßnahmen zu entwickeln und zu implementieren. Die Evaluation der Lehre ist nicht mit Qualitätsverbesserung gleichzusetzen, aber werden ihre Ergebnisse entsprechend genutzt, so leistet sie dazu einen entscheidenden Beitrag.

### **3.4. Studentische Lehrevaluation**

Studentische Lehrevaluationen erfolgen in der Regel mittels Fragebogen, mit dem die unterschiedlichen Kriterien für Qualität der Lehre erfasst werden sollen. Dabei können - wie im Marburger Evaluationsbogen - geschlossene Fragen mit offenen Fragen kombiniert werden. Die Fragestellungen können unterschiedlich weit die Qualitätsmerkmale der Lehre erfassen, ausgehend von der aktuellen Lehrpräsentation als Lehrveranstaltungsevaluation (Prozessqualität) über Organisation und Ausstattung des Studiums (Strukturqualität) bis hin zu Absolventenbefragungen im Sinne einer ex-post Evaluation (Ergebnisqualität). Neben diesen durch die Hochschule beeinflussbaren Faktoren können auch Informationen zur individuellen Studiensituation und Studienverhalten erhoben werden, z.B. die Anzahl des Vorlesungsbesuchs. Ist die Effektivität der Lehre in Form von Lerngewinn von besonderem Interesse, so können auf diesen Lehrerfolg bezogenen Prozess- und Strukturvariablen im Fragebogen erhoben werden und anschließend –wie im Marburger Evaluationsmodell - mit den Leistungstestergebnissen zusammengeführt werden (Rindermann, 2001).

Da im empirischen Anteil der Arbeit der Marburger Evaluationsbogen methodisch überprüft und die mit ihm erhobenen Daten zur Lehrevaluation ausgewertet werden, soll im folgenden Abschnitt speziell die Problematik möglicher Erhebungs- und Auswertungsartefakte, der Einfluss von Bias-Variablen sowie die allgemeine Gültigkeit der Evaluationsdaten, die durch die Befragung von Studierenden erhoben wurden, erörtert werden.

### 3.4.1. Datenerhebungs- und Auswertungsprobleme

Nach empirischen Erkenntnissen aus der Bildungsforschung hängt der Verlauf des Studiums und ein mehr oder minder erfolgreicher Abschluss in hohem Maß von Merkmalen der individuellen Studiensituation ab (Kromrey, 2001). Diese, auf die Bewertung einer Veranstaltung Einfluss nehmenden Faktoren können die Ergebnisse der Evaluation verzerren (Rindermann, 2001). Solche so genannten Bias-Variablen sind z.B. die Lebenssituation (z.B. Berufstätigkeit neben dem Studium), Interessen und Leistungsmotivation, Studienstil und Studienintensität (z.B. Anzahl der besuchten Veranstaltungen, Zeitaufwand für das Studium). Es wird daher kontrovers diskutiert, ob Studierende die Hochschullehre angemessen beurteilen können. Kromrey (2000) bezeichnet sie sogar als „Alltags-Evaluatoren“, die irgendetwas (was jeweils mit dem in der Frage angesprochenen Sachverhalt assoziiert wird) irgendwie („alles in allem“ oder „aus aktueller Erfahrung“ oder „mit Blick auf das Wesentliche“) unter irgendwelchen Gesichtspunkten (Nutzen für das Studium oder den späteren Beruf, aktuell persönliches Interesse oder abstrakt verallgemeinertes Interesse der Studierenden) bewerten. Kritikpunkte an studentischer Lehrevaluation leiten sich auch aus Problemen bei der Erhebung und Auswertung der Daten ab. Es kann dabei zu Erhebungs- und Auswerteartefakte kommen. Bei Erhebung der Daten ist zunächst der Einfluss der Teilnehmerstruktur und Teilnehmermotivation auf das Urteil zu beachten (Kromrey, 1994). Kromrey differenziert bezüglich der Teilnehmermotivation zwischen intrinsischer und extrinsischer Motivation. Intrinsisch motivierte Besucher haben ein besonderes Interesse an dem Fach bzw. der Thematik der Veranstaltung. Für extrinsisch motivierte Besucher sind externe Faktoren wie z.B. der Pflichtcharakter der Veranstaltung der ausschlaggebende Besuchsgrund. Lehrevaluationen werden gewöhnlich Mitte oder Ende des Semesters durchgeführt. Folge davon ist, dass die Ergebnisse in Wahl- und Wahlpflichtveranstaltungen durch die bis dahin stattgefundenen Selektion eine positive Bias aufweisen. Wahlpflichtveranstaltungen werden nach geraumer Zeit und damit auch zum Zeitpunkt der Evaluation von intrinsisch motivierten Studierenden besucht und damit auch von diesen positiv bewertet. Dagegen sind bei Pflichtveranstaltungen auch die extrinsisch motivierten Studierenden anwesend. Ihr Urteil fließt hier mit in die Gesamtwertung der Veranstaltung ein. Pflichtveranstaltungen werden daher systematisch negativer bewertet als freiwillige Veranstaltungen (Kromrey, 2000). Neben der Problematik des

Zeitpunktes ist auch die gewählte Methode des Datenrücklaufs Ursache einer möglichen Selektion. Werden Fragebögen mit der Bitte um Rückgabe den Studierenden lediglich mitgegeben, so werden bevorzugt die Studierenden, die motiviert sind sich mitzuteilen dieser Bitte nachkommen. Der Anteil Studierender mit mangelnder Motivation zur Rückgabe wird in den in dieser Form erhobenen Daten unterrepräsentiert sein. Eine weitere Ursache für Erhebungsartefakte kann in der Konzeption des Fragebogens selbst liegen. Unter Studierenden gibt es bei der Urteilsfindung zwei Tendenzen: Globalaussagen und differenzierte Bewertung (Kromrey, 2001). Werden zunächst durch spezielle Items Details zu einer Dimension erfragt und anschließend ein zusammenfassendes Urteil z.B. in Form einer offenen Frage erhoben, so zeigt ein Teil der Studierenden die Neigung zur Pauschalisierung. Die Tendenz der Einschätzung der Items und das zusammenfassende Gesamturteil sind gleich und damit waren die Studierenden entweder mit allem zufrieden oder unzufrieden. Die zweite Gruppe findet zu einer differenzierten Bewertung bei Beantwortung der Detail-Fragen, ist also mit Teilaspekten zufrieden oder kritisiert diese, und bildet bei der Gesamtbewertung einen subjektiven Mittelwert. Bei diesen Studierenden würde unter Verwendung kurzer Fragebögen mit allgemein formulierten Fragestellungen wie z.B. „Waren sie mit der Veranstaltung zufrieden?“ ein wichtiger Teil an Informationen verloren gehen.

### **3.4.2. Gültigkeit der Evaluationsergebnisse**

Wenn nach Kromrey (2001) Studierende „irgendetwas irgendwie unter irgendwelchen Gesichtspunkten“ beurteilen, so stellt sich im Rahmen der Auswertung von Evaluationsdaten die Fragen, ob die studentischen Urteile überhaupt valide sind, was die durch studentische Evaluationen gewonnen Daten repräsentieren und wie bedeutend der Einfluss von Bias-Variablen auf die Evaluationsergebnisse ist.

Die Ergebnisse studentischer Lehrevaluation können die spezifische Teilnehmerzusammensetzung (Kromrey, 2001), das Lehrverhalten des einzelnen Dozenten oder das Interesse am Thema selbst reflektieren. Die Evaluationsergebnisse gelten dann als valide, wenn sie das Geschehen in einer Lehrveranstaltung adäquat beschreiben. Rindermann (2001) konnte bei Vergleich studentischer Veranstaltungsurteile mit den Urteilen geschulter Fremdevaluatoren eine Korrelation auf mittlerem Niveau nachweisen und bezeichnet daher die studentische Beurteilung

zumindest als „mittelvalide“. Dennoch ist die Gültigkeit von studentischen Evaluationsergebnissen aufgrund fraglichen Einflusses von Bias-Variablen nicht unumstritten. Rindermann (2001) beschreibt z.B. die beobachteten Zusammenhänge zwischen Beurteilung der Veranstaltung und Vorinteressen sowie Besuchsgrund (intrinsisch oder extrinsisch motiviert) als bedeutend und stabil. Ebenso wie Kromrey stimmt er darin überein, dass Themen vor einer interessierten Hörschaft (Wahlpflichtveranstaltung) Lehre und Lernerfolg erleichtern und/ oder zu positiv verzerrten Beurteilungen führen können. Nicht abschließend geklärt sei allerdings die Stärke dieses Bedingungs-Wirkungs-Verhältnisses. Die Frage, ob die Motivation der Teilnehmer und damit die Teilnehmerzusammensetzung einen Einfluss auf das Gesamturteil über eine Veranstaltung hat, ist ein Aspekt der empirischen Analyse des Marburger Evaluationsbogens. Als relevante Bias-Variable auszuschließen sind dagegen nach Rindermann (2001) biographische Variablen wie Alter, Geschlecht, Semesterzahl sowie das Leistungsniveau der Studierenden (z.B. Abiturnote, Prüfungsergebnisse).

Die studentische Veranstaltungskritik kann einen Beitrag zur Verbesserung der Lehre leisten. Sie kann Hinweise auf folgende Aspekte liefern: Grad der Zufriedenheit bzw. Unzufriedenheit mit der Lehre, Schwachstellen in der Lehre sowie im Lernverhalten der Studierenden, Beliebtheit der einzelnen Themen und Akzeptanz der Lern- und Arbeitshilfen. Neben diesen eher den Lehr- und Lernprozess betreffenden Aussagen, können auch strukturelle Mängel (Räumlichkeiten, Ausstattung) aufgedeckt werden. Bewertungen von Lehrveranstaltungen durch Studierende sind allerdings nur dann im Prozess der Verbesserung der Lehre hilfreich, wenn die Ergebnisse an die Studierenden und Lehrenden rückgemeldet und diskutiert werden und die Bereitschaft und Fähigkeit zur konstruktiver Kritik und für Veränderungen auf Seiten der Lehrenden und Lernenden besteht.

### **3.5. Die Entwicklung systematischer Verfahren zur Qualitätsverbesserung an Hochschulen**

Systematische Verfahren für die Qualitätsverbesserung im Hochschulbereich etablierten sich in Europa seit Mitte der 80er Jahre. In verschiedenen europäischen Hochschulsystemen – allen voran in Frankreich, den Niederlanden und Großbritannien

– wurden Lenkungsmechanismen entwickelt und erprobt, die die bisherige allumfassende staatliche Steuerung der Hochschulen schrittweise ablösen und durch Stärkung der Hochschulautonomie neue Möglichkeiten eröffnen sollten, um auf dezentraler Ebene schneller und flexibler auf die vielfältigen gesellschaftlichen und wissenschaftlichen Herausforderungen reagieren zu können. Die Verantwortlichen in diesen Ländern waren sich darüber einig, dass staatliche Deregulierung und Stärkung der Autonomie der Hochschulen mit entsprechendem Kompetenzzuwachs nur bei gleichzeitiger Entwicklung von Verfahren zur Qualitätssicherung und –verbesserung sowie Verfahren zur Schaffung von Transparenz und Rechenschaftslegung möglich waren (Carstensen, 1997). In diese europäische Entwicklung ist auch die Diskussion über Qualität in der Lehre und die Etablierung von Evaluationsverfahren in Deutschland eingebettet, wenn gleich die ersten Evaluationsagenturen und länderübergreifenden Evaluationsnetzwerke erst in der Mitte der 90er Jahre ihre Arbeit aufnahmen. Die von ihnen zur Anwendung kommenden Evaluationsverfahren sind stark an das niederländische Modell angelehnt, welches im Folgenden kurz vorgestellt wird.

### **3.5.1. Das niederländische Evaluationskonzept**

Die zentrale Idee des niederländischen Modells ist es, die stärkere Selbststeuerung und Selbstkontrolle der Hochschule zu ermöglichen und gleichzeitig eine Steigerung des Qualitätsbewusstseins zu bewirken. Die Vereinigung der niederländischen Universitäten (vsnu) hat dafür in den 80er Jahren ein einheitliches Qualitätsbewertungssystem eingeführt, welches die an den Hochschulen angebotenen Studiengänge in einem Zyklus von 6 Jahren beurteilt.

Zunächst nehmen die Hochschulen mittels Selbstevaluationsbericht Stellung zu den auf der Prüfliste der Rektorenkonferenz aufgeführten Punkte: Studienorganisation, Stärken und Schwächen, Ausstattung des Fachbereichs, Studienerfolgsquote, studentischer Veranstaltungsbewertung und Zufriedenheit von Studierenden und Lehrenden insgesamt. Nach dieser internen Evaluation folgt der mehrtägige Besuch durch externe Gutachter, die sogenannte „peer review“. Die Gruppe der Sachverständigen (peers) setzt sich aus niederländischen und ausländischen Lehrenden, Fachkollegen aus der Berufspraxis, einem Studierenden und einem Hochschuldidaktiker zusammen. Bestandteile des Besuches sind z.B. Gespräche mit dem Dekan, Studierenden,

Lehrenden und technisch-administrativen Personal und die Besichtigung von Gebäuden, Laboratorien und Bibliotheken. Das Komitee fasst die gewonnenen Eindrücke und die Ergebnisse der Selbstevaluation in einem vorläufigen Bericht mit Empfehlungen zusammen und trifft sich nochmals mit den Repräsentanten des beurteilten Fachbereichs, um die Ergebnisse zu diskutieren und Irrtümer oder Missverständnisse zu korrigieren. Abschließend verfasst das Komitee den endgültigen Bericht für das Wissenschaftsministerium, welcher veröffentlicht wird. Der internen und externen Evaluation schließen sich Metaevaluation und Follow-up an. Das staatliche Inspektorat überprüft Methoden und Verfahren des angewandten Evaluation (Metaevaluation) und verfasst einen Bericht für das Wissenschaftsministerium, welcher ebenso sechs Monate nach Veröffentlichung des Evaluationsberichtes veröffentlicht wird. Der öffentliche Träger kann sich somit ein Urteil über die Qualität des Qualitätssicherungssystems der vnsu bilden, denn er behält die generelle Verantwortung für die Qualität in der Lehre. Das Follow-up umfasst die Vereinbarungen des Fachbereich und der Hochschulleitungen über zu ergreifende Maßnahmen zur Verbesserung der Lehre und des Studiums, zur Optimierung der Ergebnisse bzw. zur Sicherstellung bestimmter zu erreichender Standards innerhalb eines definierten Zeitraums (Kromrey, 2000). Nach sechs bis acht Monaten ist dann nach Aufforderung ein Bericht über bereits ergriffene und geplante Maßnahmen im Wissenschaftsministerium einzureichen. Zeigt sich in den folgenden Jahren keine Verbesserung, so kann es im Sinne von Sanktionen zu einer Kürzung der finanziellen Mittel für das Studienprogramm kommen. Um aber den Fachbereichen ausreichend Zeit zur Qualitätsverbesserung zu geben, sichert der Minister zu, acht Jahre keine Ressourcenentscheidungen an die Ergebnisse der Evaluation zu knüpfen (Carstensen, 1997).

Bei dem hier vorgestellten niederländischen Evaluationskonzept handelt es sich um ein national einheitlich angewandtes und in Zusammenarbeit mit einer zentralen Agentur durchgeführtes Evaluationsverfahren. Ähnliche Verfahren finden sich in Großbritannien und Frankreich. In Frankreich wird die Evaluation der Lehre und das 1985 eingeführte Verfahren zur Qualitätssicherung an französischen Hochschulen von zentraler Stelle aus, vom Comité National d'Evaluation (CNE) gesteuert (Staropoli, 1995). In Großbritannien ist seit 1997 die Quality Assurance Agency for Higher Education (QAA) die zentrale Einrichtung für Qualitätssicherung in der Lehre (Frederiks, 2001).



### 3.5.2. Netzwerke, Verbände und Agenturen zur Hochschulevaluation in Deutschland

In Deutschland wurden erst Anfang der 90er Jahre systematische Verfahren für die Evaluation der Lehre eingeführt. HRK (1995) und Wissenschaftsrat (1996) führten 1994 und 1995 zwei Pilotprojekte durch, um zum einen die Diskussion über Evaluationsverfahren auf nationaler Ebene voranzutreiben und zum anderen anerkannte Regeln für die Qualitätssicherung und –verbesserung zu erarbeiten. Das Resultat der Pilotprojekte war die Empfehlung, Evaluationsverfahren nach niederländischem Vorbild mit interner und externer Evaluation auch in Deutschland durchzuführen.

Mit der Novellierung des Hochschulrahmengesetzes 1998 wurden die Hochschulen in §6 (HRG) dann zur Evaluation der Lehre verpflichtet:

*„Die Arbeit der Hochschulen in Forschung und Lehre, bei der Förderung des wissenschaftlichen Nachwuchses sowie der Erfüllung des Gleichstellungsauftrages soll regelmäßig bewertet werden. Die Studierenden sind bei der Bewertung der Qualität der Lehre zu beteiligen. Die Ergebnisse der Bewertung sollen veröffentlicht werden.“*

Aus der allgemein gehaltenen Formulierung lässt sich allerdings nicht unmittelbar ableiten, welche Methoden wie oft zur Evaluation angewandt werden und in welcher Form die Veröffentlichung stattfinden soll. Folglich reichen die Methoden der Lehrevaluation von alleiniger interner Evaluation in Form studentischer Lehrevaluation über eine reine externe Evaluation sowie die Kombination beider Verfahren bis hin zu Evaluationen durch Unternehmensberatungen und zentrale Agenturen.

Bei einem Vergleich der deutschen mit den internationalen Evaluationsverfahren ist dementsprechend zu berücksichtigen, dass aufgrund des späten Einstiegs weniger Erfahrungen über die Umsetzung von Evaluationserfahrungen vorliegen und dass die Vielfalt der vorliegenden Evaluationskonzepte durch die Kulturhoheit der einzelnen Bundesländer zusätzlich gefördert wird (Carstensen, 1997). Dies hat zur Folge, dass eine Positionierung im Wettbewerb noch nicht in dem Maße stattgefunden haben kann wie im Ausland (Fischer-Bluhm, 2000). Seit den 90er Jahren zeigen sich allerdings in Deutschland auch zunehmend Entwicklungen hin zu einem nationalen

Evaluationsnetzwerk. Deutschlandweit kooperieren immer mehr Hochschulen in Bezug auf Evaluation ihrer Leistungen miteinander. Es wurden einzelne Agenturen, Netzwerke und Verbände von Hochschulen gegründet, die mehrstufige Evaluationsverfahren durchführen. Als erster länderübergreifender und hochschulübergreifender Verbund wurde 1994 der Verbund der norddeutschen Universitäten durch die Rektoren und Präsidenten der Universitäten Bremen, Rostock, Kiel, Hamburg und Oldenburg gegründet. Ziel war und ist die Verstärkung der Kooperation zwischen den Universitäten, insbesondere im Bereich der Lehre. 1995 hat die niedersächsische Landeshochschulkonferenz mit Unterstützung des Ministeriums für Wissenschaft und Kultur die Zentrale Evaluationsagentur (ZEVA) eingerichtet. Seit 1996 führt diese flächendeckend die Evaluation an niedersächsischen Universitäten und Fachhochschulen in Anlehnungen an das niederländischen Konzept durch. Die bisherigen Evaluationsprojekte in Niedersachsen haben dabei gezeigt, dass die gewünschte Verbesserung der Lehre und des Studiums mit diesem Verfahren zu erreichen waren (Hegewald & Reuke, 1997). Grundlage dafür war, dass durch den Evaluationsprozess Fachprofile erarbeitet wurden und so auch Transparenz in den Fächern über Stärken und Schwächen geschaffen wurde. Die Liste weiterer Agenturen und Hochschulnetzwerke, welche in den letzten Jahren gegründet wurden, lässt sich durch folgende Beispiele fortsetzen (HRK, 2000; Mittag et al. 2003):

- Das Netzwerk Bildungsforschung der Länder Bremen, Flandern, Niederlande, Niedersachsen und Nordrhein-Westfalen (seit 1995)
- Die Geschäftsstelle für Evaluation an Universitäten (seit 1996) und Fachhochschulen (seit 1997) in Nordrhein-Westfalen
- Das Evaluationsbüro für die Lehrevaluation in der Universitätspartnerschaft Halle, Jena und Leipzig (seit 1996)
- Der länderübergreifende Verbund der TU Darmstadt, der Universitäten Kaiserslautern und Karlsruhe in Verbindung mit der ETH Zürich (seit 1999)
- Evaluationsnetzwerk Wissenschaft (ENWISS) der Universitäten und Fachhochschulen in Hessen, Rheinland-Pfalz und Thüringen (seit 2001)
- Evaluationsagentur Baden-Württemberg (evalag, seit 2001)

Es haben sich in den bisherigen Entwicklungen zwei Organisationsmodelle für evaluierende Agenturen herauskristallisiert (Schreier, 2000). Auf der einen Seite sind

dies die länder- und hochschulübergreifenden Verbände wie der Verbund norddeutscher Universitäten oder die Kooperation der Universitäten Jena, Halle und Leipzig. Auf der anderen Seite gibt es mit z.B. der ZEvA Niedersachsen auch Agenturen, die auf Länderbasis die entsprechenden Hochschulen evaluieren. Ziel der Agenturen, Netzwerke und Verbände, so Schreier (2000), sollen Informationsgewinn, Erfahrungsaustausch und auf lange Sicht eine länderübergreifende Vernetzung und konzeptionelle Weiterentwicklung sein.

### **3.6. Fazit**

Qualität in der Hochschullehre ist nicht allein durch didaktisch erfolgreiche Lehrveranstaltungen erreichbar, sondern sollte in ein entsprechendes Umfeld, d.h. inhaltlich adäquat und gut strukturiertes Curriculum und günstige institutionelle Rahmenbedingungen eingebettet sein. Hierzu, gerade bei inhaltlichen Fragen, sollten nicht nur - wie durch das HRG §6 verlangt - Studierende befragt werden, sondern auch Selbst- und Fremdbeurteilungen durch Fachkollegen herangezogen werden (Rindermann, 2001). Die studentische Lehrevaluation ist für die Evaluation von Studium und Lehre nicht ausreichend und behaftet mit Erhebungs- und Deutungsproblemen. Die interne Evaluation der Lehre unter Einbezug der Studierenden ist zwar fester Bestandteil internationaler und nationaler Evaluationskonzepte, deren Ergebnisse aber durch die Daten der Lehrberichte und die Ergebnissen der externen Evaluation ergänzt werden. Eine solche Evaluation von Studium und Lehre ist zeit- und ressourcenaufwendig. Deshalb sollte sich eine durchgeführte Evaluation für die Beteiligten auch einen erkennbaren Nutzen bringen und nicht als „Selbstzweck“ durchgeführt werden. Das Potential der Evaluation als Beitrag zur Qualitätsverbesserung in der Lehre und als Instrument der hochschulinternen Steuerung ist hervorzuheben. Evaluationsergebnisse stellen die Informationsgrundlage zur Entwicklung von Maßnahmen zur Verbesserung der Qualität in Lehre und Studium dar und können auch zur Wirksamkeitskontrolle der implementierten Maßnahmen eingesetzt werden. Sie können des Weiteren als Grundlage für Entscheidungen zur Mittelvergabe nach Leistungs- und Belastungskriterien genutzt werden. Kromrey (2000) warnt allerdings davor, Evaluationen als reines Kontrollinstrument zu verwenden. Die Knüpfung von Sanktionen an „schlechte“ Ergebnisse führe zu einer eher geringen Akzeptanz des Verfahrens und darüber hinaus könne dies sogar die

Evaluationsergebnisse beeinflussen, indem kritische Informationen verschleiert und positive Informationen besonders in den Vordergrund gestellt werden.

## **4. Das Marburger Evaluationsmodell**

### **4.1. Der Marburger Fragebogen zur interne Evaluation und leistungsorientierten Mittelvergabe in der Lehre**

Im Fachbereich Humanmedizin der Philipps-Universität Marburg hat man sich im Rahmen der gesetzlichen Verpflichtung zur Evaluation der Lehre (HRG §6) für eine interne Evaluation des Lehrangebotes entschieden. Die Kommission des Studienausschusses zur leistungsbezogenen Mittelvergabe in der Lehre unter der Leitung des Fachvertreters der Medizinischen Psychologie entschied, in Orientierung an den Ausführungen Rindermanns (2001) und in Anlehnung an Donabedians Konzept zur Beurteilung der Qualität medizinischer Versorgung auch die Qualität der medizinischen Lehre unter den Aspekten Prozess-, Struktur- und Ergebnisqualität zu bewerten (Donabedian, 2003) und die Ergebnisse in die Entscheidung über Zuweisung finanzieller Mittel einfließen zu lassen. Daten zur Ergebnisqualität in einem Fachgebiet liefern dabei die Resultate der medizinischen Staatsexamen des IMPP. Struktur- und Prozessqualität soll im Rahmen studentischer Lehrevaluation mit dem Marburger Fragebogen zur Evaluation des Lehrangebotes in der Medizin bewertet werden. Die jeweiligen Ergebnisse sollen in ein Punktesystem übertragen werden, dem dadurch erreichten Rang unter den Fachgebieten wird ein Geldwertäquivalent zugeordnet. Aus dem Gesamtpool für leistungsbezogene Mittelvergabe entfallen 30% auf die Ergebnisqualität und 60 % auf Struktur- und Prozessqualität. Die verbleibenden 10% entfallen auf Sonderleistungen eines Fachgebietes, welche z.B. das Verfassen und Herausgeben von Lehrbüchern oder die didaktische Neukonzeption von Praktika oder Seminaren sein kann. Die Konzeption des Marburger Fragebogens wird im Abschnitt 5.2. ausführlich erläutert.

### **4.2. Entwicklung der empirischen Fragestellungen**

Der Marburger Fragebogen zur Evaluation des Lehrangebotes in der Medizin ist Gegenstand der psychometrischen Überprüfung dieser Arbeit. Damit das Urteil der Studierenden in die Gesamtbewertung eines Faches und der sich daran knüpfenden finanziellen Folgen einfließen kann, muss der Fragebogen verschiedene Anforderungen erfüllen, aus denen sich die folgenden Fragestellungen ableiten lassen:

1. Es soll überprüft werden, ob die ausgewählten Items zu Aspekten der Struktur- und Prozessqualität jeweils eine eindimensionale Skala darstellen. Die Eindimensionalität einer Skala ist Voraussetzung dafür, dass der Mittelwert über die Items als Urteil der Studierenden über eine Lehrveranstaltung in die Gesamtbewertung eines Faches einfließen kann. Es ist zu untersuchen, ob der Fragebogen die zwei gewünschten Dimensionen Struktur- und Prozessqualität getrennt voneinander zu erfassen vermag.
2. Es wird vermutet, dass der vorliegende Fragebogen reliabel ist. Die Reliabilität der geschlossenen Fragen soll mittels Konsistenzanalyse überprüft werden.
3. Es gilt zu überprüfen, ob und inwieweit die Motivation der Teilnehmer einen Einfluss auf das Gesamturteil über eine Veranstaltung hat. Dafür werden die Bewertungen der Wahlpflichtveranstaltungen (Vorlesungen) und die Urteile über die Pflichtveranstaltungen (Kurs, Praktika und Seminar) vergleichend untersucht. Kromry (1994) konstatiert, dass ein Zusammenhang zwischen Motivation bzw. Besuchsgrund einer Veranstaltung und der Bewertung besteht. Zu diesem Schluss kamen auch Rindermann (2001) sowie Spiel und Gössler (2000). Die Teilnehmerstruktur bestimme in hohem Maße, ob das „durchschnittliche“ Urteil für eine Veranstaltung gut oder schlecht ausfalle. Pflichtveranstaltungen werden demnach schlechter beurteilt, weil hier auch das Urteil unmotivierter Studierender in die Bewertung einfließt. Vorlesungen dagegen, die überwiegend vor einer interessierten Hörerschaft gehalten werden, erhalten mehr positive Beurteilungen. Konträr dieser Aussage gegenüberstehend sind die Untersuchungsergebnisse Bergers et al. (2003), aus denen sich die vierte Fragestellung ableitet.
4. Es ist zu untersuchen, ob die Häufigkeit der Anwesenheit in einer Veranstaltung Einfluß auf das Urteil der Studierenden bei Beantwortung der geschlossenen Fragen hat. Berger et al. (2003) stellten fest, dass die Häufigkeit des Veranstaltungsbesuches die Ergebnisse der Lehrveranstaltungsevaluation systematisch verzerrt. Dabei wurden Veranstaltungen mit Anwesenheitspflicht wie Seminare signifikant um ca. eine Drittelnote besser bewertet werden als

freiwillige Vorlesungen. Studierende mit einer angegebenen Anwesenheit von 80% in Vorlesungen bewerteten die Veranstaltung um ca. eine halbe Note besser als Studierende die diese seltener besucht haben. Der Marburger Evaluationsbogen schließt die Angabe über die Häufigkeit der Anwesenheit für Vorlesungen in Form eines Items ein. Für diesen Veranstaltungstyp soll überprüft werden, ob die Häufigkeit der Anwesenheit Einfluss auf das studentische Urteil hat. Neben Berger et al. (2003) konnten auch Spiel und Gössler (2000) einen, wenngleich geringen Zusammenhang zwischen Anwesenheit und Evaluationsergebnissen zeigen.

5. Es soll untersucht werden, ob die Höhe der Semesterzahl einen Einfluss auf das Veranstaltungsurteil der Studierenden hat. Die Ergebnisse Rindermanns (2001) zeigen insgesamt, dass Verzerrungen durch personenbezogene Variablen wie z.B. das Geschlecht, die Abiturnote und Semesterzahl ausgeschlossen werden können. Diese Feststellung soll für den Einfluss der Semesterzahl auf das Urteil der Studierenden mit den Daten dieser Studie überprüft werden. Die Ergebnisse der geschlossenen Fragen werden entsprechend den drei Studienabschnitten Vorklinik, 1. klinischer Abschnitt und 2. klinischer Abschnitt auf signifikante Unterschiede überprüft.
6. Durch eine qualitative Datenanalyse soll überprüft werden, ob sich anhand der Antworten der Studierenden auf die offenen Fragen ein Kategoriensystem entwickeln lässt, in das die Aussagen reliabel eingeordnet werden können. Neben der Bestimmung der Reliabilität des Kategoriensystems soll durch eine Frequenzanalyse ermittelt werden, welche Kategorien besonders häufig Inhalt der Meinungsäußerungen sind. Auch bei den offenen Fragen soll untersucht werden, ob sich signifikante Unterschiede in der Häufigkeit der Nennung einer Kategorie abhängig von der Semesterzahl zeigen. Für die Untersuchung der freien Meinungsäußerungen werden zwei Gruppen gebildet, Studierende des vorklinischen und Studierende der klinischen Studienabschnitte.

Die Überprüfung der Dimensionalität der Skala, der Reliabilität des Fragebogens und die Entwicklung eines Kategoriensystems sind die Schwerpunkte des statistischen Teils dieser Arbeit. Mit der Untersuchung der Bias-Variablen Anwesenheit, Motivation bzw.

Besuchsgrund und Semesterzahl sollen mögliche Verzerrungseffekte in der studentischen Lehrveranstaltungsevaluation aufgedeckt werden.



## **5. Methodik der empirischen Untersuchung**

In diesem Kapitel werden die Datengewinnung, die Konzeption des Marburger Fragebogens zur Evaluation des Lehrangebots in der Medizin und die angewandten Auswertungsverfahren erläutert. Es wird beschrieben, wie die Daten erhoben wurden, wie der Fragebogen aufgebaut ist, welche Methoden zur Überprüfung der Fragestellung und welche Kriterien zur Bestimmung der Gütekriterien verwandt worden sind.

### **5.1. Stichprobengewinnung**

Die Evaluation mittels des Marburger Fragebogens fand erstmals am Ende des Sommersemesters 2002 im Fachbereich 20 der Philipps-Universität Marburg statt. Um eine hohe Anzahl Studierender zu erreichen, wurden die Bögen in einer der anwesenheitspflichtigen Kursstunden oder zusammen mit einer Klausur ausgeteilt. Für die Studierenden des vorklinischen Abschnitts wurden für das 2. Semester die letzte Unterrichtsstunde des Histologiekurses und für das 4. Semester das letzte Biochemie-Praktikum gewählt. Da zum Wintersemester 2000/2001 das Studienjahr für den Studiengang Humanmedizin an der Philipps-Universität eingeführt wurde, gab es im Erhebungszeitraum kein erstes bzw. drittes vorklinisches Semester. Im klinischen Abschnitt erhielten die Studierenden den Fragebogen in den Kursen allgemeine Pathologie (1. klinische Semester), Mikrobiologie (2. klinische Semester) und spezielle Pathologie (3. klinische Semester). Zusammen mit einer Abschlussklausur wurden die Fragebögen an das vierte bis sechste klinische Semester ausgehändigt. Das Verteilen und Einsammeln der Bögen oblagen der beauftragten studentischen Hilfskraft und den Unterrichtenden. Auf diese Weise wurden 657 Fragebögen im Sommersemester 2002 von den Studierenden ausgefüllt und an die studentische Hilfskraft bzw. Unterrichtenden zurückgegeben.

### **5.2. Konzeption des Fragebogens**

Der Marburger Fragebogen besteht aus offenen und geschlossenen Fragen und ermöglicht eine getrennte Beurteilung von Pflichtveranstaltungen (Kurse, Seminare und Praktika) und freiwilligen Vorlesungen.

Der aus 6 Seiten bestehende Fragebogen gliedert sich in zwei Teilbögen, einen Teilbogen für Vorlesungen oder Seminare und einen Teilbogen für Praktika und Kurse. Auf jeder Seite sind die im Curriculum für das jeweilige Semester vorgesehenen Lehrveranstaltungen aufgeführt. Zu Beginn des Fragebogens ist der entsprechende Studiengang mit Humanmedizin, Zahnmedizin oder Humanbiologie anzugeben. Es folgen 12 geschlossenen Fragen, die auf einer Ratingskala von 1 (= sehr gut) bis 6 (=schlecht), bzw. 9 (= nicht vorhanden/nicht anwesend) beurteilt werden können.

Die erste Seite des Fragebogens mit dem Titel „Evaluation der Lehrstruktur“ beinhaltet 6 Items zur Strukturqualität (im weiteren Verlauf als SQ 1 bis SQ 6 bezeichnet), die sich auf folgende Punkte beziehen: *Erläuterung des Lehrkonzeptes* (SQ 1), *Abstimmung von Lehrinhalten und Prüfungsinhalten* (SQ 2), *Qualität der Unterrichtsmaterialien* (SQ 3), *Vorbereitung des Unterrichtenden* (SQ 4), *Ausfall der Veranstaltung* (SQ 5) und *Anwesenheit der Studierenden in der Veranstaltung* (SQ 6 für den Veranstaltungstyp Vorlesung) bzw. *Beitrag der Semesterhilfskräfte* (SQ 6 für Kurse, Praktika und Seminare). Die Ausprägungen der Items SQ 5 und SQ 6 werden auf einer Skala von 0% bis 100% in 20er Schritten gemessen. Item SQ 6 ist das einzige Item, welches zur Beurteilung der Strukturqualität bei Pflichtveranstaltungen und Vorlesungen unterschiedlich gewählt worden ist. Die Items SQ 1 bis SQ 5 sind bei allen Veranstaltungstypen identisch. Die 6 Items zur Prozessqualität (im weiteren Verlauf als PQ 1 bis PQ 6 bezeichnet) befinden auf der zweiten Seite des Fragebogens, welcher den Titel „Evaluation des Lehrprozesses“ trägt. Diese Items erfassen das Urteil der Studierenden über *Organisation* (PQ 1), *Didaktik* (PQ 2), *Lern- und Arbeitsatmosphäre* (PQ 3), *Prüfungsvorbereitung* (PQ 4), *Abstimmung der Lehrinhalte* (PQ 5) sowie *Prüfungs- und Praxisrelevanz des Unterrichts* (PQ 6). Zusätzlich bietet eine Seite am Ende des Fragebogens Platz für Antwortmöglichkeiten auf die offene Frage nach „Meinungen zu anderen Aspekten des Unterrichts“, „Qualität der Lehrveranstaltung“ und „ggf. Verbesserungsvorschläge“ zu jeder Veranstaltung des Semesters. Die Studierenden werden auf den vertraulichen Umgang mit ihren Angaben hingewiesen und gebeten, neben der allgemeinen Beurteilung einer von mehreren Dozenten gehaltenen Veranstaltung auch dozentenspezifische Wertungen abzugeben. Der vollständige Fragebogen ist im Anhang A einzusehen.

### 5.3. Methoden zur Ermittlung der Gütekriterien

In diesem Abschnitt des Kapitels werden die Methoden zur Ermittlung der Gütekriterien der Items vorgestellt. Die Berechnungen wurden mit dem statistischen Programm SPSS (Version 11.0) durchgeführt.

#### 5.3.1. Itemanalyse

Die Qualität eines Testes wird geprägt von der Art und Zusammensetzung seiner einzelnen Items. Zu den unter dem Begriff „Itemanalyse“ gezählten und in dieser Arbeit angewandten Verfahren zählen die Ermittlung von Mittelwerten (MW), Standardabweichungen (SD) sowie die Berechnung von Trennschärfe und Dimensionalität.

Der *arithmetische Mittelwert*  $\bar{x}$  beschreibt den Quotienten aus der Summe der Messwerte und ihrer Anzahl  $N$ .

Die *Standardabweichung*  $s$  stellt den positiven Wert unter der Wurzel aus der Streuung der Einzelmesswerte einer Messreihe um ihren Mittelwert (Varianz) dar.

Die *Trennschärfe* ist definiert als Korrelation zwischen einer einzelnen Itemantwort und dem Gesamtskalenwert, der Summe aller Items (Lienert, 1989). Sie gibt an, wie gut ein einzelnes Item das Gesamtergebnis eines Tests repräsentiert, d.h. wie gut das Testergebnis aufgrund einer einzelnen Itemantwort vorhersagbar ist. Um eine artifizielle Erhöhung der Trennschärfe durch Einbezug des aktuell betrachteten Items in die Berechnung des Gesamttestwertes auszuschließen, berücksichtigt der korrigierte *Trennschärfekoeffizient*  $r_{it}$  das aktuell betrachtete Item bei Ermittlung des Gesamttestwertes nicht. Für die für den Marburger Fragebogen gewünschte eindimensionale Skala sind hohe Trennschärfekoeffizienten mit Werten größer 0,5 angestrebt, da die geringe Trennschärfe eines Items auf einen schlechten Indikator für das angestrebte Konstrukt hinweist (Bortz & Döring, 2002).

Die *Dimensionalität* eines Tests besagt, wie viele Konstrukte mit den Testitems operationalisiert werden. Auf die Methode zur Ermittlung der Dimensionalität wird im Abschnitt 5.4.1.2. eingegangen.

### 5.3.2. Reliabilitätsanalyse

Die Reliabilität - als Grad der Genauigkeit mit welchem ein Test misst - wird durch Berechnung der internen Konsistenz unter Verwendung des Reliabilitätskoeffizienten Alpha nach Chronbach ermittelt. Dieser Koeffizient entspricht der mittleren Testhalbierungs-Reliabilität eines Tests für alle möglichen Testhalbierungen.

## 5.4. Methoden zur Überprüfung der Fragestellungen

Da der Fragebogen aus geschlossenen und offenen Fragen aufgebaut und diese getrennt voneinander analysiert wurden, werden die dabei angewandten Methoden ebenfalls getrennt beschrieben.

### 5.4.1. Analyse der geschlossenen Fragen

Der Analyse der geschlossenen Fragen ging die Selektion der Datenbasis entsprechend der Ausschlusskriterien sowie die Komplettierung und Transformation der Daten voraus. Es folgten zur Beantwortung der Fragestellung Faktorenanalyse, Varianzanalyse sowie der U-Test nach Mann und Whitney.

#### 5.4.1.1. Aufarbeitung des Datensatzes

Als Ausschlusskriterien galten mehr als 2 missing data bei der Beurteilung einer Veranstaltung oder eine angegebene Anwesenheit von unter 20% in einer Vorlesung (Item SQ 6, *Anwesenheit der Studierenden in der Veranstaltung*). Dadurch mussten 60 Fragebögen von der Analyse ausgeschlossen werden. In den verbleibenden Fällen wurden fehlende Angaben durch den Mittelwert des entsprechenden Items über alle Studierende ersetzt.

Bei Item SQ 3 (*Qualität der Unterrichtsmaterialien*) codiert die Ziffer 9 das Fehlen eines Skriptes. Bei der Auswertung wurde nach Beschluss der Kommission des Studienausschusses zur leistungsbezogenen Mittelvergabe in der Lehre die Ziffer 9 durch die Note 6 ersetzt. Die Angaben zu den Items SQ 5 und SQ 6 in Vorlesungen sowie Item SQ 5 in Kuren und Praktika wurden in eine sechsstufige Ratingskala

transformiert. Dadurch ergibt sich für SQ 6 bei 100% die Note 1, bei 20% die Note 5. Bei SQ 5 entsprechen die Angaben von 0% Unterrichtsausfall der Note 1, 50% entsprechen der Note 6.

#### 5.4.1.2. Faktorenanalyse

Um die Dimensionalität der Skala für die geschlossenen Fragen zu ermitteln, wird eine *Faktorenanalyse* durchgeführt. Die Faktorenanalyse ist ein datenreduzierendes Verfahren. Ziel ist es, Faktoren zu ermitteln, welche die beobachteten Zusammenhänge zwischen den gegebenen Variablen möglichst vollständig erklären und eine inhaltlich sinnvolle Interpretation ermöglichen. Variablen, die untereinander hoch korrelieren, werden dabei zu Faktoren zusammengefasst. Ein Faktor stellt dabei ein theoretisches Konstrukt dar, welches den wechselseitig hoch korrelierenden Variablen zugrunde liegt. Für die in dieser Arbeit durchgeführte Faktorenanalyse wurde das Verfahren der *Hauptkomponentenanalyse* (*principal component analysis; PCA*) mit der Rotationsmethode „Varimax mit Kaisernormalisierung“ verwendet.

Für die PCA gibt es mehrere Kennwerte:

- Der *Eigenwert* eines Faktors gibt an, wie viel von der Gesamtvarianz aller Variablen durch diesen Faktor erfasst wird. Faktoren mit Eigenwerten  $< 1$  sind wegen der geringen Varianzaufklärung als unbedeutend zu betrachten. Bei der PCA werden jene Faktoren extrahiert, die einen Eigenwert  $> 1$  besitzen (Kaiser-Guttman-Kriterium).
- Der *Faktorenwert* zeigt an, wie stark die Ausprägung der in einem Faktor zusammengefassten Merkmale bei einer Variablen ist.
- Die *Faktorenladung* entspricht der Korrelation zwischen Faktor und Variablen.
- Die *Kommunalität* einer Variablen gibt an, in welchem Ausmaß die Varianz dieser Variablen durch den Faktor aufgeklärt wird.

Bei der PCA werden Faktoren ermittelt, die voneinander unabhängig sind und sukzessiv die maximale Varianz erklären. Dabei erklärt der erste Faktor den größten Anteil der Gesamtvarianz. Der zweite Faktor erklärt von der verbleibenden Varianz wiederum das Maximum. Die Variablen sind so standardisiert, dass sie die Varianz 1 haben und somit entspricht die aufzuklärende Gesamtvarianz der Anzahl der Variablen.

#### 5.4.1.3. Varianzanalyse

Die *Varianzanalyse* untersucht den Einfluss von einer oder mehreren unabhängigen Variablen auf eine abhängige Variable. Dabei soll die Frage geklärt werden, ob auftretende Mittelwertunterschiede sich mit zufälligen Schwankungen erklären lassen oder einen signifikanten Unterschied bedeuten. Die in dieser Arbeit durchgeführte *einfaktorielle Varianzanalyse* (ANOVA) testet, ob sich die Mittelwerte aus verschiedenen Gruppen signifikant unterscheiden. Voraussetzung für die Durchführung einer Varianzanalyse sind Homogenität der Gruppenvarianzen und Normalverteilung der Fehlerkomponente, welche die Abweichung eines Messwertes vom jeweiligen Stichprobenmittel darstellt.

Der Varianzanalyse geht zunächst eine Homogenitätsprüfung der Gruppenvarianzen mittels Levene-Statistik voraus. Bei einem Signifikanzniveau von  $p > 0,05$  liegt eine Homogenität der Varianzen vor. Um eine Normalverteilung zu bestätigen, wird der *Kolmogorov-Smirnov-Test* angewandt. Eine berechnete Wahrscheinlichkeit von  $p < 0,05$  bedeutet, dass sich die getestete Verteilung signifikant von einer Normalverteilung unterscheidet. Bei einem Signifikanzniveau von  $p > 0,05$  kann dagegen eine Normalverteilung angenommen werden. Führt nun die Varianzanalyse zu einem signifikanten Ergebnis ( $p < 0,05$ ), dann ist davon auszugehen, dass sich mindestens eine der untersuchten Gruppe bezüglich der Mittelwerte von den anderen unterscheidet. Abschließend wird als Post-Hoc-Test die Scheffé-Prozedur angewandt. Dabei kommt es zu einem paarweisen Vergleich der Gruppen. Ziel des Scheffé-Tests ist es, anhand des Vergleichs der Mittelwerte bei signifikanten Unterschieden homogene Untergruppe zu definieren.

Mittels Varianzanalyse soll in der vorliegenden Arbeit überprüft werden, ob sich Studierende der vorklinischen Semester und Studierende klinischer Semester in ihrer

Veranstaltungsbeurteilung signifikant voneinander unterscheiden. Dabei werden die Veranstaltungstypen Vorlesung und Kurse/ Praktika getrennt untersucht. Seminare fanden nur im vierten vorklinischen Semester statt und entfallen somit aufgrund fehlender Vergleichsgruppe für diese Fragestellung. Weiterhin sollen Unterschiede in der Beurteilung der drei Veranstaltungstypen auf Signifikanz geprüft werden.

#### **5.4.1.4. U-Test nach Mann und Whitney**

Des Weiteren ist zu untersuchen, ob die Häufigkeit des Veranstaltungsbesuches einen Einfluss auf die studentische Beurteilung einer Veranstaltung hat. Studierende, die ihre Anwesenheit in der Veranstaltung mit 20% und 40% angeben, werden zu einer Gruppe zusammengefasst. Die zweite Gruppe wird von Studierenden mit einer angegebenen Anwesenheit von 60% bis 100% gebildet. Auch die beiden Extreme, eine Anwesenheit von 20% und 100%, werden miteinander verglichen. Für diese Untersuchungen wird allein der Veranstaltungstyp Vorlesung herangezogen, da die Fragebögen zu den Seminaren und Praktika als Pflichtveranstaltung nicht die Frage nach der Anwesenheitshäufigkeit enthalten.

Der für diese Fragestellung verwendete U-Test nach Mann und Whitney ist ein nonparametrischer Test zum Vergleich von zwei unabhängigen Stichproben. Dieser Test setzt keine Normalverteilung voraus und basiert auf einer gemeinsamen Rangreihe der Werte beider Stichproben. Er untersucht, ob sich die Verteilung der Daten beider Gruppen gleichen (Nullhypothese  $H_0$ ) oder nicht ( $H_1$ ).

#### **5.4.2. Analyse der offenen Fragen**

Zur Auswertung der von den Studierenden getroffenen Aussagen wurde eine Inhaltsanalyse durchgeführt. Theoretischer Leitfaden dabei waren die Arbeiten von Rust (1981) und Rustemeyer (1992). Von 657 Fragebögen konnten 447 (68%) Bögen zur Auswertung der offenen Fragen verwendet werden. Die verbleibenden 210 (32%) Fragebögen enthielten keine Meinungsäußerungen.

#### 5.4.2.1. Entwicklung des Kategoriensystems

Ziel der Inhaltsanalyse ist es, einzelne Teile eines Textes zu ausgewählten, übergreifenden Bedeutungseinheiten, den Kategorien, zuzuordnen. Kernstück einer Inhaltsanalyse ist dabei das Kategoriensystem. In dieser Arbeit wurde es induktiv erstellt, d.h. Textmaterial wurde gesichtet und anhand des Inhaltes entsprechende Kategorien entwickelt. Rustemeyer (1992) konstatiert als generelles Prinzip der Inhaltsanalyse eine intersubjektive Zuordnung der Textteile zu den Bedeutungsaspekten mit dem Ziel einer systematischen Gesamtbeschreibung der Bedeutung eines Textes. Intersubjektivität bedeutet in diesem Zusammenhang, dass verschiedene Beobachter unabhängig von ihren subjektiven Vorstellungen unter Verwendung eines vorliegenden Kategoriensystems bei der Auswertung des Textmaterials zum gleichen Ergebnis kommen und damit eine Codiererübereinstimmung gegeben ist.

Es ist zu erwarten, dass die Antworten der Studierenden sehr heterogen ausfallen und komplexe Textinhalte aufzeigen, die mehrere Themen gleichzeitig ansprechen und somit in unterschiedlichen Kategorien einzuordnen sind. Da das Aufschlüsseln des Textes und die Einordnung der einzelnen Aspekte in Kategorien den subjektiven Verstehensprozessen und der Interpretation des Codierers unterliegen, können unterschiedliche Codierer aufgrund subjektiver Wahrnehmung und Assoziation vorliegendes Textmaterial in verschiedene Kategorien einordnen.

#### 5.4.2.2. Reliabilität des Kategoriensystems

Das Problem der Reliabilität eines Kategoriensystems ist auf zwei Ebenen zu sehen (Rust, 1981), der Formulierungsebene und der Codiererebene:

- *Formulierungsebene:* Die Formulierungen der Kategorien müssen eindeutig sein. Eine Kategorie muss in Bezug auf die Fragestellung letzte Gültigkeit haben und darf sich nicht mit anderen Kategorien überschneiden.
- *Codiererebene:* Die Codierer müssen zuverlässig sein, d.h. die Zahl von Abweichungen soll möglichst gering ausfallen.



Sind die Kategorien nicht klar definiert oder überschneiden sie sich, dann ist keine zuverlässige Zuordnung möglich. Der Spielraum für Interpretationen bleibt groß und es kann Unsicherheit auf Seiten der Codierer bei der Einordnung von Aussagen entstehen. Um die Codiererübereinstimmung zu erhöhen, schlägt Rustemeyer (1992) daher eine Codiererschulung vor. Dabei wird auf Probleme des Kategoriensystems hingewiesen und der Zuordnungsprozess anhand von Beispielen den Codierern anschaulich erklärt. Die Anwendung des Kategoriensystems soll damit bei den unterschiedlichen Codierern, welche auch als Rater bezeichnet werden können, vereinheitlicht werden.

Ein Ziel der Untersuchung des Marburger Evaluationsbogens ist es, ein Kategoriensystem zu entwickeln, welches die sinnvolle Zuordnung der Textinhalte zu bedeutungsübergreifenden Kategorien reliabel ermöglicht. Nach Erstellung des Kategoriensystems soll eine Zufallsauswahl von 100 Bögen aus dem Pool von 447 Fragebögen getroffen und von zwei unabhängigen Personen bearbeitet werden, um anschließend die Codiererübereinstimmung auf zwei Arten zu bestimmen:

- *Intra-Rater-Reliabilität*: Es wird überprüft, wie hoch die Übereinstimmung der Zuordnungen der Meinungsäußerungen zu den Kategorien ist, wenn dieselbe Person zu zwei verschiedenen Zeitpunkten die Kategorisierung durchführt.
- *Inter-Rater-Reliabilität*: Es werden die Zuordnungen der Meinungsäußerungen, welche zwei voneinander unabhängige Personen vorgenommen haben, auf Übereinstimmung untersucht.

Als Übereinstimmungskoeffizient wird Cohens Kappa gewählt. Dieser Koeffizient berücksichtigt die Häufigkeit der allein per Zufall möglichen Übereinstimmungen und korrigiert die beobachteten Übereinstimmungen um diesen Zufallseffekt. Bei der Untersuchung der Interraterübereinstimmung gibt Kappa nicht nur Auskunft über die Reliabilität des Kategoriensystems, sondern auch über dessen Objektivität – dem Grad, in dem die Ergebnisse eines Testes unabhängig vom Untersucher sind. Er ist eine quantitative Maßzahl für die Präzision und Expliztheit eines Kategoriensystems (Rustemeyer, 1992).

### 5.4.2.3. Frequenzanalyse

Die durchzuführende Inhaltsanalyse dient der Darstellung der Häufigkeit bestimmter Thematiken in dem vorliegenden Textmaterial. Dadurch wird ein Überblick über die Daten gewonnen und die Frage geklärt, ob eine Thematik überhaupt genannt wird bzw. wie oft. Die Frequenzanalyse, auch Häufigkeitsanalyse genannt, hat hier eine „Erkundungsfunktion“ (Rustemeyer, 1992).

Der Häufigkeitsanalyse werden die insgesamt 447 Bögen, welche von einem Rater komplett kategorisiert wurden, zugrunde gelegt. Die Kategorien werden ausgezählt und anschließend die absoluten und relativen Häufigkeiten der Kodierungen für die positiven und negativen Kategorien der gesamten Stichprobe berechnet. Anschließend wird die Häufigkeitsanalyse unter Berücksichtigung der Studienabschnitte durchgeführt. Dabei erfolgt eine Trennung in vorklinischen und klinischen Studienabschnitt. Es soll neben den absoluten und relativen Häufigkeiten auch ermittelt werden, ob die Verteilung der Nennungen für die einzelnen Kategorien signifikante Unterschiede in den zwei Gruppen zeigt.

Neben der Frage nach der absoluten und relativen Häufigkeit der Kategorien, stellt sich auch die Frage nach den Beziehungen der Kategorien untereinander. Es soll geklärt werden, ob sich die quantitativen Häufigkeiten zwischen den Kategorien signifikant unterscheiden. Die theoretische Häufigkeitsverteilung wäre eine Gleichverteilung, bei der jede Kategorie mit der gleichen Häufigkeit genannt würde. Da man nicht von einer Normalverteilung der erhobenen Daten in der Inhaltsanalyse ausgehen kann, eignet sich der Chi-Quadrat-Test als nonparametrischer Test zum Vergleich der erhobenen Häufigkeitsverteilung mit der zu erwartenden Häufigkeitsverteilung (Rustemeyer, 1992).

## 6. Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der statistischen Untersuchung des Marburger Fragebogens zur Evaluation des Lehrangebots in der Medizin dargestellt. Nach der Stichprobenbeschreibung folgt eine ausführliche Darstellung der Resultate von Item- und Faktorenanalysen. Anschließend werden die Ergebnisse der Analyse der offenen Fragen und das entwickelte Kategoriensystem vorgestellt.

### 6.1. Stichprobenbeschreibung

Der Fragebogen wurde von den Studierenden der Humanmedizin und Zahnmedizin der Philipps-Universität Marburg im Sommersemester 2002 insgesamt 657-mal anonym ausgefüllt. Davon befanden sich 376 (57,2%) Studierende im vorklinischen Abschnitt, darunter 38 Studierende der Zahnmedizin, und 281 (42,3%) im klinischen Abschnitt ihres Studiums. Unter Berücksichtigung der Ausschlusskriterien für die Analyse der geschlossenen Fragen konnten 60 Fragebögen nicht in die Auswertung der geschlossenen Fragen eingehen. Es verblieben 597 Bögen (90,9%). Zur inhaltlichen Untersuchung der freien Meinungsäußerungen der Studierenden konnten 447 (68,0%) Bögen herangezogen werden. Die 210 verbleibenden Bögen enthielten keine Antworten auf die offenen Fragen.

### 6.2. Ergebnisse zu den Gütekriterien der 12-Item-Skala

In Tabelle 1 wurden die Ergebnisse der psychometrischen Überprüfung der 12-Item-Skala für die drei Veranstaltungstypen zusammengefasst. Die interne Konsistenz der Skala für alle Veranstaltungstypen ist mit Chronbach's alpha zwischen 0,85 bis 0,87 gut. Bei Ermittlung der Trennschärfe zeigt sich allerdings, dass nicht alle Items den angestrebten Wert für den Trennschärfekoeffizienten mit  $r_{it} > 0,5$  erfüllen. Bei Analyse der Fragebögen für Praktika weisen die Items SQ 5 ( $r_{it} = 0,26$ ) und SQ 6 ( $r_{it} = 0,22$ ) ungenügende Trennschärfen auf. Dies gilt für den Veranstaltungstyp Vorlesung neben den Items SQ 5 ( $r_{it} = 0,09$ ) und SQ 6 ( $r_{it} = 0,17$ ) auch für das Item SQ 3 ( $r_{it} = 0,27$ ). Die Elimination des Items SQ 5 würde hier eine Erhöhung der internen Konsistenz von  $\alpha = 0,85$  auf  $\alpha = 0,87$  bewirken. Bei Betrachtung der Ergebnisse für Seminare zeigt sich das gleiche Bild. Der Trennschärfekoeffizient beträgt für SQ 3  $r_{it} = 0,17$ , SQ 5  $r_{it} = 0,03$  und

SQ 6  $r_{it} = 0,13$ . Eine Erhöhung der internen Konsistenz von  $\alpha = 0,86$  auf  $\alpha = 0,88$  würde der Ausschluss von Item SQ 3 erbringen.

**Tabelle 1: Itemanalyse der 12- Item-Skala**

Veranstaltung	Chronbach`s alpha	Trennschärfe		MW	SD
		$r_{min}$	$r_{max}$		
Praktikum	0,87	0,22	0,77	2,76	0,79
Vorlesung	0,85	0,09	0,77	2,60	0,74
Seminar	0,86	0,13	0,78	2,87	0,80

Die Items SQ 3 (*Qualität der Unterrichtsmaterialien*), SQ 5 (*Ausfall der Veranstaltung*) und SQ 6 (*Anwesenheit der Studierenden in der Veranstaltung/ Beitrag der Semesterhilfskräfte*) waren bei allen drei Veranstaltungstypen die Items mit der geringsten Trennschärfe ( $r_{it} < 0,5$ ). Da Items mit geringer Trennschärfe Informationen erbringen, die nicht mit dem Gesamtergebnis übereinstimmen, sind sie aus einem eindimensional angelegten Test zu entfernen. Item PQ 2 (*Didaktik*) dagegen zeigt für Praktika, Vorlesung und Seminar die höchsten Trennschärfen mit  $r_{it}$  zwischen 0,75 und 0,78. Auch die verbleibenden Items zur Strukturqualität SQ 1, SQ 2 und SQ 4 sowie die restlichen Items zur Prozessqualität zeigen hohe Trennschärfen ( $r_{it} > 0,5$ ).

### 6.3. Ergebnisse der Faktorenanalysen

Ziel bei der Konzeption des Fragebogens war jeweils die Eindimensionalität der Skalen zur Struktur- und Prozessqualität. Die Skala zur Strukturqualität sollte dabei nur die Items SQ 1 bis SQ 5 einbeziehen, da Item SQ 6 der Anwesenheitskontrolle von Hilfskräften in Kursen bzw. der Studierenden in der Vorlesung galt. In der ersten Faktorenanalyse wurde Item SQ 6 dennoch berücksichtigt, so dass eine 12-Item-Skala zur Analyse stand.

Jeder Studierende wurde mit der Bewertung für nur eine Veranstaltung einbezogen, da sonst die Unabhängigkeit der Daten nicht gewährleistet wäre. Es ergaben sich für die Faktorenanalyse 173 Fälle „Vorlesung“, 318 Fälle „Kurs/ Praktikum“ und 106 Fälle „Seminar“.

Um die Dimensionalität des Fragebogens zu untersuchen wurde eine Faktorenanalyse (Hauptkomponentenanalyse mit Varimax-Rotation, Abbruchkriterium Eigenwert  $< 1$ ) der 12-Item-Skala durchgeführt. Wie die Ergebnisse zu den Gütekriterien bereits andeuten, liegt der Ausschluss der drei Items SQ 3, SQ 5 und SQ 6 nahe, um eine eindimensionale Skala zur Strukturqualität zu erreichen.

### 6.3.1. Faktorenanalytische Ergebnisse der 12-Item-Skala

Die Faktorenanalyse wurde über die drei Veranstaltungstypen Vorlesung, Seminar sowie Kurse und Praktika zunächst für die 12-Item-Skala durchgeführt. Anschließend folgte die Analyse getrennt für die jeweils aus 6 Items bestehenden Skalen zur Struktur- und Prozessqualität.

#### 6.3.1.1. Faktorenstruktur der Veranstaltungstypen

Die faktorenanalytischen Ergebnisse der 12-Item-Skala über alle Veranstaltungstypen sind in Tabelle 2 zusammengefasst. Die Faktorenanalyse über die 12 Items für den Veranstaltungstyp **Vorlesung** führt zu einer 3-Faktor-Lösung (Eigenwerte 5,32; 1,13 und 1,04), welche 62,4% der Gesamtvarianz erklärt. Die 6 Items zum Lehrprozess laden zusammen mit den Items SQ 1 (*Erläuterung des Lehrkonzepts*), SQ 2 (*Abstimmung von Lehrinhalt und Prüfungsinhalt*) und SQ 4 (*Vorbereitung des Dozenten*) auf den ersten Faktor (FA 1). Der zweite Faktor (FA 2) beinhaltet die Items SQ 3 (*Qualität der Unterrichtsmaterialien*) und SQ 5 (*Ausfall der Veranstaltung*). Auf den dritten Faktor (FA 3) lädt lediglich SQ 6, welches die Häufigkeit der Vorlesungsbesuche erfasst.

Auch bei der Auswertung der **Seminare** erbringt die Analyse eine Dreidimensionalität des Fragebogens (Eigenwerte 5,29; 1,21 und 1,11), welche 63,4% der Gesamtvarianz erklärt. Auf den ersten Faktor laden wieder alle Prozessitems und zusätzlich die drei Items SQ 1, SQ 2 und SQ 4. Die Items SQ 3 und SQ 5 stellen die zweite Dimension dar und auf den dritten Faktor lädt allein Item SQ 6.

**Tabelle 2: Hauptkomponentenanalyse mit Varimax-Rotation**

VL	FA 1	FA 2	FA 3	SEM	FA 1	FA 2	FA 3	PR	FA 1	FA 2	FA 3
<u>Skala 1</u>				<u>Skala 1</u>				<u>Skala 1</u>			
SQ 1	<b>0,79</b>	-0,06	-0,01	SQ 1	<b>0,77</b>	0,10	0,15	SQ 1	<b>0,60</b>	0,50	0,15
SQ 2	<b>0,74</b>	0,07	0,09	SQ 2	<b>0,54</b>	-0,06	0,31				
SQ 4	<b>0,77</b>	0,05	-0,14	SQ 4	<b>0,78</b>	-0,07	-0,07	SQ 4	<b>0,69</b>	0,32	-0,03
PQ 1	<b>0,75</b>	-0,13	0,06	PQ 1	<b>0,80</b>	0,16	-0,22	PQ 1	<b>0,60</b>	0,43	0,31
PQ 2	<b>0,85</b>	-0,05	-0,01	PQ 2	<b>0,87</b>	-0,28	0,08	PQ 2	<b>0,72</b>	0,33	0,31
PQ 3	<b>0,77</b>	0,34	-0,05	PQ 3	<b>0,85</b>	0,00	0,05	PQ 3	<b>0,79</b>	0,14	0,12
PQ 4	<b>0,83</b>	0,08	0,11	PQ 4	<b>0,68</b>	-0,15	0,43	PQ 4	<b>0,70</b>	0,33	-0,02
PQ 5	<b>0,66</b>	0,23	0,31	PQ 5	<b>0,74</b>	0,23	0,07	PQ 5	<b>0,60</b>	0,38	0,12
PQ 6	<b>0,63</b>	-0,18	0,10	PQ 6	<b>0,73</b>	0,12	0,08	PQ 6	<b>0,78</b>	-0,75	0,02
<u>Skala 2</u>				<u>Skala 2</u>				<u>Skala 2</u>			
SQ 3	0,34	<b>0,67</b>	-0,22	SQ 3	0,14	<b>0,76</b>	0,06	SQ 2	0,42	<b>0,57</b>	-0,06
SQ 5	0,27	<b>0,58</b>	-0,13	SQ 5	-0,04	<b>0,72</b>	-0,04	SQ 3	0,09	<b>0,75</b>	0,08
<u>Skala 3</u>				<u>Skala 3</u>				<u>Skala 3</u>			
SQ 6	0,12	-0,04	<b>0,93</b>	SQ 6	0,01	0,06	<b>0,91</b>	SQ 5	-0,11	0,34	<b>0,78</b>
								SQ 6	0,27	-0,28	<b>0,66</b>

Legende: VL = Vorlesung, SEM = Seminar, PR = Kurse und Praktika. FA= Faktorladung des jeweiligen Items zur Strukturqualität (SQ) bzw. Prozessqualität (PQ): FA 1 = Ladung auf den ersten Faktor, FA 2 = Ladung auf den zweiten Faktor, FA 3 = Ladung auf den dritten Faktor.

Für **Kurse und Praktika** bestätigt sich ebenfalls die Dreidimensionalität des Fragebogens (Eigenwerte 5,17; 1,20 und 1,02), welche 61,52% der Gesamtvarianz erklärt. Wiederum laden die Prozessitems auf den ersten Faktor, jedoch im Gegensatz zur Beurteilung der Vorlesung und Seminare laden nur zwei Strukturitems auf den ersten Faktor: SQ 1 und SQ 4. Die zweite Dimension erfasst die Items SQ 2 und SQ 3. SQ 5 und SQ 6 laden gemeinsam auf einen dritten Faktor.

Die stets unter dem ersten Faktor zusammengeschlossenen Items der Prozessqualität weisen mit Faktorladungen zwischen 0,59 und 0,86 mittlere bis hohe Korrelation auf. Item PQ 6 (*Prüfungs- und Praxisrelevanz des Unterrichts*) stellt dabei mit Ladungen von 0,59 bis 0,74 unter den Prozessitems das Item der geringsten Korrelation dar. Die höchste Korrelation zeigt Item PQ 2 (*Didaktik*) mit Faktorladungen > 0,84 über alle Veranstaltungen. Die Strukturitems SQ 1, SQ 2 und SQ 4 korrelieren bei Faktorladungen > 0,7 für Vorlesungen und Seminare sowie Faktorladungen > 0,6 für

SQ 1 und SQ 4 bei Kursen und Praktika gut mit dem ersten Faktor. Die verbleibenden Strukturitems verteilen sich auf den zweiten und dritten Faktor, wobei einzig das Item SQ 6 mit Werten zwischen 0,67 bis 0,82 über alle Veranstaltungen auf den letzten Faktor lädt. Dieses Item war als Kontrollinstrument für die Anwesenheit der Studierenden in der Veranstaltung bzw. der Hilfskräfte in den Praktika angelegt und galt damit nicht direkt der Messung der Strukturqualität. Erwartungsgemäß zeigt sich keine Korrelation zu den anderen Items.

### 6.3.1.2. Faktorenstruktur der Skalen zur Struktur- und Prozessqualität

Die 6 Items zur **Prozessqualität** laden bei Betrachtung der verschiedenen Unterrichtsveranstaltungen immer auf einen gemeinsamen Faktor, weshalb anzunehmen ist, dass es sich bei dem Teilfragebogen zur Prozessqualität der Lehre um einen eindimensionalen Fragebogen handelt. Die unter Einbezug der 6 Items zur Prozessqualität durchgeführten Faktorenanalysen bestätigen diese Annahme. Die durch diese eine Dimension erklärte Gesamtvarianz beträgt für Praktika 59,4% (Eigenwerte 3,564 und 0,719), für Vorlesungen 59,81% (Eigenwerte 3,589 und 0,76) und für Seminare 64,59% (Eigenwerte 3,875 und 0,66). Die einzelnen Items laden mit Werten zwischen 0,64 bis 0,87 auf den extrahierten Faktor. Item PQ 2 weist dabei stets die höchste Ladung auf.

Die 6-Item-Skala zur Messung der **Strukturqualität** zeigt im Gegensatz dazu keine Eindimensionalität, auch wenn die drei Items SQ 1 (*Erläuterung des Lehrkonzeptes*), SQ 2 (*Abstimmung von Lehrinhalten und Prüfungsinhalten*) und SQ 4 (*Vorbereitung des Dozenten*) das gleiche Konstrukt zu messen scheinen wie die Prozessitems. Sie laden für Vorlesung und Seminar auf einen gemeinsamen Faktor mit den Items zur Prozessqualität. Die neun Items SQ 1, SQ 2, SQ 4 und PQ 1 bis PQ 6 könnten zu einem Gesamtscore zusammengefasst werden. Voraussetzung dafür wäre die Eindimensionalität dieser 9-Item-Skala. Die Notwendigkeit der Elimination der Items SQ 3, SQ 5 und SQ 6, die bereits durch ihre geringe Trennschärfe auffielen, hat sich mit den Ergebnissen der Faktorenanalyse bestätigt.

### 6.3.2. Faktorenanalytische Ergebnisse der 9-Item-Skala

Da sich bei allen Veranstaltungstypen eine Dreidimensionalität in der Faktorenanalyse der 12-Item-Skala zeigte, wurden die Berechnungen unter Ausschluss der Strukturitems SQ 3, SQ 5 und SQ 6 wiederholt. Die Faktorenanalyse der daraus resultierenden 9-Item-Skala bestätigt die Eindimensionalität der Skala für alle drei Veranstaltungstypen. Die durch den extrahierten Faktor erklärte Gesamtvarianz beträgt für Praktika 53,6% (Eigenwerte 4,82 und 0,80), für Vorlesungen 57,3% (Eigenwerte 5,15 und 0,78) und für Seminare 58,17% (Eigenwerte 5,24 und 0,89). Die Items laden auf den extrahierten Faktor mit Werten zwischen 0,57 bis 0,86. Der extrahierte Faktor kann als „Index der allgemeinen Lehrqualität“ bezeichnet werden.

### 6.4. Ergebnisse zu den Gütekriterien der 9-Item- Skala

Die Überprüfung der Gütekriterien der 9-Item-Skala wurde ebenfalls getrennt nach Veranstaltungstypen durchgeführt. Die Ergebnisse sind in Tabelle 3 zusammengefasst.

**Tabelle 3: Itemanalyse der 9-Item-Skala**

Veranstaltung	Chronbach`s alpha	Trennschärfe		MW	SD
		$r_{\min}$	$r_{\max}$		
Praktikum	0,89	0,52	0,78	2,86	0,92
Vorlesung	0,90	0,55	0,79	2,61	0,87
Seminar	0,91	0,49	0,81	2,92	0,99

Bei dem Veranstaltungstyp **Praktikum** zeigte sich für die 9-Item-Skala eine gute interne Konsistenz mit Chronbach`s  $\alpha = 0,89$ . Die Items SQ 2 mit  $r_{it} = 0,52$  und PQ 6 mit  $r_{it} = 0,53$  wiesen die geringste Trennschärfe auf. Das Item mit der höchsten Trennschärfe ist PQ 2. Die Ergebnisse der psychometrischen Überprüfung der 9-Item-Skala für den Veranstaltungstyp **Vorlesung** zeigten mit Chronbach`s  $\alpha = 0,90$  eine hohe interne Konsistenz und die Trennschärfe der Items liegt zwischen  $r_{it} = 0,55$  und  $r_{it} = 0,79$ . Dabei weisen die Items PQ 5 mit  $r_{it} = 0,56$  und PQ 6 mit  $r_{it} = 0,55$  die geringste, Item PQ 2 erneut die höchste Trennschärfe auf. Auch für **Seminare** zeigt sich mit  $\alpha = 0,91$  eine hohe interne Konsistenz der Skala. Das Item der geringsten Trennschärfe ist in diesem Fall SQ 2 („Abstimmung von Lehrinhalten und Prüfungsinhalten“) mit  $r_{it} = 0,49$ . Wieder ist Item PQ 2 mit  $r_{it} = 0,81$  das Item mit der höchsten Trennschärfe.



Die 9-Item-Skala zur Messung des Konstruktes „Index der Allgemeine Lehrqualität“ zeigt über alle Veranstaltung eine hohe interne Konsistenz. Die ausgewählten Items erfüllen den Anspruch auf hohe Trennschärfe mit  $r_{it} > 0,5$ . Dabei zeigte sich, dass Item PQ 2, welches Bezug auf die Didaktik einer Veranstaltung nimmt, die höchste Trennschärfe über die drei Veranstaltungstypen aufweist. Im Gegensatz dazu lag der ermittelte Trennschärfekoeffizient für die Items SQ 2 und PQ 6 über die Veranstaltungstypen mit Werten für  $r_{it}$  zwischen 0,49 und 0,55 am niedrigsten.

## **6.5. Ergebnisse der Varianzanalyse**

Im folgenden Abschnitt werden die Ergebnisse der einfaktoriellen Varianzanalyse (ANOVA) vorgestellt. Als Voraussetzung für eine einfaktorielle Varianzanalyse gelten Normalverteilung und Homogenität der Gruppenvarianzen. Es war zu untersuchen, ob sich die Mittelwerte zwischen den Veranstaltungstypen bzw. den Studienabschnitten unterscheiden und sich damit ein Hinweis auf einen möglichen Einfluss von Studiendauer und Motivation auf die Veranstaltungsevaluation ergibt. Den Berechnungen wurden die Mittelwerte über die 9-Item-Skala zugrunde gelegt.

### **6.5.1. Varianzanalyse für die Veranstaltungstypen**

Eine Normalverteilung der Varianzen ist anzunehmen und die Levene-Statistik bestätigt die Homogenität der Varianzen (Levene-Statistik = 2,15;  $df = 2$ ,  $p > 0,05$ ). Die Varianzanalyse zeigt, dass ein signifikanter Unterschied ( $F_{2,594} = 5,47$ ;  $p < 0,05$ ) bezüglich der Mittelwerte zwischen den Veranstaltungstypen existiert. Es wurden mittels Scheffé-Test zwei Untergruppen definiert. Das 95%-Konfidenzintervall bei Vergleich der Veranstaltungstypen Praktikum und Seminar schließt die Null ein und es daher ist davon auszugehen, dass sie sich nicht unterscheiden. Bei Vergleich mit dem Veranstaltungstyp Vorlesung liegt die Null nicht im Intervall (vgl. Tabelle 4) und die Nullhypothese muss in diesem Fall abgelehnt werden. Es liegt somit keine Gleichheit zwischen der Untergruppe bestehend aus Kursen/ Praktika und Seminar und der Untergruppe Vorlesung vor.

**Tabelle 4: Ergebnisse Post-Hoc-Test (Scheffé-Prozedur) Veranstaltungstypen**

	Typ I	Typ II	Mittlere Differenz (I - II)	Standard -fehler	Signifikan z	95% Konfidenz-intervall	
						Untergrenze	Obergrenze
Scheffé-Prozedur	<b>PR</b>	<b>SEM</b>	-0,06	0,10	0,83	-0,32	0,20
		<b>VL</b>	0,25	0,09	0,15	<b>0,04</b>	<b>0,47</b>
	<b>SE</b>	<b>VL</b>	0,32	0,11	0,02	<b>0,04</b>	<b>0,60</b>
	<b>M</b>						

Legende: PR = Praktikum; SEM = Seminar; VL = Vorlesung

Der Vorlesung als erste Untergruppe stehen die Veranstaltungstypen Praktika und Seminar als zweite homogene Untergruppe gegenüber. Während die Vorlesungen (MW = 2,61; SD = 0,99) signifikant besser beurteilt wurden, unterscheiden sich Praktika (MW = 2,86; SD = 0,92) und Seminare (MW = 2,92; SD = 0,87) in ihrer Bewertung kaum voneinander.

### 6.5.2. Varianzanalyse für die Studienabschnitte

Für **Vorlesungen** zeigen sich in den einzelnen Studienabschnitten keine signifikanten Unterschiede der Mittelwerte ( $F_{2,170} = 2,27$ ;  $p > 0,05$ ). Der Scheffé-Test würde somit nur eine homogene Gruppe ermitteln, in der die drei Studienabschnitte enthalten wären. Für die Bewertungen der **Kurse und Praktika** ergibt die Varianzanalyse dagegen ein signifikantes Ergebnis ( $F_{2,315} = 18,95$ ;  $p < 0,05$ ). Mittels Scheffé-Prozedur werden vorklinischer und erster klinischer Abschnitt zu einer Untergruppe zusammengefasst und dem zweiten klinischen Abschnitt gegenüber gestellt. Während bei Vergleich der Vorklinik mit dem ersten klinischen Abschnitt von Gleichheit auszugehen ist, besteht ein signifikanter Unterschied zwischen dieser ersten Gruppe und der Gruppe des zweiten klinischen Abschnittes. Das 95%-Konfidenzintervall schließt die Null jeweils nicht mit ein und von Gleichheit zwischen den Gruppen kann folglich nicht ausgegangen werden (vgl. Tabelle 5).

Studierende des zweiten klinischen Abschnittes bewerteten Kurse und Praktika anscheinend besser als Studierende der Vorklinik und des ersten klinischen Abschnittes (vgl. Tabelle 6). Die der Varianzanalyse für Praktika vorausgehende Überprüfung auf

Homogenität der Varianzen zeigte allerdings, dass sich die Varianzen in den Gruppen signifikant unterscheiden (Levene-Statistik = 6,153;  $df = 2$ ;  $p < 0,05$ ). Damit ist der F-Test unzuverlässig und ebenso wie die Einzelvergleiche nach Scheffé mit Vorsicht zu interpretieren.

**Tabelle 5: Ergebnisse Post-Hoc-Test (Scheffé-Prozedur) Studienabschnitte**

	Typ I	Typ II	Mittlere Differenz (I - II)	Standardfehler	Signifikanz	95% Konfidenzintervall	
						Untergrenze	Obergrenze
Scheffé-Prozedur	<b>VK</b>	<b>1</b>	0,25	0,12	0,11	-0,04	0,55
		<b>2</b>	0,82	0,13	0,00	<b>0,49</b>	<b>1,10</b>
	<b>1</b>	<b>2</b>	0,56	0,16	0,02	<b>0,18</b>	<b>0,95</b>

Legende: VK = Vorklinischer Abschnitt; 1 = 1. klinischer Abschnitt; 2 = 2. klinischer Abschnitt.

**Tabelle 6: Zusammenfassung Vorlesung/Praktikum**

Abschnitt	Vorlesung			Praktikum		
	N	MW	SD	N	MW	SD
Vorklinik	47	2,67	0,84	190	3,06	0,84
1. klin. Abschnitt	73	2,72	0,96	73	2,81	1,03
2. klin. Abschnitt	53	2,39	0,74	55	2,24	0,70

Legende: N = Anzahl der einfließenden Urteile, MW = Mittelwert über die 9-Item-Skala, SD = Standardabweichung.

## 6.6. Ergebnisse des U-Testes nach Mann und Whitney

Um die Frage zu beantworten, ob die Häufigkeit des Veranstaltungsbesuches einen Einfluss auf das Urteil der Studierenden haben könnte, wurde der U-Test nach Mann und Whitney durchgeführt. Es wurde der Veranstaltungstyp Vorlesung untersucht, da der entsprechende Fragebogen mit Item SQ 6 die Anwesenheitshäufigkeit der Studierenden erfasst. Den Berechnungen wurden die Mittelwerte über die 9-Item-Skala zugrunde gelegt.

Es muss festgestellt werden, dass der U-Test nicht zu einem signifikanten Ergebnis führt, weder für den Vergleich zwischen den Gruppen mit einem angegebenen Besuch

von 20% und 100% ( $U = 254,5$ ;  $p = 0,19$ ) noch für eine Vergleich der Gruppen mit einem Besuch von  $< 50\%$  und  $> 50\%$  ( $U = 2914,0$ ;  $p = 0,24$ ). Von den 173 Studierenden, die mit einem Urteil über die Vorlesung ( $MW_{\text{gesamt}} = 2,6$ ;  $SD = 0,87$ ) eingegangen sind, gaben 31 einen Besuch von 100% ( $MW_{100} = 2,45$ ;  $SD = 0,94$ ) und 21 Studierende einen Besuch von 20% ( $MW_{20} = 2,7$ ;  $SD = 0,95$ ) an. Insgesamt besuchten 117 Studierende die Veranstaltungen mit einer Häufigkeit  $> 50\%$  ( $MW_{>50} = 2,56$ ;  $SD = 0,83$ ), 56 mit einer Häufigkeit von  $< 50\%$  ( $MW_{<50} = 2,70$ ;  $SD = 0,95$ ). Bei einem Vergleich der Mittelwerte zeigt sich, dass die vergebenen „Noten“ zwar bei häufigem Vorlesungsbesuch besser scheinen als bei einer als selten angegebenen Anwesenheit, allerdings ist dieser Unterschied zu gering, um zu einem signifikanten Ergebnis zu führen.

## **6.7. Ergebnisse der Analyse der offenen Fragen**

Zur Auswertung der sehr heterogen ausgefallenen Antworten auf die offenen Fragen ist eine Kategorienliste erstellt worden, die es erlaubte, die vielfältigen Meinungsäußerungen inhaltlich einzuordnen. Anschließend wurde die Güte dieser Kategorienliste überprüft.

### **6.7.1. Ergebnisse der Kategorienbildung**

Es wurden folgende 5 Hauptkategorien entwickelt, die sich sowohl für die Klassifikation positiver als auch negativer Meinungsäußerungen eignen:

1. Organisation des Unterrichts
2. Dozent als Unterrichtender und als Person
3. Didaktik
4. Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz
5. Medieneinsatz, Bild- und Arbeitsmaterialien

Diese Hauptkategorien wurden weiter untergliedert, um die Meinungsäußerungen differenzierter darzustellen. Es ergaben sich dabei 24 Unterkategorien für die positiven und 34 für die negativen Äußerungen. Die Kategorienliste mit Erläuterungen zu den einzelnen Kategorien und Subkategorien ist im Anhang B einzusehen.

### 6.7.2. Ergebnisse zu den Gütekriterien der Kategorienliste

Um die Reliabilität der Kategorisierung anhand der erstellten Kategorienliste zu überprüfen, kamen zwei Verfahren zur Anwendung.

1. Intra-rater Reliabilität: Es wird überprüft, wie hoch die Übereinstimmung der Zuordnungen der Meinungsäußerungen zu den Kategorien ist, wenn dieselbe Person zu zwei verschiedenen Zeitpunkten die Kategorisierung durchführt
2. Inter-rater Reliabilität: Es wird das Ausmaß bestimmt, in dem zwei voneinander unabhängige Personen die Antworten denselben Antwortkategorien zuordnen.

Beiden Verfahren wurde die Kategorisierung der freien Meinungsäußerungen von jeweils 100 zufällig ausgewählten Fragebögen zugrunde gelegt. Da die Fragebögen auf die Veranstaltungen des jeweiligen Semesters Bezug nehmen, welche in ihrer Anzahl variieren und in den Bögen Meinungsäußerungen oft nur zu einer Auswahl von Veranstaltungen enthalten sind, ergibt sich eine unterschiedliche Anzahl beurteilter Veranstaltungen pro Bogen. In den 100 zufällig ausgewählten Bögen zur Bestimmung der Intra-rater Reliabilität wurde insgesamt 454 mal eine Veranstaltung beurteilt. Die 100 Bögen zur Ermittlung der Inter-rater Reliabilität enthielten 761 Veranstaltungsurteile. Cohens Kappa wurde als Reliabilitätsindex gewählt. Die Berechnung erfolgte für die einzelnen Sub- und Hauptkategorien über eine Vierfeldertafel. Die Tabellen 7 und 8 zeigen die Ergebnisse der Intra- und Inter-rater Reliabilität für die Hauptkategorien. Die Übereinstimmungen für die jeweiligen Subkategorien werden im Anhang C und D in den Tabellen 11 und 12 dargestellt.

#### 6.7.2.1. Ergebnisse der Intra-Rater-Reliabilität

Die Güte der Übereinstimmung ist mit Ergebnissen von Cohens Kappa zwischen 0,61 bis 0,80 als „substantial“ und mit Werten zwischen 0,81 und 1,00 als „almost perfect“ zu werten (Rustemeyer, 1992). Cohens Kappa nimmt für die positiven Hauptkategorien Werte zwischen 0,69 und 0,93 ein, bei einem Gesamtkappa von 0,80. Für die negativen Hauptkategorien ergeben sich Werte zwischen 0,62 und 0,77 mit einem Gesamtkappa

von 0,73. Werden alle Kategorien gemeinsam betrachtet, so beträgt Cohens Kappa 0,76. Damit zeigt die sich eine gute bis sehr gute Übereinstimmung. Bei Betrachtung der Ergebnisse zu den Subkategorien erstreckt sich das Spektrum von fehlender bis vollkommener Übereinstimmung. Kappa nimmt für die positiven Subkategorien Werte zwischen 0,61 und 0,91 und für die negativen Subkategorien Werte zwischen 0,30 und 0,95 ein.

***Tabelle 7: Intra-Rater-Reliabilität der Hauptkategorien***

<b>Positive Hauptkategorien</b>	<b>0/0</b>	<b>0/1</b>	<b>1/0</b>	<b>1/1</b>	<b>Cohens Kappa</b>
<b>+1: Organisation des Unterrichts</b>	447	1	1	5	0,83
<b>+2: Dozent als Unterrichtender und als Person</b>	298	35	22	99	0,69
<b>+3: Didaktik</b>	380	12	15	47	0,74
<b>+4: Themen und Inhalte</b>	305	22	6	121	0,85
<b>+5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>	446	1	-	7	0,93
<b>Summe</b>	1876	71	44	279	<b>0,80</b>
<b>Negative Hauptkategorien</b>					
<b>-1: Organisation des Unterrichts</b>	305	23	20	106	0,77
<b>-2: Dozent als Unterrichtender und als Person</b>	351	30	19	54	0,62
<b>-3: Didaktik</b>	282	33	18	121	0,74
<b>-4: Themen und Inhalte</b>	380	27	5	42	0,69
<b>-5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>	434	3	7	10	0,66
<b>Summe</b>	1752	116	69	518	<b>0,73</b>
<b>Gesamt</b>	3628	187	113	615	<b>0,76</b>

Legende: Vierfeldertafel jeweils für positive und negative Hauptkategorien: 0/0 - beide Rater hielten die Kategorie bei einer Aussage für nicht gegeben. 1/0 und 0/1 – jeweils nur einer der Rater hat eine Aussage dieser Kategorie zugeordnet. 1/1 – beide Rater stimmen mit ihrer Zuordnung überein. Summe = Gesamtkappa der jeweiligen Hauptkategorie. Gesamt = Kappa über alle Hauptkategorien, bezogen auf eine Zufallsauswahl von 100 Fragebögen. Vgl. Anhang C Tabelle 11: Intra-Rater-Reliabilität der Subkategorien.

### 6.7.2.2. Ergebnisse der Inter-Rater-Reliabilität

Die Tabelle 8 zeigt die Ergebnisse zur Inter-Rater-Reliabilität der Hauptkategorien. Cohens Kappa fällt mit 0,83 für die positiven und 0,77 für die negativen Hauptkategorien und Betrachtung aller Kategorien mit 0,80 höher aus als bei Überprüfung der Intra-Rater-Reliabilität. Für die Subkategorien der Interraterübereinstimmung (vgl. Anhang D Tabelle 12) ergeben sich neben völliger Übereinstimmung und Nichtübereinstimmung für Cohens Kappa Werte zwischen 0,40 und 0,93 sowohl für die positiven als auch die negativen Subkategorien.

**Tabelle 8.: Inter-Rater-Reliabilität der Hauptkategorien**

Positive Hauptkategorien	0/0	0/1	1/0	1/1	Cohens Kappa
<b>+1: Organisation des Unterrichts</b>	750	1	2	8	0,84
<b>+2: Dozent als Unterrichtender und als Person</b>	538	42	21	160	0,78
<b>+3: Didaktik</b>	633	26	18	84	0,76
<b>+4: Themen und Inhalte</b>	541	17	8	195	0,92
<b>+5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>	704	22	2	33	0,72
<b>Summe</b>	3166	108	51	480	<b>0,83</b>
Negative Hauptkategorien					
<b>-1: Organisation des Unterrichts</b>	593	30	24	114	0,77
<b>-2: Dozent als Unterrichtender und als Person</b>	634	24	15	88	0,79
<b>-3: Didaktik</b>	375	76	55	255	0,65
<b>-4: Themen und Inhalte</b>	661	26	8	66	0,77
<b>-5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>	702	7	4	48	0,89
<b>Summe</b>	2965	163	106	571	<b>0,77</b>
<b>Gesamt</b>	6131	271	157	1051	<b>0,80</b>

Legende: Vierfeldertafel für positive und negative Hauptkategorien: 0/0 - beide Rater hielten die Kategorie bei einer Aussage für nicht gegeben. 1/0 und 0/1 – jeweils nur einer der Rater hat eine Aussage dieser Kategorie zugeordnet. 1/1 – beide Rater stimmen mit ihrer Zuordnung überein. Summe = Gesamtkappa der jeweiligen Hauptkategorie. Gesamt = Kappa über alle Hauptkategorien, bezogen auf eine Zufallsauswahl von 100 Fragebögen.

### 6.7.3. Ergebnisse der Häufigkeitsanalyse

Die 447 Bögen, welche Meinungsäußerungen enthielten, wurden von einem Rater komplett kategorisiert und als gesamte Stichprobe sowie getrennt in vorklinischen und klinischen Abschnitt auf die absoluten und relativen Häufigkeiten der Kategoriennennung untersucht.

#### 6.7.3.1. Frequenzanalyse der gesamten Stichprobe

Die Antworten auf die offenen Fragen wurden von einem Rater vollständig kategorisiert. Den Meinungsäußerungen wurde insgesamt 3752 mal eine Kategorien zugeordnet. Dabei wurden 1620 (43,1%) positive und 2134 (56,9%) negative Kategorien vergeben. Die Ergebnisse der Kodierungen für die Hauptkategorien sind in den Tabellen 9 und 10 getrennt nach positiven und negativen Kategorien zusammengefasst. Die Kodierungen für die positiven und negativen Subkategorien können im Anhang E Tabelle 13 eingesehen werden.

**Tabelle 9: Frequenzanalyse der positiven Hauptkategorien**

Positive Hauptkategorien	Häufigkeit	
	absolut	relativ in %
<b>+1: Organisation des Unterrichts</b>	30	<b>1,8</b>
<b>+2: Dozent als Unterrichtender und als Person</b>	559	34,5
<b>+3: Didaktik</b>	322	19,9
<b>+4: Themen und Inhalte</b>	638	<b>39,4</b>
<b>+5: Medieneinsatz, Bild- und Arbeitsmaterialien</b>	71	4,4
<b>Gesamt</b>	<b>1620</b>	<b>100</b>

Für die positiven Hauptkategorien reichen die relativen Häufigkeiten von 1,8% für Kategorie +1 (*Organisation des Unterrichts*) bis 39,4% für Kategorie +4 (*Themen und Inhalte*). Die am häufigsten vorkommenden Subkategorien waren +2.1.3. (*Dozent gibt effektiven Unterricht*) mit einer relativen Häufigkeit von 19,6% und +2.2.1. (*Dozent ist motiviert*) mit 6,2% für die Hauptkategorie +2 (*Dozent als Unterrichtender und als Person*). Für die sich auf Didaktik beziehende Hauptkategorie +3 wurden die Subkategorien +3.1. (*übersichtlicher und systematischer Aufbau der Veranstaltung*) und



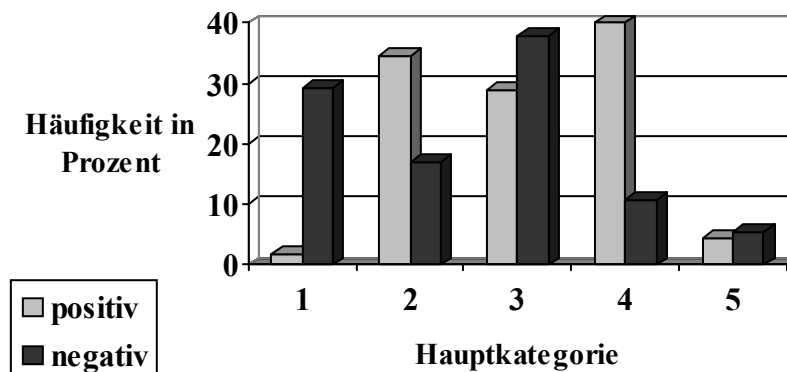
+3.2. (*Veranstaltung verständlich und praxisnah*) mit 7,1% und 6,7% relativ am häufigsten genannt (vgl. Anhang E Tabelle 13). In der Kategorien +1 (*Organisation des Unterrichts*) und +5 (*Medieneinsatz, Bild- und Arbeitsmaterialien*) zeigen die Subkategorie +1.2. (*zeitliche Aspekte*) mit 1,3% und +5.3. (*gute Skripte, Handouts*) mit 2,8% die für die jeweilige Hauptkategorie häufigste Nennung.

Bei den negativen Hauptkategorien (vgl. Tabelle 10) zeigen sich relative Häufigkeiten von 5,4% für die Kategorie –5 (*Medieneinsatz, Bild- und Arbeitsmaterialien*) bis 37,8% für die Kategorie –3 (*Didaktik*). Bei der jeweiligen Hauptkategorie zeigen sich für folgende Subkategorien die häufigsten Nennungen: für Kategorie –1 (*Organisation des Unterrichts*) Subkategorie –1.5. (*fehlende personelle Kontinuität des Dozenten*) mit einer relativen Häufigkeit von 6.5%, Kategorie –2 (*Dozent als Unterrichtender und als Person*) –2.1.3. (*Dozent gestaltet den Unterricht nicht effektiv*) mit 5.2%, Kategorie –3 (*Didaktik*) –3.1. (*Veranstaltung unübersichtlich und unsystematisch*) mit 10,4% und für die Kategorie –5 (*Medieneinsatz, Bild- und Arbeitsmaterialien*) Subkategorie –5.3.1. (*fehlende Skripte*) mit 1,5%.

**Tabelle 10.: Frequenzanalyse der negativen Hauptkategorien**

Negative Hauptkategorien	Häufigkeit	
	absolut	relativ in %
<b>-1: Organisation des Unterrichts</b>	624	29,2
<b>-2: Dozent als Unterrichtender und als Person</b>	363	17,0
<b>-3: Didaktik</b>	806	<b>37,8</b>
<b>-4: Themen und Inhalte</b>	225	10,6
<b>-5: Medieneinsatz, Bild- und Arbeitsmaterialien</b>	116	<b>5,4</b>
<b>Gesamt</b>	<b>2134</b>	<b>100</b>

Als Ergebnis der Frequenzanalyse ist festzuhalten, dass die einzelnen Haupt –und Subkategorien sehr unterschiedlich besetzt sind. Von den Studierenden wurden negative Aspekte häufiger genannt als positive. Dabei überwog die Kritik an der Didaktik (37,8%), speziell an dem unübersichtlichen, unsystematischen Aufbau der Veranstaltung durch die häufige Nennung der Subkategorie –3.1. (10,4%). In Abbildung 4 werden die Ergebnisse der Frequenzanalyse für die positiven und negativen Hauptkategorien zusammenfassend veranschaulicht.



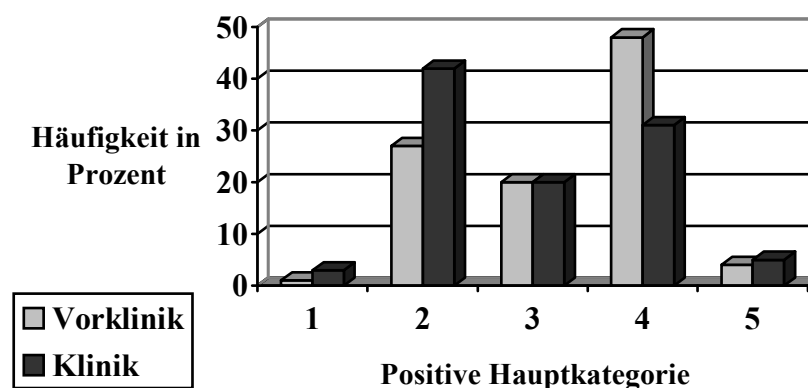
*Abbildung 4: Häufigkeitsanalyse der gesamten Stichprobe*

Anhand der Graphik lässt sich bereits feststellen, dass sowohl für die positiven als auch die negativen Hauptkategorien keine Gleichverteilung, bei der jede Kategorie mit 20% vertreten wäre, vorliegt. Die Ergebnisse des Chi-Quadrat-Testes bestätigen diese Annahme. Die  $\chi^2$ -Werte von 939,1 für positive und 759,3 für negative Kategorien bei einem Freiheitsgrad von 4 sind mit  $p < 0,01$  höchst signifikant. Die Nullhypothese muss folglich abgelehnt werden, d.h. es liegt keine Gleichverteilung vor. Die Kategorien +2 und +4 sind überdurchschnittlich, die Kategorien +1 und +5 unterdurchschnittlich bei den positiven Hauptkategorien vertreten. Bei den negativen Kategorien sind es die Aspekte zur Organisation (Kategorie –1) und Didaktik (Kategorie –3), welche signifikant häufiger genannt wurden. Die verbleibenden negativen Hauptkategorien wurden unterdurchschnittlich repräsentiert.

### 6.7.3.2. Frequenzanalyse getrennt für Vorklinik und Klinik

253 (56,6%) Fragebögen des vorklinischen und 194 (43,4%) Bögen des klinischen Studienabschnitts wurden gewertet. Die Ergebnisse der Auszählung sind in den Abbildungen 5 und 6 veranschaulicht. Im Anhang F Tabelle 14 sind die Daten für die positiven und in Tabelle 15 für die negativen Hauptkategorien dargestellt. Die Tabellen 16 und 17 im Anhang G enthalten die Ergebnisse der Frequenzanalyse der Subkategorien getrennt in Vorklinik und Klinik. Von den 1620 vergebenen positiven Kategorien entfielen 814 (50,2%) auf Aussagen der vorklinischen Semester und 806

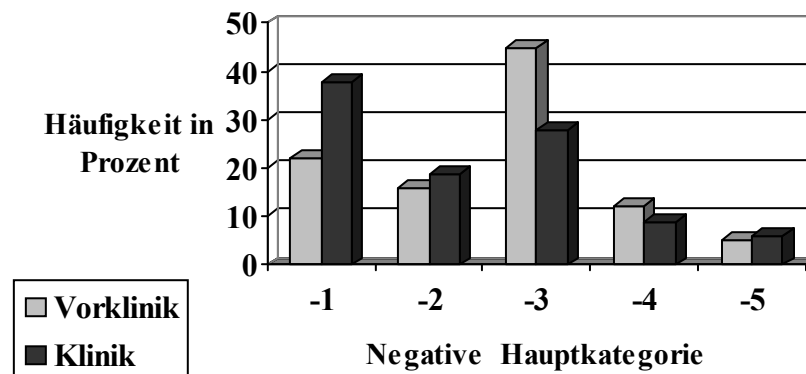
(49,8%) auf Aussagen Studierender im klinischen Abschnitt. Bei Vergleich des vorklinischen mit dem klinischen Studienabschnitt ist festzuhalten, dass die Hauptkategorien +1 und +5 jeweils sehr selten, die Kategorie +3 nahezu gleichhäufig und die Kategorien +2 und +4 am häufigsten zugeordnet wurden. Mit 42 % sind die Aspekte der zweiten Kategorie (*Dozent als Unterrichtender und als Person*) von Studierenden der klinischen Semester am häufigsten genannt worden. Studierende des vorklinischen Abschnitts nannten mit 48% die vierte Kategorie (*Themen und Inhalte*) am häufigsten.



**Abbildung 5: Häufigkeitsanalyse der positiven Hauptkategorien**

Die  $\chi^2$ -Werte betragen für die Vorklinik 594,2 und für die Klinik 453,0 und stehen bei einem Freiheitsgrad von 4 mit  $p < 0,01$  für ein höchst signifikantes Ergebnis. Es liegt damit keine Gleichverteilung der Kategorienhäufigkeiten vor. Die Kategorien +2 und +4 werden überdurchschnittlich häufig von beiden Gruppen genannt.

Mit insgesamt 2134 Nennungen überwogen bei der Inhaltsanalyse der gesamten Stichprobe die negativen Kategorien. 1208 (56,6%) dieser Nennungen kamen aus dem vorklinischen und 926 (43,4%) aus dem klinischen Abschnitt. Während die Studierenden der Vorklinik die Didaktik (Kategorie -3) mit 45% am stärksten kritisierten, waren die Studierenden im klinischen Abschnitt mit der Organisation des Unterrichtes (Kategorie -1, 38%) am wenigsten zufrieden (vgl. Abb. 6).



*Abbildung 6: Häufigkeitsanalyse der negativen Hauptkategorien*

Der Chi-Quadrat-Test führt mit  $\chi^2=579,4$  für die Vorklinik und  $\chi^2=332,8$  für die Klinik bei einem Freiheitsgrad von 4 auf dem 1% Niveau auch bei den negativen Hauptkategorien zu einem signifikanten Ergebnis. Die Nullhypothese, also eine Gleichverteilung, muss für beide Gruppen erneut abgelehnt werden.

## **7. Diskussion**

Die methodische Untersuchung des Marburger Evaluationsbogens hatte zum Ziel, die Dimensionalität der Skalen zur Struktur- und Prozessqualität sowie die Reliabilität des Fragebogens zu prüfen. Zusätzlich wurde der mögliche Einfluss von Motivation, Anwesenheitshäufigkeit in einer Veranstaltung und Semesterzahl auf das Urteil der Studierenden untersucht. Weiterhin sollte ein Kategoriensystem zur reliablen Einordnung der freien Meinungsäußerungen entwickelt und eine inhaltliche Analyse der Meinungsäußerungen vorgenommen werden. Im folgenden Abschnitt werden die im vorangegangenen Kapitel dargestellten Ergebnisse im Hinblick auf ihren Aussagewert und den daraus zu ziehenden Schlussfolgerungen kritisch diskutiert.

### **7.1. Diskussion der Stichprobe**

Bei der Stichprobengewinnung wurde auf das Verteilen der Fragebögen in nicht anwesenheitspflichtigen Vorlesungen verzichtet. Die Freiwilligkeit des Veranstaltungsbesuchs kann zu einer Selektion der Studierenden führen. Es ist anzunehmen, dass solche Veranstaltungen vorzugsweise von besonders interessierten und motivierten Studierenden besucht werden (Kromrey, 2000). Um zu einer repräsentativen Stichprobe zu gelangen und eine möglichst hohe Anzahl Studierender zu erreichen, wurden Veranstaltungen mit Anwesenheitspflicht, z.B. die letzten Semestertermine von Kursen, Praktika und Seminaren gewählt bzw. die Fragebögen im Rahmen einer Klausur ausgeteilt. Pro Kurs bzw. Praktikum gibt es allerdings 1 bis 2 Fehltermine, die wiederum von weniger motivierten Studierenden am Ende der Kurse z.B. zu Gunsten der Prüfungsvorbereitungszeit wahrgenommen werden könnten. Eine Selektion besonders motivierter Studierender ist somit nicht vollkommen auszuschließen.

### **7.2. Diskussion der Gütekriterien**

Die Reliabilität der geschlossenen Fragen wurde über die interne Konsistenz der Skala bestimmt. Eine gute interne Konsistenz der 12-Item-Skala für alle Veranstaltungen konnte mit Werten für Chronbach's alpha zwischen 0,85 bis 0,87 nachgewiesen werden. Die Ermittlung der Trennschärfen erbrachte allerdings, dass die Items SQ3, SQ 5 und SQ 6 für alle Veranstaltungstypen und zusätzlich Item SQ 3 für die Veranstaltungstyp

en Vorlesung und Seminar ungenügende Trennschärfen mit  $r_{it} < 0.5$  zeigten. Die Notwendigkeit der Elimination dieser Items wurde durch die Ergebnisse der Faktorenanalyse bestätigt (vgl. Kap. 7.3.).

Die psychometrische Überprüfung der nach Itemelimination verbleibenden 9-Item-Skala zeigte für alle Veranstaltungstypen eine höhere interne Konsistenz mit Chronbach's alpha zwischen 0,88 und 0,91 sowie hohe Trennschärfe der verbleibenden Items mit  $r_{it} > 0,5$ . Eine zufrieden stellende Messgenauigkeit des Marburger Evaluationsbogens konnte somit nachgewiesen werden.

### **7.3. Dimensionalität des Fragebogens**

Der Marburger Fragebogen soll als Instrument zur Bewertung der Veranstaltung hinsichtlich Lehrstruktur und Lehrprozess eingesetzt werden und die Übertragung der Urteile der Studierenden in ein Punktesystem ermöglichen. Dieses soll Basis für eine leistungsorientierte Mittelvergabe in der Lehre im Fachbereich Medizin der Philipps-Universität Marburg sein. Für den aus 2 Teilen bestehenden Fragebogen war vorgesehen, über die Items SQ 1-5 und über die Items PQ 1-6 jeweils den Mittelwert zu bilden. Item SQ 6 im Fragebogen zum Veranstaltungstyp Vorlesung bezieht sich auf die Anwesenheit der Studierenden in der Veranstaltung, welche wiederum von einer Vielzahl so genannter Biasvariablen wie Interesse, Lernverhalten, Vorkenntnisse und außeruniversitäre Rahmenbedingungen abhängig ist (Spiel & Gössler, 2000). Item SQ 6 des Praktikumsfragebogens diente zur Anwesenheitskontrolle der Hilfskräfte. Dennoch wurde bei der ersten Durchführung dieses Item berücksichtigt, so dass eine 12-Item-Skala zur Analyse stand.

#### **7.3.1. Dimensionalität der 12-Item-Skala**

Die Dimensionalität des Fragebogens wurde zunächst unter Einschluss aller Items bei alle Veranstaltungstypen untersucht. Idealerweise hätte sich eine zweidimensionale Skala nachweisen lassen sollen, bei der die Items zur Strukturqualität auf den ersten, die Items zur Prozessqualität auf den zweiten Faktor laden. Es zeigte sich allerdings eine dreidimensionale Skala und es folgte die Faktorenanalyse differenziert nach den Teilfragebögen zur Struktur- und Prozessqualität.

Für die Items zur Messung der Prozessqualität konnte damit die gewünschte Eindimensionalität für alle Veranstaltungstypen nachgewiesen werden. Dieser Teilfragebogen scheint damit der Zielsetzung, dass diese Items das gleiche Konstrukt messen, zu entsprechen.

Die Skala zur Messung der Strukturqualität zeigt sich dagegen dreidimensional. Dabei luden die Items SQ1, SQ2 und SQ 4 auf einen ersten gemeinsamen Faktor. SQ 3 und SQ 5 auf einen zweiten und SQ 6 auf einen dritten Faktor. Um die gewünschte Eindimensionalität zu erhalten waren somit die Items SQ 3, SQ 5 und SQ 6 auszuschließen. Eine mögliche Ursache für die Mehrdimensionalität kann in der Art der Datenauswertung liegen. Item SQ 3 bezieht sich auf die Qualität der Skripte, die Veranstaltungen häufig nicht vorhanden sind. Ein fehlendes Skript, von den Studierenden mit Ziffer „9“ im Fragebogen gekennzeichnet, wurde mit der Note 6 gewertet. Somit erniedrigt bei einem sonst als „gut“ bewertetem Seminar das fehlendem Skript mit der Codierung als Note 6 nicht nur den Mittelwert über die Items des Fragebogens. Item SQ 3 fällt auch aus dem Bewertungsprofil „gut“ heraus, zeigt folglich eine geringe Korrelation zu den anderen Items und lädt damit auf einen separaten Faktor. Daher ist Item SQ 3 für die generelle Beurteilung der Veranstaltung ungeeignet. Item SQ 5 geht der Frage nach wie viel Unterricht ausgefallen ist und dient damit der Kontrolle des Unterrichtenden bzw. der Kontrolle der zeitlichen Organisation des Unterrichts. Fand der Unterricht regelmäßig statt und die Studierenden gaben den Unterrichtsausfall als gering an, so wurde dieses Item mit einer guten bis sehr guten Note codiert, auch wenn Konzept, Inhalte, Skripte und die Vorbereitung der Dozenten eine im Vergleich dazu eher schlechte Beurteilung erhielten. Bei Analyse der Daten fiel auf, dass insgesamt wenig Unterricht ausfiel, so dass sich bei Item SQ 5 eine geringe Varianz ergibt. Dies spiegelt sich ebenfalls in der niedrigen Korrelation zu den andern Items wieder und führt zur Ladung auf einen separaten Faktor. Die Mehrdimensionalität der Strukturskala ist vermutlich in einem Auswertungsartefakt begründet.

Item SQ 6, welches lediglich als Kontrollinstrument für die Anwesenheit der Studierenden in Vorlesungen bzw. das Vorhandensein von Hilfskräften konzipiert war, lud wie erwartet auf einen separaten Faktor. Die Anwesenheit von Studierenden in einer Veranstaltung ist von einer Vielzahl anderer Faktoren abhängig, die nicht von der Struktur und damit von der Hochschule beeinflussbar sind (Spiel & Gössler 2000;

Kromrey 2000). Item SQ 6 liefert damit keine relevanten Daten zur Strukturqualität in Vorlesungen. Auch stehen nicht in jedem Kurs oder Praktikum Hilfskräfte zur Verfügung, was im Rahmen der Auswertung zu einer Codierung mit der Note 6 führte. Somit entsteht wiederum ein Auswertungsartefakt. Die Ergebnisse der Faktorenanalyse bestätigen nur die Richtigkeit des im Vorfeld festgelegten Ausschlusses von Items SQ 6 bei der Mittelwertbildung über alle Items.

Die faktorenanalytischen Ergebnisse der 12-Item-Skala zeigen nicht nur die Mehrdimensionalität der Skala zur Messung der Strukturqualität. Es musste auch festgestellt werden, dass die Items SQ 1, SQ 2 und SQ 4 zusammen mit den Items der Prozessqualität auf einen gemeinsamen Faktor luden. Die Ursache dafür könnte auf der einen Seite in der Auswahl der Items, welche nicht genügend zwischen Prozess- und Strukturqualität zu trennen vermögen, gesehen werden. Damit wäre der Fragebogen in der vorliegenden Konzeption nicht geeignet zur getrennten Erfassung von Prozess- und Strukturqualität. Auf der anderen Seite ist es auch möglich, dass der Großteil der Studierenden nicht zwischen Prozess- und Strukturqualität bei Bearbeitung des Fragebogens differenziert und ein Globalurteil über die zu bewertende Veranstaltung abgeben hat.

### 7.3.2. „Index der allgemeinen Lehrqualität“

Es zeigte sich, dass sich die bei der Fragebogenkonstruktion vorgenommene Unterscheidung in Prozess- und Strukturmerkmale in den Ergebnissen der Faktorenanalyse nicht wieder findet. Die drei Items der Strukturskala SQ 1 (*Bekanntmachung des Lehrkonzepts*), SQ 2 (*Abstimmung von Lehr- und Prüfungsinhalt*) und SQ 4 (*Vorbereitung der Dozenten*) messen offensichtlich das gleiche Konstrukt wie die Items der Prozessskala. Um den Fragebogen dennoch als Instrument zur Bewertung der Lehrqualität zu nutzen, wurden die auf einen Faktor ladenden neun Items zu einer Skala zusammengefasst und näher untersucht. Dabei zeigten sich eine gute interne Konsistenz (vgl. Kap. 7.2.) und die Eindimensionalität der Skala. Das Konstrukt kann als „Index der allgemeinen Lehrqualität“ bezeichnet werden. Die Bildung eines Mittelwerts über diese 9-Item-Skala ist nun möglich und dadurch auch die angestrebte Übertragung der Urteile der Studierenden in ein Punktesystem. Die übrigen Items der Strukturskala sollten nicht in den Summenwert eingehen, sondern als Deskriptoren einzelner



struktureller Aspekte der Lehre verwendet werden (Basler & Krebs, 2003). Dabei handelt es sich um die Qualität der Skripten (SQ 3), den Unterrichtsausfall (SQ 5) und die Häufigkeit des eigenen Vorlesungsbesuchs bei Vorlesungen (SQ 6) bzw. den Beitrag der Semesterhilfskräfte zum Lernerfolg bei Kursen und Praktika (SQ 6). Somit fließen Unterrichtsausfall, Beitrag von Hilfskräften und Existenz sowie Qualität von Skripten nicht direkt in die Beurteilung der allgemeinen Lehrqualität ein. Das muss allerdings nicht daran liegen, dass diese Kriterien für die Beurteilung der Qualität der Lehre irrelevant sind. Dieser Befund könnte auch durch eine mangelnde Varianz des Urteils bei diesen Kategorien oder durch deutlich abweichende Verteilungsformen dieser Items zu erklären sein, so dass er in einem statistischen Artefakt begründet wäre. Eine Betrachtung der Daten zeigt in der Tat, dass sehr wenig Unterricht ausfällt, so dass sich bei diesem Item eine geringe Varianz ergibt (vgl. Kap. 7.3.1.). Darüber hinaus werden in vielen Veranstaltungen keine Skripten verteilt, so dass auch dieses Item für eine generelle Beurteilung der Veranstaltungen ungeeignet ist. Ebenso sind nicht in allen Veranstaltungen Hilfskräfte vorhanden.

#### **7.4. Diskussion der Varianzanalyse und des U-Tests**

Berger et al. (2003) konstatierten, dass die Häufigkeit des Veranstaltungsbesuches die Lehrveranstaltungsevaluation systematisch verzerrt. So würden Veranstaltungen mit Anwesenheitspflicht wie Seminare und Kurse, die immer besucht werden, signifikant besser beurteilt als freiwillige Vorlesungen. Demgegenüber stehen die Untersuchungen Kromreys (1994), welche zu dem Schluss führten, dass Pflichtveranstaltungen schlechter beurteilt werden aufgrund der Teilnehmerstruktur. Bei der Beurteilung von Pflichtveranstaltungen fließe das negative Urteil unmotivierter Studierender ein, während bei freiwilligen Veranstaltungen vorwiegend interessierte Studierende bewerten. Es galt der Frage nachzugehen, ob die Motivation der Studierenden, die Häufigkeit des Veranstaltungsbesuches oder die Semesteranzahl Einfluss auf die Urteilsbildung der Studierenden nehmen. Mittels Varianzanalyse konnte gezeigt werden, dass die nicht anwesenheitspflichtigen Vorlesungen besser beurteilt wurden als die Pflichtveranstaltungen (Kurse, Praktika und Seminare), welches der Schlussfolgerung Kromreys (1994) entspricht. Berger et al. (2003) gingen von der Annahme aus, dass Pflichtveranstaltungen besser beurteilt wurden, da die Studierenden häufiger anwesend waren als bei freiwilligen Veranstaltungen. Um einen möglichen

Einfluss der Häufigkeit eines Veranstaltungsbesuches näher zu betrachten, wurden die Daten der Vorlesungen einem U-Test nach Mann und Whitney unterzogen, da der Fragebogen für Vorlesungen mit Item SQ 6 konkrete Angaben zur Anwesenheitshäufigkeit enthielt. Es konnte allerdings kein signifikanter Unterschied in den Urteilen der Studierenden, welche häufig ( $> 50\%$ ) oder selten ( $< 50\%$ ) die Veranstaltung besuchten, nachgewiesen werden.

Ein Einfluss der Semesterzahl und damit des Alters und der Erfahrung der Studierenden auf die Beurteilung von Wahlpflichtveranstaltungen (Vorlesungen) konnte nicht nachgewiesen werden. Die Mittelwerte über die 9-Item-Skala der dafür vorgenommenen Einteilung in drei Studienabschnitte Vorklinik, 1. klinischer und 2. klinischer Abschnitt zeigten keinen signifikanten Unterschied. Dies bestätigt die Aussage Rindermanns (2001), der die Semesterzahl als Bias weitestgehend ausschloss. Bei Bewertung der Pflichtveranstaltungen (Kurse und Praktika) zeigte sich allerdings, dass Studierende des zweiten klinischen Abschnittes diese mehr als eine halbe Note besser bewerteten als Studierende der Vorklinik und des ersten klinischen Abschnittes. Eine mögliche Erklärung könnte in der Organisationsstruktur und den Inhalten des Unterrichtes liegen. Die Kurse in den klinischen Fächern und Unterricht am Patientenbett bieten mehr Praxisbezug und könnten daher mehr das Interesse und die Motivation wecken als Vorlesungen und Experimente zu vorklinischen Sachverhalten. Ein höherer Anteil intrinsisch motivierter Studierenden würde dann auch zu einer besseren Bewertung führen (Kromrey, 1994).

### **7.5. Diskussion der Kategorisierung**

Aus den freien Meinungsäußerungen der Studierenden wurde ein Kategoriensystem entwickelt, welches die folgenden 5 Hauptkategorien enthält:

1. Organisation des Unterrichts
2. Didaktik
3. Dozent als Unterrichtender und als Person
4. Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz
5. Medieneinsatz, Bild- und Arbeitsmaterialien.

Diese fünf Hauptkategorien wurden unterteilt in 24 positive und 34 negative Subkategorien. Eine reliable Einordnung der Antworten in dieses Kategoriensystem war möglich. Die Abweichung in der Anzahl der Unterkategorien für positive und negative Aussagen liegt darin begründet, dass die Studierenden bei Zufriedenheit eher global eine Veranstaltung als allgemein „gut“ beurteilen, als einzelne Aspekte aufzuzählen, die ihnen besonders zusprachen. So wurde bei der Auswertung häufiger das pauschale Urteil „der Unterricht bzw. die Vorlesung war gut/ interessant“ gefunden, als eine differenzierte Aussage, aus denen sich Subkategorien hätten entwickeln lassen. Dagegen wurde Kritik wesentlich differenzierter zum Ausdruck gebracht, was zur Konsequenz hatte, dass insgesamt 34 negative Subkategorien entwickelt werden konnten.

### **7.5.1. Reliabilität des Kategoriensystems**

Zur Bestimmung der Reliabilität des Kategoriensystems wurde Cohens Kappa gewählt. Für die Hauptkategorien konnten mit Cohens Kappa von 0,76 und 0,80 jeweils eine gute Intra- und Interraterreliabilität nachgewiesen werden. Es zeigte sich in beiden Fällen höhere Übereinstimmung für Zuordnungen zu den positiven Kategorien. Dabei ist zu beachten, dass das Maß der Übereinstimmung von der Differenziertheit des Kategoriensystems abhängig ist (Rustemeyer, 1992). Je differenzierter die Kategorien sind, umso ungünstiger fällt Cohens Kappa aus. Dies ist der Fall für die negativen Kategorien, die über zehn mehr Subkategorien als die positiven Kategorien verfügen.

### **7.5.2. Diskussion der Ergebnisse der Frequenzanalysen**

Bei der Frequenz- und Inhaltsanalyse zeigte sich insgesamt die Tendenz zu negativen Äußerungen (56,9%). Dabei wurde Didaktik (37,8%) und Organisation (29,2%) von allen Studierenden am häufigsten bemängelt und somit sowohl Aspekte der Prozess- als auch der Strukturqualität kritisiert. Positiv beurteilt wurden besonders die Themen und Inhalte (39,4%) sowie die Dozenten (34,5%). Es galt nun mögliche Unterschiede bei der Kategoriennennung zwischen dem vorklinischen und klinischen Abschnitt zu untersuchen.

Die Frequenzanalyse zeigte, dass in beiden Gruppen positive Aussagen etwa gleichhäufig vorkamen. Am häufigsten positiv mit „gut“, „interessant“ und

„praxisrelevant“ wurden von Studierenden der Vorklinik die Themen und Inhalte (48%) der Veranstaltungen beurteilt. Für Studierende der Klinik stand diese Kategorie an zweiter Stelle. An erster Stelle in dieser Gruppe waren Nennungen zur Kategorie 2 (42%), welche den Dozenten als Unterrichtenden und Person zum Inhalt hatten. Freundlichkeit, Motivation und guter, effektiver Unterricht wurden positiv erwähnt. Oftmals kam es auch zu dem pauschalen Urteil, dass der Dozent „gut“ sei. Erst an dritter Stelle der positiven Nennungen steht für beide Gruppen die Didaktik. Von den Studierenden wird als positiv empfunden, dass eine Veranstaltung systematisch und übersichtlich strukturiert ist, die Wissensvermittlung verständlich und praxisnah durch Fallbeispiele oder Patientenvorstellungen gestaltet ist und dabei auf das Wesentliche konzentriert bleibt. Fasst man diese Ergebnisse zusammen, so ist festzustellen, dass das Thema einer Veranstaltung, die Lehrpräsentation durch den Dozenten und die allgemeine Didaktik als Aspekte der Prozessqualität bei der Beurteilung von Veranstaltungen besonders stark von den Studierenden wahrgenommen wurden. Dagegen wurden strukturelle Aspekte wie die Organisation der Lehre nur selten (Kategorie 1 mit 1,8%) positiv genannt.

Im Rahmen der negativen Nennungen kritisierten die Studierenden die Organisation und die Didaktik am stärksten. Dabei überwog die Kritik aus den Reihen der Vorklinik. Sie bemängelten die Didaktik am häufigsten. Die Veranstaltungen wurden als unsystematisch, unstrukturiert, unverständlich (z.B. aufgrund zu schnellen Vorgehens oder Vorraussetzung von nicht bestehenden Vorkenntnissen) und praxisfern empfunden. Die Studierenden der Klinik sind am wenigsten mit der Organisation des Unterrichts, speziell zeitlichen Aspekten (Kategorie -1.2.) zufrieden. Es käme häufig zu Terminüberschneidungen mit anderen Kursen oder Vorlesungen, Dozenten erschienen nicht pünktlich und es entstünden Wartezeiten bis hin zu Unterrichtsausfall.

Worin sind nun die Ursachen der dargestellten Unterschiede in den Nennungen zu sehen? Sowohl die Studierenden der Vorklinik als des klinischen Abschnittes zeigten bei der Analyse der Hauptkategorien die gleichen Schwerpunkte. Die Unterschiede in der Häufigkeit der Nennung von Kategorien spiegeln die Unterschiede in den Unterrichtsstrukturen zwischen Vorklinik und Klinik wieder, welche den Lehrprozess maßgeblich beeinflussen. In der Vorklinik finden die meisten Unterrichtsveranstaltungen in Form von Vorlesungen, Praktika und Seminaren mit häufig gleich

bleibenden und unterrichtserfahrenen Dozenten statt. Dagegen ist der Unterricht im klinischen Abschnitt in Form von Vorlesungen und Kursen in Groß- und Kleingruppen organisiert. Dieser in Kursform gestaltete Stundenplan bringt in einem hohen Maß organisatorische und inhaltliche Probleme. Zwar finden die Vorlesungen zu festen Zeitpunkten an festgelegten Wochentagen statt, die Kurstermine variieren dagegen in Wochentag und Zeitpunkt. Dabei kann es zu Überschneidung mit den Vorlesungsterminen kommen. Dementsprechend groß ist die Kritik an der Organisation des Unterrichts. Weiterhin behandeln Vorlesungen zwar die im Vorfeld durch einen Vorlesungsplan bekannt gegebenen Themen. Für Kurse gibt es aber keinen vorgegebenen Themenkatalog, die Inhalte richten sich meist nach den aktuell in der Klinik vorliegenden Krankheitsbildern. Weiterhin finden die Kurse häufig auf wechselnden Stationen und mit wechselnden Dozenten statt, die der Doppelbelastung von Unterrichten und Stationsarbeit ausgesetzt sind und z.T. erst kurz vor Veranstaltungsbeginn über ihre Unterrichtsaufgabe informiert wurden. Fehlender Themenkatalog, mangelnde Vorbereitungszeit und Doppelbelastung der Dozenten wirken sich negativ auf die Lehrqualität der Kurse aus. Umso mehr fällt es für Studierende der Klinik ins Gewicht, dass der Dozent motiviert ist und effektiven Unterricht geben kann, was wiederum eine entsprechende Qualifikation fachlich wie didaktisch voraussetzt. Detaillierter können Schlussfolgerungen aus der Analyse der Subkategorien gezogen werden. Auch wenn bei den positiven Nennungen bezüglich der Subkategorien keine nennenswerten Unterschiede aufgezeigt werden konnten, so zeigten sich Unterschiede bei der Nennung negativer Subkategorien. Diese weisen, wie auch die Häufigkeit der Nennungen in Hauptkategorien auf studienabschnittsspezifische organisatorische und strukturelle Probleme in der Lehre hin. Mit 9,1% der negativen Nennungen wurde an der fehlenden Kontinuität des Dozenten (-1.5.) am häufigsten aus Reihen der sich im klinischen Abschnitt befindenden Studierenden Kritik geübt. Ebenso wurde die zeitliche Organisation (-1.2.1. ; 5,7%) und die langen Wartezeiten (-1.2.3. ; 4%) kritisiert. Die vorgegebene Anzahl an Kursterminen wurde eher als zu gering (5,7%) als zu viel (3,1%) empfunden. Auch in der Vorklinik ist die Organisation der zweithäufigste Kritikpunkt der Studierenden. Sie empfinden die Veranstaltungen als zu früh für ihren Wissensstand (-1.6.1.; 6,8%) und die Terminanzahl, im Gegensatz zum klinischen Abschnitt, eher zu hoch (-1.6.3.; 3,9%) als zu gering (-1.6.2.; 1,7%). Letzteres kann als Ausdruck einer Überforderung der Studierenden im vorklinischen Abschnitt mit dem Curriculum sowohl zeitliche als auch inhaltlich interpretiert werden.

Dies spiegelt auch die häufige Aussage wieder, dass in Veranstaltungen großes fachliches Wissen vorausgesetzt werde, Grundkenntnisse im Vorfeld jedoch nicht vermittelt würden (-3.5.1.; 3,4%) und damit der Unterrichtsstoff nicht nachvollziehbar sei (-3.5.2.; 3,6%). Referate werden von einem Teil der Studierenden als ungeeignete Unterrichtsform betrachtet (-3.9.; 4,5%), da diese häufig zu hochspezialisierten Themen vorgetragen würden. Auch wird den Studierenden der Vorklinik der Zusammenhang zwischen Übungen, Experimenten, Aufgaben und der Praxis zu wenig deutlich (-3.4.; 4,5%).

## **7.6. Kritische Auseinandersetzung mit den Ergebnissen**

Bei der Datenerhebung dieser Arbeit wurde auf die Austeilung der Fragebögen in Vorlesungen verzichtet, um den durch die Selektion besonders interessierter Studierenden auftretende Verzerrungseffekte weitestgehend zu vermeiden. Wie in Abschnitt 7.1. dargestellt ist eine Selektion besonders interessierter Studierender und damit Erhebungsartefakte jedoch nicht vollständig auszuschließen. Es konnte gezeigt werden, dass der Fragebogen reliabel ist, in seiner ursprünglichen Konzeption allerdings nicht geeignet war, Struktur- und Prozessqualität getrennt voneinander zu erfassen. Eine Reduktion der Skala auf 9 Items, welche das gleiche Konstrukt, einen „Index der allgemeinen Lehrqualität“ messen, ermöglicht allerdings die Bildung von Mittelwerten über alle Items und damit die angestrebte Überführung der Ergebnisse des Fragebogens in ein Punktesystem zur leistungsorientierten Mittelvergabe. Aufgrund der in Abschnitt 7.3.1. erläuterten Datenauswertung kann bei Betrachtung der faktorenanalytischen Ergebnisse das Vorliegen von Auswertungsartefakten nicht ausgeschlossen werden. Mit der Untersuchung der Bias-Variablen Motivation, Anwesenheit und Semesterzahl sollten mögliche Verzerrungseffekte auf die studentische Lehrevaluation aufgedeckt werden (vgl. Abschnitt 7.4.). Es konnte gezeigt werden, dass freiwillige Veranstaltungen signifikant besser beurteilt wurden als Pflichtveranstaltungen. Somit kann vermutet werden, dass die Motivation des Evaluierenden einen Einfluss auf deren Urteil haben kann. Ein solcher Einfluss auf das Veranstaltungsurteil konnte dagegen für die Anwesenheitshäufigkeit nicht nachgewiesen werden. Dem dazu durchgeführten U-Test nach Mann und Whitney ist anzumerken, dass die Fallzahlen relativ gering waren, was wiederum Einfluss auf die Teststärke hat. Je größer die Teststärke ist, umso wahrscheinlicher wird ein tatsächlich vorhandener Unterschied zu einem signifikanten Ergebnis führen. Aufgrund der geringen Fallzahl liegt hier allerdings eine geringe

Teststärke vor, welche einen Fehler 2. Art (Irrtumswahrscheinlichkeit) zur Folge haben kann. Es kann daher nicht sicher ausgeschlossen werden, dass dies bei den vorliegenden Ergebnissen der Fall ist. Das würde bedeuten, dass der doch vorhandene Effekt der Anwesenheitshäufigkeit auf das Urteil mit den vorliegenden Ergebnissen nicht nachgewiesen werden konnte. Für den Einfluss der Semesterzahl auf das Veranstaltungsurteil wurden Varianzanalysen durchgeführt, die für den Veranstaltungstyp Vorlesung keine signifikanten Unterschiede in den Semesterabschnitten und damit keinen Einfluss der Semesterzahl auf das Urteil vermuten lässt. Anders fielen die Ergebnisse für den Veranstaltungstyp Praktika und Kurse aus. Hier scheint ein Einfluss der Semesterzahl nachweisbar zu sein. Die der Varianzanalyse für Praktika und Kurse vorausgehende Überprüfung auf Homogenität der Varianzen zeigte, dass sich diese in den Gruppen Vorklinik, 1. klinischer Abschnitt und 2. klinischer Abschnitt signifikant unterscheidet. Aufgrund der fehlenden Homogenität der Varianzen ist der sich anschließende F-Test unzuverlässig und ebenso wie die Einzelvergleiche nach Scheffé mit Vorsicht zu interpretieren. Mit den vorliegenden Ergebnissen zum Veranstaltungstyp Praktika und Kurse ist folglich weder eine Bestätigung noch eine Widerlegung der Aussage Rindermanns (2001) möglich.

## 8. Zusammenfassung und Fazit

Deutsche Hochschulen stehen national und international in einem zunehmenden Wettbewerb um Studierende, wissenschaftliche Mitarbeiter, Lehrende und finanzielle Mittel. Vor diesem Hintergrund nimmt die Diskussion über die Qualität in der Lehre und die daraus resultierenden Maßnahmen im Hinblick auf Qualitätsstandards, Qualitätssicherung und Qualitätsverbesserung in der Lehre und um die Anwendung wirtschaftsspezifischer Qualitätsmanagementsysteme ständig zu. Ziel der vorliegenden Arbeit war es, den Beitrag von Evaluationen zur Qualitätsverbesserung der Lehre im Kontext des Qualitätsmanagements der Hochschule zu diskutieren und vor diesem Hintergrund den Marburger Fragebogen zur Evaluation des Lehrangebotes in der Medizin im empirisch-statistischen Anteil der Arbeit zu untersuchen. Die Überprüfung der Dimensionalität der Skalen zur Struktur- und Prozessqualität, der Reliabilität des Fragebogens und die Entwicklung eines Kategoriensystems zur reliablen Einordnung der freien Meinungsäußerungen sowie deren inhaltliche Analyse waren dabei die Schwerpunkte. Zusätzlich sollte der mögliche Einfluss von Anwesenheitshäufigkeit in einer Veranstaltung, Motivation und Semesterzahl auf das Urteil der Studierenden untersucht werden.

Vor der Diskussion der Umsetzungsmöglichkeiten eines Qualitätsmanagements in den Hochschulbereich waren die Begriffe Qualität und Qualität in der Lehre zu erörtern. Qualität allgemein stellt eine Übereinstimmung von Leistungen und Anforderungen aus den verschiedenen Perspektiven an ein Produkt oder eine Dienstleistung dar. Diese Anforderungen entsprechen den zu erfüllenden Qualitätsmerkmalen, welche anwendungs- und zielgruppenorientiert definiert werden müssen. Zur Festlegung von Qualitätsmerkmalen sind grundsätzlich alle Personen berechtigt, die ein Interesse an einem Produkt oder einer Dienstleistung haben, so z.B. Leistungserbringer, Kunden, Verbraucher und Geldgeber. Da sich die Anforderungen und Kundenbedürfnisse in einem ständigen Wandel befinden, ist Qualität als eine dynamische Größe zu betrachten, die im Prozess der Qualitätsverbesserung stetig neu zu definiert ist. Der vorliegenden Arbeit wurde der Qualitätsbegriff Donabedians (1980) zugrunde gelegt und auf die Hochschullehre übertragen. Donabedian schlüsselt den Qualitätsbegriff in die drei Dimensionen Struktur-, Prozess- und Ergebnisqualität auf, die im Marburger Evaluationsmodell ihre Anwendung finden. Strukturqualität umfasst die strukturellen Charakteristika und Rahmenbedingungen einer Einrichtung, die der Erbringung einer



Dienstleistung wie der Lehre dienen, z.B. die Zahl und die Ausbildung der Mitarbeiter sowie die Qualität und Quantität weiterer zur Erstellung der Leistung notwendiger Ressourcen wie Organisation, finanzielle Ausstattung, Infrastruktur und Gebäude. Mit Prozessqualität sind alle Arbeitsabläufe gemeint, die der Durchführung und der Erbringung der Dienstleistung dienen. Auf den Hochschulbereich übertragen umfasst Prozessqualität z.B. den Ablauf und die Didaktik der Unterrichtsveranstaltungen. Die Ergebnisqualität lässt schließlich auf den Erfolg oder Misserfolg der Leistung schließen. In Hochschulen können zur Bestimmung der Ergebnisqualität der Lehre die Prüfungsergebnisse herangezogen werden. Qualität in der Lehre ist ebenso eine Frage der Lehrkraft, der Lehrpräsentation, der Entstehung und organisatorischen Umsetzung des Lehrplans wie eine Frage der Gestaltung des sozialen Umfeldes und der Ausstattung. Aus diesen Anforderungen lassen sich Qualitätsmerkmale für gute Lehre auf den Ebenen Struktur, Prozess und Ergebnis ableiten, wobei zu beachten ist, dass auch die Qualitätskriterien für gute Lehre einem dynamischen Prozess unterliegen. Auf Ebene der Strukturqualität sind es die Lehrausstattung und das Curriculum, die entsprechend zu optimieren sind. Unter der Ausstattung der Lehre werden die räumlichen Ausstattungen, die Bibliothek, die Arbeits- und Lernbedingungen und das studentische Umfeld verstanden. Das Curriculum ist so zu strukturieren, dass eine inhaltlich aufeinander abgestimmte Abfolge von Veranstaltungen ohne terminliche und inhaltliche Überschneidungen oder Lücken möglich ist. Prozessqualität in der Lehre entspricht der Qualität der Lehrpräsentation. Qualität in der Lehre setzt hier didaktische Kompetenz des Lehrpersonals voraus. Kriterien einer „guten“ Didaktik sind Motivation, Engagement, rhetorische Kompetenz, strukturierte, verständliche Erklärungen und Darstellungen der Lehrinhalte sowie die Fähigkeit Interesse zu wecken. Die Güte der Lehrpräsentation ist z.T. allerdings auch abhängig von der strukturellen Qualität. Darunter sind die verwendeten Mittel, Materialien und Skripte zur Lehrpräsentation als auch die Bekanntmachung von Ablauf, Zielen und Anforderungen der Lehrveranstaltungen und Prüfungen zu verstehen. Ergebnisqualität in der Lehre bedeutet neben entsprechend guten Prüfungsergebnissen auch die Bewährung der Absolventen im Arbeitsmarkt. Die akademische Lehre ist darauf ausgerichtet, Studierende auf die spätere Berufstätigkeit durch Vermittlung der entsprechenden Sachkenntnisse und Fertigkeiten vorzubereiten und somit den erfolgreichen Eintritt in das Berufsfeld zu ermöglichen. Je mehr die Lernziele des Studienganges aus dem

Berufsprofil abgeleitet werden, umso mehr kann die Lehre der oben genannten Funktion gerecht werden und eine entsprechend gute Ergebnisqualität hervorbringen.

Die Einführung von Qualitätsmanagement an Hochschulen wird als eine mögliche Strategie zur Qualitätssicherung und Qualitätsverbesserung vielfach diskutiert. Unter Qualitätsmanagement versteht man die aufeinander abgestimmten Tätigkeiten zum Lenken und Leiten einer Organisation, die darauf abzielen, die Qualität der Produkte oder der angebotenen Dienstleistung zu verbessern. Wollen Hochschulen QM einführen, so sind die Konsequenzen daraus die Entwicklung eines eigenen Leitbildes und eigener Qualitätsziele, die Erarbeitung von Strategien zu deren Umsetzung und die Einführung von Prüfverfahren, welche die Wirkung der ergriffenen Maßnahmen zur Qualitätsverbesserung betrachten. In der Praxis findet üblicherweise ein Kreislaufprozess im Sinne des von Edward Deming (1986) beschriebenen Deming-Zyklus mit den vier Phasen der Planung, Durchführung, Prüfung und Qualitätsverbesserung Anwendung. Im Hochschulbereich lassen sich daraus vier Aufgabengebiete ableiten: Bestandsaufnahme, Festlegung von Entwicklungszielen, Maßnahmen zur Qualitätssteigerung und Überprüfung der Wirksamkeit von Reformaktivitäten (Arnold, 2003). Qualitätsverbesserung in der Lehre beginnt mit einer Bestandsaufnahme der aktuellen Situation. In der Hochschullehre kann sich diese in Anlehnung an Donabedians Qualitätsdimensionen auf strukturelle Bedingungen (Ausstattung, Personal, Organisationsstruktur), auf Prozesse (Planung und Präsentation des Lehrangebots, Beratung und Betreuung der Studierenden) und auf Ergebnisse (Prüfungsnoten, Studiendauer, Zufriedenheit der Studierenden) beziehen. Der Analyse der Ausgangssituation schließt sich die Festlegung von Entwicklungszielen und Planung von Aktivitäten zur Umsetzung dieser Ziele an. Die Maßnahmen müssen nicht sofort das gesamte System Hochschule ergreifen, sondern können in Form von Reformprojekten zu einzelnen Themen eine Art Testphase durchlaufen. Am Ende dieser Phase steht die Prüfung der Wirksamkeit der eingesetzten Reformmaßnahme. Zusammen mit den Struktur- und Prozessdaten der allgemeinen Bestandsaufnahme vervollständigen die Daten der Evaluation der Reformaktivitäten im Sinne von Ergebnisdaten das Qualitätsbild der Hochschule. Aus diesem können wiederum neue Qualitätsverbesserungsmaßnahmen erarbeitet werden. Der Kreislauf wird erneut durchlaufen mit dem Ziel einer kontinuierlichen Qualitätsverbesserung in der Lehre.

Evaluationen sind neben Revisionen und Controlling ein mögliches Prüfverfahren im Qualitätsmanagement. Sie sind das in der Praxis übliche Prüfverfahren im Prozess der Qualitätsverbesserung an Hochschulen und stellen systematische Untersuchung zum Nutzen oder Güte eines Gegenstandes oder Programms unter Anwendung von sozialwissenschaftlichen Forschungsmethoden dar. Evaluationen an Hochschulen dienen dem Informationsgewinn im Sinne einer Stärken-Schwächen-Analyse (Erkenntnisfunktion), der Entstehung eines Diskurses über Qualität von Lehre und Studium und der Entwicklung eines Qualitätsbewusstseins (Dialogfunktion) und daraus resultierend der Initialisierung von Maßnahmen zur Veränderung und Verbesserung von Organisationsstruktur und Lehrqualität. Evaluationen können eingesetzt werden zur Überprüfung der Wirksamkeit implementierter Maßnahmen (Kontrollfunktion). Weiterhin schaffen sie Transparenz über Leistungen sowie verwendete Mittel und dienen damit der Rechenschaftslegung der Hochschule gegenüber den Stakeholdern (Legitimationsfunktion). Somit werden Evaluationen Bestandteil eines Gestaltungsprozesses und leisten einen wichtigen Beitrag zur Qualitätsverbesserung und -sicherung. Nicht unumstritten ist die Rolle der Evaluation als hochschulinternes Kontroll- und Steuerungsinstrument, wenn finanzielle Anreize und Sanktionen damit verbunden sind. Um gültige Ergebnisse zu liefern, bedarf es der Einsicht in die Notwendigkeit einer Evaluation und der Akzeptanz des Verfahrens. Von geringer Akzeptanz ist ein Evaluationsverfahren, das als reines Kontrollinstrument eingesetzt wird und Sanktionen zur Folge hat. Damit werden die Evaluierten verleitet, kritische Informationen zu verschleiern und positive in den Vordergrund zu spielen (Kromrey, 2001).

Der Fachbereich Humanmedizin der Philipps-Universität Marburg hat sich im Rahmen der Qualitätsverbesserung in der Lehre für eine interne Evaluation der Lehrqualität differenziert in Struktur-, Prozess- und Ergebnisqualität entschieden. Dafür wurde am Ende des Sommersemesters 2002 erstmalig der Marburger Fragebogen zur Evaluation der Lehrangebote in der Medizin eingesetzt, der getrennt nach den Veranstaltungstypen Vorlesung, Kurse und Praktikum und Seminar eine differenzierte Beurteilung der Lehrstruktur und des Lehrprozesses durch die Studierenden ermöglichen sollte. Um eine möglichst hohe Anzahl Studierender zu erreichen, wurden die Fragebögen in Pflichtveranstaltungen oder zusammen mit einer Klausur ausgehändigt. Auf diese Weise liefen 657 Fragebögen zurück, von denen unter Berücksichtigung der Ausschlusskriterien 597 Bögen (90,9%) zur Analyse der geschlossenen Fragen und 447

Bögen (68%) zur Entwicklung eines Kategoriensystems sowie zur inhaltlichen Analyse der freien Meinungsäußerungen herangezogen werden konnten. Als Datenbasis galten die Urteile der Studierenden. Die Untersuchung der geschlossenen Fragen beinhaltete eine Itemanalyse mit Bestimmung der Mittelwerte, Standardabweichungen sowie die Berechnung der Trennschärfe der Items. Die Reliabilität des Fragebogens wurde durch Berechnung der internen Konsistenz unter Verwendung des Reliabilitätskoeffizienten Chronbach's alpha ermittelt. Anschließend wurde die Dimensionalität des Fragebogens mittels Faktorenanalyse (Hauptkomponentenanalyse mit Varimax-Rotation, Abbruchkriterium Eigenwert  $< 1$ ) untersucht. Zur Untersuchung der Bias-Variablen Häufigkeit des Veranstaltungs-besuchs, Motivation und Semesterzahl wurden einfaktorielle Varianzanalysen (ANOVA) und der U-Test nach Mann und Whitney durchgeführt. Die Reliabilität des Kategoriensystems wurde mit zwei Verfahren, der Intraraterreliabilität und der Interraterreliabilität überprüft. Dabei wurde Cohens Kappa als Reliabilitätsindex gewählt. Anschließend wurden zur inhaltlichen Untersuchung Frequenzanalysen für die gesamte Stichprobe und getrennt in vorklinischen und klinischen Abschnitt durchgeführt.

Die Ergebnisse der deskriptiv-statistischen Untersuchung der ursprünglichen 12-Item-Skala zeigten zwar eine zufrieden stellende interne Konsistenz, drei Items fielen jedoch durch eine mangelhafte Trennschärfe auf. Es waren die Items SQ 3 (*Qualität der Unterrichtsmaterialien*), SQ 5 (*Ausfall der Veranstaltung*) und SQ 6 (*Anwesenheit der Studierenden in der Veranstaltung bzw. Beitrag des Semesterhilfskräfte*). Diese Items luden in der Faktorenanalyse auf separate Faktoren und die Skala zur Strukturqualität war damit dreidimensional. Dagegen konnte durch Faktorenanalyse die gewünschte Eindimensionalität der Skala zur Prozessqualität bestätigt werden. Mit dem Ziel eine eindimensionale Skala zur Strukturqualität zu erhalten, wurden die drei Items SQ 3, SQ 5 und SQ 6 eliminiert. Durch diesen Schritt wurde bei der Analyse der verbleibenden 9-Item-Skala die gewünschte Eindimensionalität der Strukturskala erreicht. Zusätzlich musste festgestellt werden, dass die Items SQ 1, SQ 2 und SQ 4 mit den Items zur Prozessqualität auf einen gemeinsamen Faktor luden. Folglich messen die verbleibenden 9 Items das gleiche Konstrukt, welches als „Index der allgemeinen Lehrqualität“ bezeichnet wurde. Die getrennte Erfassung von Prozess- und Strukturqualität ist mit dem Marburger Evaluationsbogen nach den vorliegenden Ergebnissen nicht möglich. Dennoch kann er als Evaluationsinstrument genutzt werden.

Durch die Eindimensionalität der 9-Item-Skala ist es zulässig, einen Mittelwert über die Items zu bilden, welcher anschließend in das Punktesystem übertragbar ist.

Mit den Ergebnissen der Varianzanalysen zur Untersuchung des Einflusses von Motivation und Semesterabschnitt auf das Veranstaltungsurteil konnte gezeigt werden, dass ein solcher Einfluss lediglich für die Motivation der Teilnehmer besteht. Es zeigte sich, dass Vorlesungen als Wahlpflichtveranstaltungen mit 0,3 Notenpunkten (MW 2,61, SD 0,87) signifikant besser beurteilt wurden als Seminare (MW 2,92, SD 0,99) und Kurse und Praktika (MW 2,86, SD 0,92) mit Anwesenheitspflicht. Dies bestätigt die Ergebnisse Kromreys (1994), der als Ursache dafür die Selektion der besonders interessierten und motivierten Studierenden bei Wahlpflichtveranstaltungen sieht, welche positiver bewerten würden als unmotivierte Kommilitonen. Die Varianzanalyse zur Untersuchung des Einflusses des Semesterabschnittes führte für den Veranstaltungstyp Vorlesung zu einem nicht signifikanten Ergebnis und damit konnte kein Einfluss des Semesterabschnittes auf das Urteil nachgewiesen werden. Für den Veranstaltungstyp Kurse und Praktika konnte aufgrund fehlender Homogenität der Varianzen mittels Varianzanalyse ein Einfluss des Semesterabschnittes auf das Urteil weder bewiesen noch widerlegt werden.

Durch die Auswertung der freien Meinungsäußerungen konnte ein Kategoriensystem, bestehend aus 5 Hauptkategorien (1. Organisation des Unterrichts, 2. Dozent als Unterrichtender und als Person, 3. Didaktik, 4. Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz, 5. Medieneinsatz, Bild- und Arbeitsmaterialien) mit jeweils positiver und negativer Ausprägung sowie 24 positiven und 34 negativen Subkategorien erarbeitet werden. Das Kategoriensystem ermöglichte eine reliable Zuordnung der Meinungsäußerungen. Cohens kappa zeigte mit Werten von 0,76 (Intraraterreliabilität) und 0,80 (Interraterreliabilität) für die Zuordnung der Aussagen zu Hauptkategorien eine gute Reliabilität. Bei den Subkategorien kam es sowohl zu fehlender bis vollkommener Übereinstimmung der Zuordnungen der Rater. Den Aussagen wurde insgesamt 3752-mal eine Kategorie zugeordnet. Dabei entfielen für die gesamte Stichprobe 1620 (43,1%) auf positive Kategorien und 2134 (56,9%) auf negative Kategorien. Für die Frequenzanalysen wurden die positiven und negativen Kategorien getrennt voneinander betrachtet. Es war festzustellen, dass negative Aspekte von den Studierenden häufiger genannt wurden, wobei die Kritik an der Didaktik und

speziell an dem von Studierenden als unübersichtlichen und unsystematischen empfundenen Aufbau von Veranstaltungen überwog. Auch die Organisation des Unterrichts wurde häufig kritisiert. Positiv hervorgehoben wurden die Themen und Inhalte sowie der Dozent als Unterrichtender und als Person. Zusammenfassend konnte festgestellt werden, dass das Thema einer Veranstaltungen, die Lehrpräsentation durch den Dozenten und die allgemeine Didaktik als Aspekte der Prozessqualität sowie die Organisation der Lehre als Merkmal der Strukturqualität bei der Beurteilung von Veranstaltungen besonders stark von den Studierenden wahrgenommen und entsprechend häufig beurteilt wurden.

Abschließend ist festzustellen, dass der in der vorliegenden Arbeit überprüfte Marburger Evaluationsbogen unter Verwendung der 9-Item-Skala zur Messung eines „Allgemeinen Index der Lehrqualität“ als Evaluationsinstrument über eine hinreichende Messgenauigkeit verfügt. Er erfüllt damit die Voraussetzungen, um einen Beitrag im Prozess der kontinuierlichen Verbesserung der Qualität in der Lehre zu leisten und als ein Rational für die leistungsbezogene Mittelvergabe herangezogen zu werden.

## 9. Abstract

Ziel der vorliegenden Arbeit war es, den Beitrag von Evaluationen zur Qualitätsverbesserung in der Hochschullehre zu diskutieren und vor diesem Hintergrund den Marburger Fragebogen zur Evaluation des Lehrangebotes in der Medizin im empirisch-statistischen Teil der Arbeit zu untersuchen. Die Überprüfung der Dimensionalität der Skalen zur Struktur- und Prozessqualität, der Reliabilität des Fragebogens und die Entwicklung eines Kategoriensystems zur reliablen Einordnung der freien Meinungsäußerungen sowie deren inhaltliche Analyse waren dabei die Schwerpunkte. Zusätzlich sollte der Einfluss so genannter Bias-Variablen auf das Urteil der Studierenden überprüft werden.

Der Marburger Fragebogen wurde als Instrument zur Evaluation des Lehrangebotes in der Medizin im Sommersemester 2002 erstmalig eingesetzt. Getrennt nach Veranstaltungstypen (Vorlesung, Kurs/ Praktikum, Seminar) und bestehend aus offenen sowie geschlossenen Fragen sollte er eine differenzierte Beurteilung von Lehrstruktur und Lehrprozess ermöglichen. Als Datenbasis der vorliegenden Querschnittsstudie galten die Urteile der Studierenden. Die methodische Untersuchung der geschlossenen Fragen beinhaltete die Item- und Reliabilitätsanalyse (interner Konsistenz, Reliabilitätskoeffizienten Chronbach's alpha) sowie die Dimensionalitätsprüfung mittels Faktorenanalyse (Hauptkomponentenanalyse mit Varimax-Rotation, Abbruchkriterium Eigenwert  $< 1$ ). Zur Untersuchung der drei Bias-Variablen Häufigkeit des Veranstaltungsbesuchs, Teilnehmermotivation und Semesterzahl wurden einfaktorielle Varianzanalysen (ANOVA) und der U-Test nach Mann und Whitney durchgeführt. Die Reliabilität des Kategoriensystems wurde mittels Intra- und Interraterreliabilität (Reliabilitätsindex Cohens Kappa) überprüft. Zur inhaltlichen Untersuchung wurden Frequenzanalysen sowohl für die gesamte Stichprobe als auch getrennt für vorklinischen und klinischen Abschnitt durchgeführt.

Die Ergebnisse der deskriptiv-statistischen Untersuchung der ursprünglichen 12-Item-Skala zeigten zwar eine zufriedenstellende interne Konsistenz und die Eindimensionalität der Skala zur Prozessqualität, jedoch fielen die Items SQ 3 (*Qualität der Unterrichtsmaterialien*), SQ 5 (*Ausfall der Veranstaltung*) und SQ 6 (*Anwesenheit der Studierenden in der Veranstaltung bzw. Beitrag des Semesterhilfskräfte*) durch

mangelhafte Trennschärfe und Ladung auf separate Faktoren in der Faktorenanalyse auf. Nach Elimination dieser Items war bei Analyse der daraus resultierenden 9-Item-Skala festzustellen, dass alle 9 Items auf einen gemeinsamen Faktor luden und folglich das gleiche Konstrukt messen, welches als „Index der allgemeinen Lehrqualität“ bezeichnet wurde. Die getrennte Erfassung von Prozess- und Strukturqualität ist mit dem Marburger Evaluationsbogen nach den vorliegenden Ergebnissen nicht möglich.

Anhand der freien Meinungsäußerungen konnte ein Kategoriensystem, bestehend aus 5 Hauptkategorien (1. Organisation des Unterrichts, 2. Dozent als Unterrichtender und als Person, 3. Didaktik, 4. Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz, 5. Medieneinsatz, Bild- und Arbeitsmaterialien) mit jeweils positiver und negativer Ausprägung und dazugehörigen Subkategorien erarbeitet werden. Durch dieses Kategoriensystem war eine reliable Zuordnung der Meinungsäußerungen möglich. Die Frequenzanalysen zeigten, dass negative Aspekte häufiger genannt wurden und dabei die Kritik an der Didaktik (Kategorie 3), speziell an dem als unübersichtlichen und unsystematischen empfundenen Veranstaltungsaufbau überwog. Positiv hervorgehoben wurden die Themen, Inhalte (Kategorie 4) und der Dozent (Kategorie 2). Zusammenfassend konnte festgestellt werden, dass das Thema einer Veranstaltung, die Lehrpräsentation durch den Dozenten und die allgemeine Didaktik als Aspekte der Prozessqualität sowie die Organisation der Lehre als Merkmal der Strukturqualität bei der Beurteilung von Veranstaltungen besonders stark von den Studierenden wahrgenommen und entsprechend häufig beurteilt wurden. Bei der Untersuchung der drei Bias-Variablen konnte mittels Varianzanalyse lediglich ein Einfluss der Teilnehmermotivation auf das Veranstaltungsurteil nachgewiesen werden, denn es zeigte sich, dass Vorlesungen als Wahlpflichtveranstaltungen signifikant besser beurteilt wurden als anwesenheitspflichtige Veranstaltungen.

Abschließend ist festzustellen, dass der in der vorliegenden Arbeit überprüfte Marburger Evaluationsbogen unter Verwendung der 9-Item-Skala zur Messung eines „Allgemeinen Index der Lehrqualität“ als Evaluationsinstrument über eine hinreichende Messgenauigkeit verfügt. Er erfüllt damit die Voraussetzungen, um einen Beitrag im Prozess der kontinuierlichen Verbesserung der Qualität in der Lehre zu leisten und als ein Rational für die leistungsbezogene Mittelvergabe herangezogen zu werden.



## 10. Literaturverzeichnis

Arnold, E. (2003). Qualitätsentwicklung im Bereich der Lehre: Anforderungen und Strategien auf der Ebene der Fachbereiche.

Quelle: <http://evanet.his.de/evanet/forum/pdf-position/ArnoldPosition.pdf>

Basler, H.-D., Krebs, K. (2003). Marburger Fragebogen zur Evaluation des Lehrangebots in der Medizin. In: Zeitschrift für medizinische Psychologie 3/2003.

S. 99-103

Berger, U., Schleußner, C., Strauß, B. (2003). Umfassende Lehrevaluation in der Medizin – eine Aufgabe für psychosoziale Fächer? In: Psychotherapie Psychosomatik Medizinische Psychologie 2/ 2003. Stuttgart: Thieme Verlag. S. 71-78

Block, H.-J. (2000). Vorwort. In: HRK (Hrsg.): „...und im Sterben immer der Erste“. Beiträge zur Hochschulpolitik 5/ 2000. Bonn: HRK

Bortz, J., Döring N. (2002). Forschungsmethoden und Evaluation. (3. Auflage). Berlin: Springer-Verlag

Bundesgesetzblatt, Teil 1, Nr. 44, 3. Juli 2002, S. 2405 – 2436.

Carstensen, D. (1997). Gemeinsamkeiten und Unterschiede der Evaluationspraxis in den Ländern des Bildungsnetzwerkes Nordwesteuropas. In: HIS Kurzinformation. A12/97: Wirksamkeit der internen und externen Evaluation von Lehre und Studium. Mit Berichten aus den Ländern Bremen, Flandern, Niederlande, Niedersachsen und Nordrhein-Westfalen. Hanover: S. 42-48

Dembski, M.; Lorenz, T. (1995). Zertifizierung von Qualitätsmanagementsystemen bei Bildungsträgern. (2. Auflage) Renningen-Malmsheim: expert-Verlag

Deming, W.E. (1986). Out of the Crisis. (2. Auflage). Cambridge/ Mass.: Massachusetts Institute of Technology (MIT) Press 1986

DIN Deutsches Institut für Normierung e.V. DIN 55 350 (1987). Teil 11: Begriffe zu Qualitätsmanagement und Statistik – Grundbegriffe des Qualitätsmanagements. Berlin

DIN Deutsches Institut für Normierung e.V. DIN EN ISO 8402 (1995). Qualitätsmanagement. Begriffe. Berlin

Donabedian, A. (1980). Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration Press

Donabedian, A. (2003). An introduction to quality assurance in health care. Oxford: Oxford University Press.

El Hage, N. (1996). Lehrevaluation und studentische Veranstaltungskritik. Projekte, Instrumente und Grundlagen. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (Hrsg.). Bonn.

Fischer-Bluhm, K. (2000). Qualitätsentwicklung als Antwort auf die Individualisierung im Bildungsbereich – am Beispiel der Hochschulen. In: Forum Bildung (Hrsg.): Erstkongress des Forum Bildung am 14. und 15. Juli 2000 in Berlin. Bonn. S.680-693

Frederiks, M. (2001). Qualitätssicherung in Großbritannien. Beiträge zur Hochschulpolitik 10/ 2001. Bonn: HRK

Hannen, C.; Lindemann, T. (1996). Qualitätsmanagement für die Praxis. Berlin: Cornelsen Verlag

Hegewald, B.; Reuke, H.; Carstensen, D.; Reisert, R. (1997). Wirksamkeitsanalyse zur Evaluation von Lehre und Studium an niedersächsischen Hochschulen. In: HIS-Kurzinformation A12/97. Wirksamkeit der internen und externen Evaluation von Lehre und Studium. Mit Berichten aus den Ländern Bremen, Flandern, Niederlande, Niedersachsen und Nordrhein-Westfalen. Hannover. Hannover. S. 25-30

Heid, H. (2000). Qualität. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. Zeitschrift für Pädagogik. Beiheft 41. S.41-51

HRK (1995). Zur Evaluation im Hochschulbereich unter besonderer Berücksichtigung der Lehre. Entschließung des 176. Plenums der HRK vom 3. Juli 1995. Bonn: HRK

HRK (2000). Wegweiser 2000 durch die Qualitätssicherung in Lehre und Studium. Dokumente & Informationen 2/2000. Bonn: HRK

HRK (2003). Wegweiser 2003. Qualitätssicherung an Hochschulen. Sachstandbericht und Ergebnisse einer Umfrage des Projektes Qualitätssicherung. Beiträge zur Hochschulpolitik 7/ 2003. Bonn: HRK

Jaster, H.J. (1997). Von der Bedeutung der Qualität in der Industrie zum Qualitätsbegriff in der Medizin. In: Jaster, H.J. (Hrsg.): Qualitätssicherung im Gesundheitswesen. Stuttgart, New York, S. 9-33

Kamiske, G.F.; Brauer, J.-P. (1995). Qualitätsmanagement von A-Z: Erläuterungen moderner Begriffe des Qualitätsmanagements (2. Auflage). München, Wien: Carl Hanser Verlag

Kern, H. (2000). Rückgekoppelte Autonomie. Steuerungselemente in lose gekoppelten Systemen. In: Hanft, A. (Hrsg.). Hochschulen managen? Zur Reformierbarkeit der Hochschulen nach Managementprinzipien. Neuwied: Luchterland-Verlag. S.25-38

Kromrey, H. (1994). Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In: Mohler, P.P. (1994) Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung. Münster: Waxmann-Verlag, S. 91-114

Kromrey K. (2000). Qualität und Evaluation im System Hochschule. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen: Leske + Budrich-Verlag. S.233-256

Kromrey, H. (2001). Studierendenbefragung als Evaluation der Lehre? – Anforderungen an Methodik und Design. In U. Engel (Hrsg.): Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre. Frankfurt: Campus. S.11-47

Landfried, K. (2001) Impulsreferat. In: HRK (Hrg.): Wettbewerb – Profilbildung - Evaluation. Qualitätssicherung von Lehre und Studium in Gegenwart und Zukunft. Beiträge zur Hochschulpolitik 6/2001. Bonn: HRK. S. 53-57

Lienert, A. (1989). Testaufbau und Testanalyse. (4.Auflage) München, Weinheim: Psychologie Verlags Union

Lüthje, J. (2000). Systemische Qualitätsentwicklung durch Evaluation. In Laske, S., Habersam, M., Kappler, E. (Hrsg.) Qualitätsentwicklung in Universitäten. Konzepte, Prozesse, Wirkungen. München, Mering: Rainer Hampp Verlag 2000. S.119-133

Mertens, D.M. (1998). Research methods in education and psychology: Integration diversity with quantitative and qualitative approaches. Thousand Oaks, CA: Sage

Mittag, S.; Bornmann, L.; Daniel, H.-D. (2003). Evaluation von Studium und Lehre an Hochschulen. Handbuch zur Durchführung mehrstufiger Evaluationsverfahren. Münster: Waxmann Verlag

Moewes, M. (2002). Rahmenbedingungen für QM an Hochschulen.

Quelle: <http://www.hs-bremen.de/Uploaded/Eintrag24408/homepageaufsatz1.doc>

Multrus, F. (1995). Zur Lehr- und Studienqualität. Dimensionen, Skalen und Befunde des Studierendensurveys. Hefte zur Bildungs- und Hochschulforschung, Nr. 12. Arbeitsgruppe Hochschulforschung. Universität Konstanz

Müller-Böling, D.; Schreiterer, U. (1999). Hochschulmanagement durch Zielvereinbarungen. Perspektiven eines neuen Steuerungsinstruments. In: Fredrowitz, J./ Krasny, K./ Ziegele, F. (Hrsg.). Hochschulen und Zielvereinbarungen – neue Perspektiven Autonomie. Gütersloh: Bertelsmann. S. 9-25

Oess, A. (1993). Total quality management: Die ganzheitliche Qualitätsstrategie. (3.Auflage) Wiesbaden: Gabler Verlag. S. 59-99

Pepels, W. (1996). Qualitätscontrolling bei Dienstleistungen. München: Vahlen

Pfitzinger, E. (1998). Der Weg von DIN EN ISO 9000 ff zu Total Quality Management (TQM). Berlin: Beuth

Reinhardt, G.; Lindemann, U.; Heinzl, J. (1996). Qualitätsmanagement. Ein Kurs für Studium und Praxis. Berlin, Heidelberg, New York: Springer Verlag

Rindermann, H. (1999). Die studentische Beurteilung von Lehrveranstaltungen – Forschungsgegenstand und Implikation. Eingeladener Gastvortrag auf dem Symposium Evaluierung an der Universität - zwischen Qualitätsmanagement und Selbstzweck. Graz. 02.12.1999

Rindermann, H. (2001). Lehrevaluation – Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen. Landau: Verlag Empirische Pädagogik

Ritz, A. (2003). Evaluation von New Public Management. Grundlagen und empirische Ergebnisse der Bewertung von Verwaltungsreformen in der schweizerischen Bundesverwaltung. Bern, Stuttgart, Wien: Paul Haupt Verlag. S.25-95

Rossi, P.H.; Freeman, H.E.; Hofmann, G. (1988). Programmevaluation: Einführung in die Methoden angewandter Sozialforschung. Stuttgart . Enke-Verlag

Rossi, P.H.; Freeman, H.E.; Hofmann, G.; Lipsey M.W. (1999). Evaluation: A Systematic Approach (6. Auflage). Thousand Oaks, London, New Dehli

Rust, H. (1981). Methoden und Probleme der Inhaltsanalyse. Eine Einführung. Literaturwissenschaft im Grundstudium Bd. 13. Tübingen: Gunter Narr Verlag

Rustemeyer, R.(1992). Praktisch-methodische Schritte der Inhaltsanalyse. Münster: Aschaffendorffsche Verlagsbuchhandlung

Schenker-Wicki, A. (1996). Evaluation von Hochschulleistungen: Leistungsindikatoren und Performance measurements. Wiesbaden. Deutscher Universitäts-Verlag GmbH

Schilling, G. (1994). Qualitätsbeurteilung und Verbesserung von Lehre und Studium – eine aktuelle Aufgabe. In: HRK (Hrsg.): Qualität von Studium und Lehre. Dokumente zur Hochschulreform. 91/1994. Bonn: HRK. S.13-23.

Schreier, G. (2000). Voneinander lernen – Einige Bemerkungen zum Stand der Qualitätssicherung in Lehre und Studium. In: HKR (Hrsg.): Voneinander lernen. Hochschulübergreifende Qualitätssicherung in Netzwerken und Verbänden. 2. Nationales Expertenseminar zur Qualitätssicherung. Bonn, 23./24. September 1999. Beiträge zur Hochschulpolitik 7/2000. Bonn: HRK. S. 11-19

Spiel, C.; Gössler, P.M. (2000) Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende. Zeitschrift für Pädagogische Psychologie 2000; 14. S. 38-47

Staropoli, A. (1995). Evaluation im Zentralstaat: Das französische Modell. In: Müller-Böhlting, D. (Hrsg.): Qualitätssicherung in Hochschulen. Gütersloh: Verlag Bertelsmann Stiftung

Stockmann, R. (2000). Evaluation in Deutschland. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen: Leske + Budrich-Verlag. S.11-35

Voegelin, L. (2001). Qualitätsverbesserung in der Lehre. In: HRK (Hrsg.): Auf dem Weg zum Qualitätsmanagement. Erfahrungen und Perspektiven. 2. Arbeitstag zur Evaluation an Hochschulen. Berlin. 12./13. Februar 2001. Beiträge zur Hochschulpolitik 14/2001. Bonn: HRK. S.33-38

Weiss, C. (2000) Begrüßung. In: HRK (Hrsg.). Leitbild der Hochschule. Qualität der Lehre. Beiträge zur Hochschulpolitik 2/2002. Bonn: HRK

Widmer, T. (2000). Qualität der Evaluation. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen: Leske + Budrich-Verlag. S.77-101.

Willems, J.; Gijssels, W.; de Brie D. (1994). Qualitätssorge in der Lehre. Leitfaden für die studentische Lehrevaluation. Hochschulwesen Wissenschaft und Praxis. Roland Richter (Hrsg.) Neuwied, Kriftel, Berlin: Luchterhand Verlag;

Wissenschaftsrat (1996). Empfehlung zur Stärkung der Lehre durch Evaluation. Berlin: Wissenschaftsrat

ZevA (2003). Zentrale Evaluations- und Akkreditierungsinstitution Hannover.

<http://www.zeva.uni-hannover.de/index.htm>

## **10. Anhang**

Anhang A: Marburger Fragebogen zur Evaluation des Lehrangebots in der Medizin

Anhang B: Kategoriensystem zur Einordnung der Meinungsäußerungen

Anhang C: Intra-Rater-Reliabilität der Subkategorien

Anhang D: Inter-Rater-Reliabilität der Subkategorien

Anhang E: Frequenzanalyse der gesamten Stichprobe; Subkategorien

Anhang F: Frequenzanalyse der Gruppen; Hauptkategorien

Anhang G: Frequenzanalyse der Gruppen; Subkategorien

Anhang H: Lebenslauf, Verzeichnis der akademischen Lehrer; Danksagung











## Anhang B

### Kategoriensystem zur Einordnung der Meinungsäußerungen

---

Die Kategorien wurden in fünf thematisch unterschiedliche Hauptkategorien unterteilt. Jeder Hauptkategorie sind Subkategorien untergeordnet. Nur Kategorie 4 besteht lediglich aus der Hauptkategorie ohne Subkategorien aufzuweisen. Je nach Inhalt werden die Aussagen den positiven oder den negativen Kategorien zugeordnet.

Hauptkategorie 1: Organisation des Unterrichts

Hauptkategorie 2: Dozent als Unterrichtender und als Person

Hauptkategorie 3: Didaktik

Hauptkategorie 4: Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz

Hauptkategorie 5: Medieneinsatz, Bild- und Arbeitsmaterialien

<b>Hauptkategorie +1: Organisation des Unterrichts</b>
--

Organisation und Ablauf des Unterrichts werden von den Studierenden positiv bewertet. Dabei lassen sich folgende Subkategorien bilden.

#### **Kategorie +1.1.**

##### **Räumlichkeiten**

Für den Unterricht stehen angemessene Räume (entsprechende Größe, Helligkeit, Belüftung, Akustik) zur Verfügung.

#### **Kategorie +1.2.**

##### **Zeitliche Aspekte**

Die Veranstaltung ist zeitlich gut organisiert, d.h. sie findet weder zu früh noch zu spät statt und es besteht die Möglichkeit für Pausen bei längeren Veranstaltungen. Es wird ausreichend Zeit für die Experimente bzw. Patientenbesprechung eingeräumt.

**Kategorie +1.3.****Gruppengröße**

Die Studierenden empfinden die Größe der Gruppe angemessen. Es ist keine Massenveranstaltung. Sie haben die Möglichkeit sich in den Unterricht einzubringen.

**Kategorie +1.4.****Abstimmung der Dozenten untereinander (inhaltlich/ terminlich)**

Bei Wechsel der Dozenten wird das Konzept der Veranstaltung beibehalten. Es kommt nicht zur Wiederholung bereits behandelter Themen. Wenn der Dozent verhindert ist, wird er durch einen entsprechend vorbereiteten Kollegen vertreten.

**Kategorie +1.5.****Personelle Kontinuität auf Dozentenseite**

Besonders positiv bewerten die Studierenden, dass nur ein Dozent durch die gesamte Veranstaltung führt. Sie können sich auf die Vortragsweise und die Anforderungen des Dozenten – auch in Prüfungssituationen – einstellen. In ihm haben sie einen gleichbleibenden Ansprechpartner, der für die Veranstaltung verantwortlich ist.

**Kategorie +1.6.****Platzierung und Gewichtung der Veranstaltung**

Für die Studierenden ist es wichtig entsprechend ihrem Kenntnisstand Veranstaltungen besuchen zu können. Das Curriculum ist im Semester entsprechend organisiert. Die hier aufgeführten Veranstaltungen sind sowohl zeitlich als auch inhaltlich für das entsprechende Semester gut auf die Studierenden abgestimmt. Die Termine entsprechen in Anzahl und Dauer den Erwartungen der Studierenden.

**Hauptkategorie +2: Dozent als Unterrichtender und als Person**

In dieser Kategorie wird der Dozent nach seinen Fähigkeiten als Unterrichtender und seinem Auftreten beurteilt. Gelingt es ihm, eine positive Atmosphäre zu schaffen, wird der Unterricht oftmals als effektiv beschrieben.

**Kategorie +2.1.****Dozent als Unterrichtender**+2.1.1. Dozent schafft eine vertrauensvolle und sachliche Atmosphäre

Dem Unterrichtenden gelingt es, eine für die Studierenden entspannte Lern –und Arbeitsatmosphäre zu schaffen. Sie werden weder von ihm vor Kommilitonen bloß gestellt, noch wird Druck oder Zwang auf sie von Seite des Dozenten ausgeübt.

+2.1.2. Guter Unterrichtsstil

Als positiv erleben es die Studierenden, wenn sie aktiv am Unterricht beteiligt werden, Raum für Diskussion gelassen wird und sie die Möglichkeit haben, Fragen zu stellen. Der Dozent nimmt sich Zeit, um auf die gestellten Fragen zu antworten.

+2.1.3. Effizienter Unterricht

Der Dozent versteht es, den Unterrichtsstoff verständlich zu erklären. In diese Kategorie fällt auch die allgemeine Aussage, dass der Dozent „gut“ sei. Die Studierenden fühlen sich nicht durch die Anforderungen/ Fragen überfordert und geben den Lerneffekt als groß an.

**Kategorie +2.2.****Dozent als Person**+2.2.1. Dozent ist motiviert

Der Dozent wird von den Studierenden als motiviert und am Unterrichten interessierte wahrgenommen.

#### +2.2.2. Spaß am Fach und am Unterrichten

Dem Dozenten ist es gelungen durch Spaß am Fach und am Unterrichten, die Studierenden für das Thema zu interessieren. Er wird als fachliche Autorität anerkannt.

#### +2.2.3. Angenehmes Auftreten

Höfliche Dozenten stoßen auf Seite der Studierenden auf besonders positive Resonanz. Persönliche Sympathie als auch adäquater Umgang mit den Patienten fallen in diese Kategorie.

#### +2.2.4. Humor

In dieser Kategorie wird die humorvolle Art des Dozenten hervorgehoben. Sie wird als positiv gewertet, da sie die Atmosphäre auflockert.

<b>Hauptkategorie +3: Didaktik gut</b>
--

In dieser Kategorie beurteilen die Studierenden die Didaktik. Dabei wird hervorgehoben, dass Art und Weise der Unterrichtsgestaltung und der Wissensvermittlung die Studierenden anspricht.

#### **Kategorie +3.1.**

##### **Aufbau der Veranstaltung systematisch und übersichtlich**

Die Studierenden beurteilen das in sich geschlossene Konzept der Veranstaltung positiv. Die Themen sind zu Beginn der Veranstaltung bekannt. Es kommt nicht zu Wiederholungen. Inhalte von Kurs und begleitender Vorlesung sind aufeinander abgestimmt. Eine Einführungsveranstaltung, die über Inhalte und Ablauf informiert, wird von Seiten der Studierenden begrüßt.

#### **Kategorie +3.2.**

##### **Veranstaltung verständlich und praxisnah**

Der Unterricht wird durch Patientenvorstellungen, Fallbeispiele und Experimente anschaulich und praxisnah gestaltet. Zusammenhänge werden verdeutlicht.



**Kategorie +3.3.****Patientenvorstellung positiv**

Die Patientenvorstellungen und Fallbeispiele werden in dieser Kategorie als besonders positiv beurteilt. Studierenden haben die Möglichkeit, selbst im Unterricht aktiv zu sein, d.h. die Anamnese zu erheben und den Patienten körperlich zu untersuchen.

**Kategorie +3.4.****Aufgaben, Experimente und praktische Übungen positiv**

Das Lösen von Aufgaben und die Durchführung von Experimenten bewerten die Studierenden als geeignet, Lehrinhalte zu verstehen und zu vertiefen bzw. zu üben.

**Kategorie +3.5.****Dem Wissensstand der Studierenden angepasste Vermittlung der Lehrinhalte**

Die Studierenden können der Veranstaltung inhaltlich folgen, denn sie fühlen sich von der Stoffmenge, der Vortragsgeschwindigkeit und dem Niveau der Veranstaltung nicht überfordert.

**Kategorie +3.6.****Konzentration auf das Wesentliche**

In der Veranstaltung wird auf eine effiziente, kompakte und auf das Wesentliche konzentrierte Wissensvermittlung Wert gelegt. Die Studierenden beurteilen das Fehlen von Exkursen als positiv.

**Kategorie +3.7.****Klausur positiv**

Die Studierenden wünschen sich in dem bewerteten Fach eine Klausur. Diese wird hier im positiven Sinne als Anreiz zum Lernen und zum Üben von IMPP-Fragen betrachtet.

**Hauptkategorie +4:****Themen, Inhalte interessant und von Prüfungs- und Praxisrelevanz**

In dieser Kategorie schildern die Studierenden, dass sie mit den Inhalten der Veranstaltung zufrieden sind bzw. sie diese interessant finden. Neben dem pauschalen Urteil, dass die Veranstaltung „gut“ sei, werden Prüfungs- und Praxisrelevanz hervorgehoben.

**Hauptkategorie +5: Medieneinsatz, Bild –und Arbeitsmaterialien**

Die Studierenden bewerten in dieser Kategorie den Medieneinsatz in der Veranstaltung als gelungen.

**Kategorie +5.1.****Gute Qualität der Materialien**

Die gezeigten Folien, Dias, Videos und PC-gestützten Präsentationen sind von überzeugender Qualität und Aktualität. Schriftart und Schriftgröße, farbige Gestaltung, Diagramme und Bilder sind optisch ansprechend.

**Kategorie +5.2.****Vielfältiger und sinnvoller Einsatz**

Hier wird hervorgehoben, dass die Medien vielfältig und häufig eingebaut werden und den Unterricht sinnvoll ergänzen und damit zum besseren Verständnis der Inhalte beitragen.

**Kategorie +5.3.****Gute Skripte/ Handouts**

Die Studierenden begrüßen in ausreichender Menge zur Verfügung gestellte Skripte oder entsprechende Handouts, wobei sie deren Gestaltung und Inhalt als hilfreich zum

Nacharbeiten und Lernen empfinden. Der Lernstoff wird kompakt und verständlich präsentiert, die weiterführenden Abbildungen ergänzen das Konzept sinnvoll.

### **Hauptkategorie –1: Organisation des Unterrichts negativ**

#### **Kategorie -1.1.**

##### **Räumlichkeiten**

Die räumlichen Gegebenheiten waren nicht zufriedenstellend. Die Räume werden als zu eng beschrieben oder bieten eine schlechte Akustik. Auch fehlende Räumlichkeiten oder dem Zweck nicht angemessene Unterrichtsorte (z.B. Flur einer Station in der Klinik) fallen in diese Kategorie.

#### **Kategorie -1.2.**

##### **Zeitliche Aspekte**

###### -1.2.1. Zeitliche Organisation negativ

Die Studierenden kritisieren an der zeitlichen Organisation einer (Pflicht-)Veranstaltung die häufigen Terminüberschneidungen mit anderen Veranstaltungen, welche sie dann nicht besuchen können. Zusätzlich bemängeln sie den häufigen Ausfall des Unterrichts.

###### -1.2.2. Ungünstiger Zeitpunkt

Die Veranstaltung findet entweder zu früh, zu spät oder in der Mittagszeit statt, weshalb sie weniger frequentiert wird.

###### -1.2.3. Lange Wartezeiten

Die Studierenden beschreiben in dieser Kategorie ihre Unzufriedenheit über die z.T. erhebliche Verzögerung mit der eine Veranstaltung beginnt oder den Ausfall der Veranstaltung. Als Gründe werden folgende Punkte angegeben: Dozent nicht anwesend/ muss erst informiert werden/ keine zuständige Person auffindbar.

**Kategorie -1.3.****Gruppengröße**

Die Gruppe ist zu groß und der Unterricht gewinnt den Charakter einer Massenveranstaltung. Den Studierenden fehlt die Möglichkeit, sich in den Unterricht einzubringen.

**Kategorie -1.4.****Abstimmung der Dozenten untereinander (inhaltlich/ terminlich)**

Die Veranstaltung, die von wechselnden Dozenten gehalten wird, ist in sich nicht stimmig, d.h. ein roter Faden fehlt. Es kommt häufig zu Wiederholungen bereits besprochener Inhalte oder zur Auslassung von Themen. Ist der Dozent verhindert, wird er von weniger vorbereiteten Kollegen vertreten, die u.U. nicht über ihre Vertretungsaufgabe vorab informiert wurden.

**Kategorie -1.5.****Fehlende personelle Kontinuität auf Dozentenseite**

Ein ständiger Wechsel der Dozenten wird negativ gewertet. Die Studierenden können sich nicht auf den Dozenten, seine Vortragsweise und Anforderungen einstellen. Das Niveau des Unterrichts ist nicht konstant.

**Kategorie -1.6.****Platzierung und Gewichtung der Veranstaltung****-1.6.1. Platzierung negativ**

Die Studierenden empfinden die Veranstaltung mit ihren Inhalten als zu früh für ihren Wissensstand im Curriculum.

**-1.6.2. Zu wenige Termine**

Das Angebot an Terminen und deren vorgegebene Dauer erfüllen nicht die Erwartungen. Die Studierenden wünschen sich hier mehr Termine und mehr Zeit in der

entsprechenden Veranstaltung, um Aufgaben adäquat zu lösen, um Patienten genau zu untersuchen und um deren Krankheitsbilder zu diskutieren.

#### -1.6.3. Zu viele Termine

Im Gegensatz zu der oben beschriebenen Kategorie sind Anzahl und Dauer der Termine in einem Fach oder zu einem Thema zu zeitintensiv. Zu viele Wiederholungen verursachen hier Langeweile. Es besteht der Wunsch nach Pausen bei mehrstündigen Veranstaltungen.

### **Hauptkategorie –2: Dozent als Unterrichtender und als Person negativ**

#### **Kategorie -2.1.**

##### **Dozent als Unterrichtender**

#### -2.1.1. Unangenehme Atmosphäre

Es herrscht eine unruhige und angespannte Unterrichtsatmosphäre. Die Studierenden fühlen sich nicht akzeptiert oder bei Nichtwissen einer Antwort vor Kommilitonen bloßgestellt. Der Dozent vermittelt den Eindruck, dass die Studierenden ihn stören („Dozent hat keine Zeit...“). Sie trauen sich nicht, Fragen zu stellen.

#### -2.1.2. Unterrichtsstil negativ

Die Studierenden wünschen sich, in den Unterricht aktiv mit einbezogen zu werden und die Möglichkeit Fragen zu stellen. Frontalveranstaltungen oder der vom Dozenten gepflegte Monologstil schließt sie aus der Veranstaltung aus.

#### -2.1.3. Unterricht nicht effektiv

Es gelingt dem Dozenten nicht, den Stoff verständlich zu erklären und zu vermitteln oder es werden Inhalte besprochen, die die Erwartungen der Studierenden nicht erfüllen. Der Dozent ist schlecht vorbereitet und die Studierenden geben den Lerneffekt als gering an.

#### -2.1.4. Überforderung in der Unterrichtssituation

Der Dozent ist aus Sicht der Studierenden mit der Aufgabe des Unterrichtens überfordert. Als Gründe dafür geben sie an:

- Doppelbelastung Unterricht/ Stationsdienst
- Müdigkeit, da so eben der Nachtdienst beendet wurde
- Mangel an fachlicher Kompetenz, Erfahrung, Planung im Unterricht

Die Studierenden wünschen einen Leitfaden für Dozenten, in dem festgelegt ist, welche Inhalte den Studierenden zu vermitteln sind.

## **Kategorie -2.2.**

### **Dozent als Person**

#### -2.2.1. Unmotivierter Dozent

Fehlende Motivation auf Dozentenseite schlägt sich auf die Unterrichtsqualität nieder. Er gibt sich wenig Mühe, die Studierenden für das Fach zu motivieren und interessieren. Die Studierenden beschreiben hier den Dozenten als uninteressiert und gelangweilt beim Unterricht.

#### -2.2.2. Unangemessenes Verhalten

In dieser Kategorie werden Verhaltensweisen des Dozenten beschrieben, die den Studierenden negativ aufgefallen sind. Neben persönlicher Antipathie wird „arrogantes“ oder „herrschaftliches“ Verhalten den Studierenden gegenüber aufgezählt. Der Situation nicht angemessenes Verhalten bei Patientenvorstellungen (Mangel an Einfühlungsvermögen und Taktgefühl) wird besonders scharf kritisiert.

### **Hauptkategorie –3: Didaktik nicht gut**

Diese Kategorie fasst Meinungsäußerungen der Studierenden zusammen, die allgemein die Unzufriedenheit mit der Art und Weise der Unterrichtsgestaltung und Wissensvermittlung zum Ausdruck bringen.

## **Kategorie -3.1.**

### **Unübersichtlicher, unsystematischer Aufbau der Veranstaltung**

Die Veranstaltung lässt ein durchgehendes Konzept vermissen. Themen bauen nicht aufeinander auf, es kommt zu Wiederholungen, Kurs und begleitende Vorlesung sind nicht aufeinander abgestimmt. Studierende wünschen eine klare Gliederung des

Unterrichts oder eine Einführungsveranstaltungen zur Bekanntgabe der Themen vor Kursbeginn.

### **Kategorie -3.2.**

#### **Veranstaltung wenig verständlich und praxisfern**

In dieser Kategorie werten die Studierenden die Veranstaltung als zu trocken und theoretisch. Es fehlt an Fallbeispielen und Patientenvorstellungen.

### **Kategorie -3.3.**

#### **Patientenvorstellung negativ**

##### -3.3.1. Patientenvorstellung zu selten/ zu kurz/ fehlt

Die Studierenden wünschen mehr Patientenvorstellungen.

##### -3.3.2. Kein Einbezug der Studierenden

Die Studierenden wünschen sich, mit in den Unterricht eingebunden zu werden, vor allem in der Form, dass sie die Anamnese erheben und körperlich untersuchen dürfen.

### **Kategorie -3.4.**

#### **Aufgaben, Experimente und Übungen negativ**

Die Studierenden beschreiben das Lösen von Aufgaben und die Experimente in Kursen als zeitraubend und nicht sinnvoll. Der Zusammenhang zwischen Aufgaben, Experimenten und Lehrinhalten bzw. Praxis wird ihnen wenig deutlich.

### **Kategorie -3.5.**

#### **Dem Wissensstand der Studierenden nicht angepasste Vermittlung der Lehrinhalte**

##### -3.5.1. Die Veranstaltung überfordert die Studierenden

Ein zu hoher Anspruch oder das Voraussetzen großen fachlichen Wissens wird negativ beschrieben. Zum Verständnis notwendige Grundkenntnisse werden im Vorfeld nicht vermittelt. Die Studierenden fühlen sich vom Stoff und seiner Vermittlung überfordert.

### -3.5.2. Zu schnelles Vorgehen oder zu viel Stoff

Die Stoffdichte im Unterricht ist zu hoch und den Studierenden fällt es schwer, dem Unterricht zu folgen. Des Weiteren wird zu schnell im Stoff vorangegangen. Es gibt kaum Zeit für schriftliche Notizen.

### -3.5.3. Zu viele Fremdwörter

Es werden zu viele noch nicht bekannte Fachbegriffe und nicht vorab erklärte Abkürzungen benutzt.

## **Kategorie -3.6.**

### **Veranstaltung zu weitschweifig**

Die Studierenden wünschen sich eine kompaktere Darstellung des Stoffgebietes ohne oberflächliche Ausschweifungen oder zu viele kleine Details.

## **Kategorie -3.7.**

### **Unterrichtsform ineffizient**

Der Veranstaltungstyp bzw. die Form in der der Unterricht stattfindet (z.B. als Seminar, praktische Übung) wird von den Studierenden als nicht geeignet zum Vermitteln, Üben und Vertiefen der Lehrinhalte eingestuft.

## **Kategorie -3.8.**

### **Klausur negativ**

Eine Klausur in der Veranstaltung wird negativ gewertet. Die Studierenden beschreiben die Wirkung der Klausur als Reduktion des Interesses am Fach auf einen reinen Scheinerwerb.

## **Kategorie -3.9.**

### **Referate negativ**

Die Studierenden beschreiben Referate im Unterricht als weniger sinnvoll, da Referate zu sehr im Vordergrund stehen und Studierende die Dozentenrolle übernehmen müssen



und zudem hochspezialisierte Themen vorgetragen werden, die die Kommilitonen überfordern.

**Hauptkategorie –4:****Themen, Inhalte uninteressant und von geringer Prüfungs- und Praxisrelevanz**

Die Themenwahl spricht die Studierenden weniger an. Sie beurteilen die Veranstaltung als uninteressant, zu speziell oder ohne bzw. von geringer Prüfungs- und Praxisrelevanz. Der Besuch der Veranstaltung „bringe“ nichts.

**Hauptkategorie –5: Medieneinsatz, Bild –und Arbeitsmaterialien nicht gut**

In dieser Kategorie beschreiben die Studierenden, dass sie die Medienwahl und deren Qualität und Einsatz im Unterricht nicht überzeugt hat.

**Kategorie -5.1.****Qualität der Materialien mangelhaft**

Die gezeigten Folie, Dias, Videos und PC-gestützten Präsentationen sind von mangelhafter Qualität, was u.U. auf ihr Alter zurückzuführen ist. Schriftart und Schriftgröße, farbige Gestaltung, Diagramme und Bilder sind nicht ansprechend oder zu überladen.

**Kategorie -5.2.****Einsatz nicht sinnvoll**-5.2.1. Einsatz zu häufig

Der Einsatz der Medien wird als zu häufig und zu sehr im Mittelpunkt stehend beschrieben. Besonders negativ bewertet werden PC-gestützte Präsentationen, die aufgrund technischer Probleme erst mit längerer Verzögerung beginnen.

-5.2.1. Einsatz zu selten

Die Studierenden wünschen, öfter Bilder oder Dias zu sehen, um Sachverhalte besser verstehen und lernen zu können.

**Kategorie -5.3.****Skripte mangelhaft oder fehlend**-5.3.1. Skripte fehlen

Die Studierenden wünschen in ausreichender Menge zur Verfügung stehende Skripte oder entsprechende „Handouts“.

-5.3.2. Mangelhafte Gestaltung und Inhalte

Gestaltung und Inhalt werden als nicht hilfreich zum Nacharbeiten und Lernen empfunden. Die Abbildungen und Diagramme sind unübersichtlich oder nicht verständlich.

**Kategorie -5.4.****Lehrmaterialien und Lernbedingungen mangelhaft**

In dieser Kategorie bewerten die Studierenden einen Mangel an Lehrmaterialien (z.B. Mikroskope, Mikroskopierkästen, Spiegel) und Arbeitsplätzen negativ.

## Anhang C

### Intra-Rater-Reliabilität der Subkategorien

**Tabelle 11: Intra-Rater-Reliabilität; Subkategorien**

Positive Subkategorien	0/0	0/1	1/0	1/1	Kappa
<b>1: Organisation des Unterrichts</b>					
+1.1.	454	-	-	-	-
+1.2.	449	1	1	3	0,75
+1.3.	454	-	-	-	-
+1.4.	454	-	-	-	-
+1.5.	453	-	-	1	1,00
+1.6.	453	-	-	1	1,00
<b>2: Dozent als Unterrichtender und als Person</b>					
+2.1.1.	447	2	2	3	0,60
+2.1.2.	450	-	2	2	0,67
+2.1.3.	354	25	5	70	0,78
+2.2.1.	432	4	5	13	0,73
+2.2.2.	445	1	4	4	0,61
+2.2.3.	447	2	1	4	0,72
+2.2.4.	447	1	3	3	0,60
<b>3: Didaktik</b>					
+3.1.	433	4	6	11	0,68
+3.2.	428	5	6	15	0,72
+3.3.	442	1	1	10	0,91
+3.4.	449	1	-	4	0,89
+3.5.	451	-	1	2	0,80
+3.6.	452	-	1	1	0,67
+3.7.	451	1	-	2	0,80
+3.8.	452	-	-	2	1,00
<b>4: Themen und Inhalte</b>	305	22	6	121	0,85
<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>					
+5.1.	453	-	-	1	1,00
+5.2.	452	-	-	2	1,00
+5.3.	449	1	-	4	0,89

Negative Subkategorien	0/0	0/1	1/0	1/1	Kappa
<b>1: Organisation des Unterrichts</b>					
-1.1.	451	-	-	3	1,00
-1.2.1.	439	1	7	7	0,63
-1.2.2.	443	-	1	10	0,95
-1.2.3.	439	2	4	9	0,74
-1.3.	450	1	-	3	0,86
-1.4.	447	3	2	2	0,44
-1.5.	419	4	5	29	0,84
-1.6.1.	432	2	1	19	0,92
-1.6.2.	436	5	-	13	0,83
-1.6.3.	435	5	-	14	0,84
<b>2: Dozent als Unterrichtender und als Person</b>					
-2.1.1.	439	2	3	10	0,79
-2.1.2.	449	-	3	2	0,57
-2.1.3.	413	19	2	20	0,63
-2.1.4.	443	4	5	2	0,30
-2.2.1.	439	2	5	8	0,69
-2.2.2.	438	3	1	12	0,85
<b>3: Didaktik</b>					
-3.1.	409	8	2	35	0,86
-3.2.	430	8	4	12	0,65
-3.3.1.	451	1	-	2	0,80
-3.3.2.	451	-	1	2	0,80
-3.4.	445	2	2	5	0,71
-3.5.1.	442	4	1	7	0,73
-3.5.2.	450	1	-	3	0,86
-3.5.3.	453	-	-	1	1,00
-3.6.	445	1	4	4	0,61
-3.7.	418	4	1	31	0,92
-3.8.	449	-	-	5	1,00
-3.9.	433	4	3	14	0,80
<b>4: Themen und Inhalte</b>	380	27	5	42	0,69

<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>					
-5.1.	451	-	2	1	0,50
-5.2.1.	453	-	-	1	1,00
-5.2.2.	449	1	3	1	0,33
-5.3.1.	449	1	2	2	0,57
-5.3.2.	450	-	-	4	1,00
-5.4.	452	1	-	1	0,67

Legende: Vierfeldertafel jeweils für positive und negative Subkategorien: 0/0 gibt die Zellengröße für die Fälle an, wo beide Rater die Kategorie für nicht gegeben hielten; 0/1 und 1/0, dass nur einer der beiden Rater die Kategorie zuordnete, und 1/1, dass beide Rater bezüglich der Kategorienzuweisung übereinstimmten; Kappa = Intraraterreliabilitätsindex für die jeweilige Subkategorie.

## Anhang D

### Inter-Rater-Reliabilität der Subkategorien

**Tabelle 12: Inter-Rater-Reliabilität; Subkategorien**

Positive Subkategorien	0/0	0/1	1/0	1/1	Kappa
<b>1: Organisation des Unterrichts</b>					
+1.1.	761	-	-	-	
+1.2.	753	1	2	5	0,77
+1.3.	760	-	-	1	1,00
+1.4.	760	-	-	1	1,00
+1.5.	760	-	-	1	1,00
+1.6.	760	-	-	1	1,00
<b>2: Dozent als Unterrichtender und als Person</b>					
+2.1.1.	751	1	2	7	0,82
+2.1.2.	743	3	3	12	0,80
+2.1.3.	628	27	8	98	0,82
+2.2.1.	726	6	1	28	0,88
+2.2.2.	753	1	5	2	0,40
+2.2.3.	750	3	2	6	0,70
+2.2.4.	753	1	-	7	0,93
<b>3: Didaktik</b>					
+3.1.	710	6	4	41	0,88
+3.2.	719	9	7	26	0,75
+3.3.	756	1	1	3	0,75
+3.4.	752	1	1	7	0,87
+3.5.	755	4	2	-	-
+3.6.	750	4	2	5	0,62
+3.7.	759	-	-	2	1,00
+3.8.	759	1	1	-	-
<b>4: Themen und Inhalte</b>	541	17	8	195	0,92
<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>					
+5.1.	749	2	-	10	0,91
+5.2.	754	1	1	5	0,83
+5.3.	740	2	1	18	0,92

Negative Subkategorien	0/0	0/1	1/0	1/1	Kappa
<b>1: Organisation des Unterrichts</b>					
-1.1.	755	-	-	6	1,00
-1.2.1.	740	2	8	11	0,68
-1.2.2.	744	1	2	14	0,90
-1.2.3.	745	5	2	9	0,72
-1.3.	758	-	-	3	1,00
-1.4.	757	-	3	1	0,40
-1.5.	724	6	3	28	0,86
-1.6.1.	737	3	5	16	0,79
-1.6.2.	743	6	1	11	0,75
-1.6.3.	739	7	-	15	0,81
<b>2: Dozent als Unterrichtender und als Person</b>					
-2.1.1.	750	2	2	7	0,78
-2.1.2.	751	3	1	6	0,75
-2.1.3.	712	10	2	37	0,85
-2.1.4.	748	2	5	6	0,63
-2.2.1.	741	2	1	17	0,92
-2.2.2.	737	5	4	15	0,76
<b>3: Didaktik</b>					
-3.1.	671	14	4	72	0,88
-3.2.	707	10	8	36	0,79
-3.3.1.	758	1	2	-	-
-3.3.2.	761	-	-	-	1,00
-3.4.	721	8	9	23	0,72
-3.5.1.	737	5	5	14	0,73
-3.5.2.	715	6	8	32	0,81
-3.5.3.	760	-	1	-	-
-3.6.	729	4	8	20	0,76
-3.7.	711	14	7	29	0,73
-3.8.	733	3	3	22	0,88
-3.9.	753	1	-	7	0,93
<b>4: Themen und Inhalte</b>	661	26	8	66	0,77

<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>					
-5.1.	750	3	-	8	0,84
-5.2.1.	755	-	1	5	0,91
-5.2.2.	758	1	-	2	0,80
-5.3.1.	745	1	1	14	0,93
-5.3.2.	747	1	1	12	0,92
-5.4.	752	1	1	7	0,87

Legende: Vierfeldertafel jeweils für positive und negative Subkategorien: 0/0 gibt die Zellengröße für die Fälle an, wo beide Rater die Kategorie für nicht gegeben hielten; 0/1 und 1/0, dass nur einer der beiden Rater die Kategorie zuordnete, und 1/1, dass beide Rater bezüglich der Kategorienzuweisung übereinstimmten; Kappa = Inter-Rater-Reliabilitätsindex für die jeweilige Subkategorie.



## Anhang E

Frequenzanalyse der gesamten Stichprobe; Subkategorien

---

**Tabelle 13: Frequenzanalyse; Subkategorien**

Positive Subkategorien	Kodierungen	
	absolut	in %
<b>1: Organisation des Unterrichts</b>		
+1.1.	-	-
+1.2.	21	<b>1,3</b>
+1.3.	3	0,2
+1.4.	1	0,1
+1.5.	3	0,2
+1.6.	2	0,1
<b>2: Dozent als Unterrichtender und als Person</b>		
+2.1.1.	39	2,4
+2.1.2.	28	1,7
+2.1.3.	318	<b>19,6</b>
+2.2.1.	101	6,2
+2.2.2.	30	1,9
+2.2.3.	24	1,5
+2.2.4.	19	1,2
<b>3: Didaktik</b>		
+3.1.	115	<b>7,1</b>
+3.2.	108	6,7
+3.3.	37	2,3
+3.4.	23	1,4
+3.5.	5	0,3
+3.6.	13	0,8
+3.7.	15	0,9
+3.8.	6	0,4
<b>4: Themen und Inhalte</b>	<b>638</b>	<b>39,4</b>
<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>		
+5.1.	14	0,8
+5.2.	11	0,7

+5.3.	46	<b>2,8</b>
<b>Gesamt</b>	<b>1620</b>	<b>100</b>

<b>Negative Subkategorien</b>		
<b>1: Organisation des Unterrichts</b>		
-1.1.	23	1,1
-1.2.1.	69	3,2
-1.2.2.	39	1,8
-1.2.3.	47	2,2
-1.3.	31	1,5
-1.4.	30	1,4
-1.5.	139	<b>6,5</b>
-1.6.1.	96	4,5
-1.6.2.	74	3,5
-1.6.3.	76	3,6
<b>2: Dozent als Unterrichtender und als Person</b>		
-2.1.1.	46	2,2
-2.1.2.	25	1,2
-2.1.3.	111	<b>5,2</b>
-2.1.4.	40	1,9
-2.2.1.	72	3,4
-2.2.2.	69	3,2
<b>3: Didaktik</b>		
-3.1.	221	<b>10,4</b>
-3.2.	92	4,3
-3.3.1.	18	0,8
-3.3.2.	7	0,3
-3.4.	68	3,2
-3.5.1.	57	2,7
-3.5.2.	47	2,2
-3.5.3.	4	0,2
-3.6.	52	2,4
-3.7.	120	<b>5,6</b>
-3.8.	60	2,8
-3.9.	60	2,8
<b>4: Themen und Inhalte</b>	<b>225</b>	<b>10,5</b>

<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>		
-5.1.	20	0,9
-5.2.1.	13	0,6
-5.2.2.	9	0,4
-5.3.1.	33	1,5
-5.3.2.	27	1,3
-5.4.	14	0,7
<b>Gesamt</b>	<b>2134</b>	<b>100</b>

Legende: Kodierungen absolut = absolute Häufigkeit, Kodierungen in % = relative Häufigkeit der Zuordnung einer Subkategorie; Gesamt = Gesamthäufigkeit der Kodierungen für positive bzw. negative Meinungsäußerungen.

## Anhang F

Frequenzanalyse der Gruppen; Hauptkategorien

**Tabelle 14: Positive Hauptkategorien - Häufigkeitsanalyse Vorklinik - Klinik**

Positive Hauptkategorie	Kodierungen Vorklinik		Kodierungen Klinik	
	absolut	in %	absolut	in %
<b>+1: Organisation des Unterrichts</b>				
	7	<b>0,9</b>	23	<b>2,9</b>
<b>+2: Dozent als Unterrichtender und als Person</b>				
	220	27,0	339	<b>42,0</b>
<b>+3: Didaktik</b>	164	20,1	158	19,6
<b>+4: Themen und Inhalte</b>	391	<b>48,0</b>	247	30,6
<b>+5: Medieneinsatz, Bild- und Arbeitsmaterialien</b>				
	32	3,9	39	4,8
<b>Gesamt</b>	<b>814</b>	<b>100</b>	<b>806</b>	<b>100</b>

**Tabelle 15: Negative Hauptkategorien - Häufigkeitsanalyse Vorklinik - Klinik**

Negative Hauptkategorie	Kodierungen Vorklinik		Kodierungen Klinik	
	absolut	in %	absolut	in %
<b>-1: Organisation des Unterrichts</b>				
	269	22,3	355	<b>38,3</b>
<b>-2: Dozent als Unterrichtender und als Person</b>				
	188	15,6	175	18,9
<b>-3: Didaktik</b>	548	<b>45,4</b>	258	27,9
<b>-4: Themen und Inhalte</b>	142	11,7	83	9,0
<b>-5: Medieneinsatz, Bild- und Arbeitsmaterialien</b>				
	61	<b>5,0</b>	55	<b>5,9</b>
<b>Gesamt</b>	<b>1208</b>	<b>100</b>	<b>926</b>	<b>100</b>

Legende: Kodierungen absolut = absolute Häufigkeit der Zuordnung einer Hauptkategorie; Kodierungen in % = relative Häufigkeit der Zuordnung einer Hauptkategorie bezogen auf die Gesamtheit aller positiven bzw. negativen Nennungen; Gesamt = Gesamthäufigkeit der Kodierungen für negative bzw. positive Meinungsäußerungen.

## Anhang G

Frequenzanalyse der Gruppen; Subkategorien

**Tabelle 16: Frequenzanalyse der Gruppen; positive Subkategorien**

Positive Subkategorien	Kodierungen Vorklinik		Kodierungen Klinik	
	absolut	in %	absolut	in %
<b>1: Organisation des Unterrichts</b>				
+1.1.	-	-	-	-
+1.2.	5	<b>0,6</b>	16	<b>2,0</b>
+1.3.	-	-	3	0,4
+1.4.	1	0,1	-	-
+1.5.	-	-	3	0,4
+1.6.	1	0,1	1	0,1
<b>Summe</b>	<b>7</b>	<b>0,9</b>	<b>23</b>	<b>2,9</b>
<b>2: Dozent als Unterrichtender und als Person</b>				
+2.1.1.	15	1,8	24	3,0
+2.1.2.	14	1,7	14	1,7
+2.1.3.	150	<b>18,4</b>	168	<b>20,8</b>
+2.2.1.	23	2,8	78	9,6
+2.2.2.	8	1,0	22	2,7
+2.2.3.	6	0,7	18	2,2
+2.2.4.	4	0,5	15	1,9
<b>Summe</b>	<b>220</b>	<b>27,0</b>	<b>339</b>	<b>42,0</b>
<b>3: Didaktik</b>				
+3.1.	67	<b>8,2</b>	48	<b>6,0</b>
+3.2.	60	7,4	48	<b>6,0</b>
+3.3.	2	0,2	35	4,3
+3.4.	11	1,4	12	1,5
+3.5.	2	0,2	3	0,4
+3.6.	7	0,9	6	0,7
+3.7.	9	1,1	6	0,7
+3.8.	6	0,7	-	-
<b>Summe</b>	<b>164</b>	<b>20,1</b>	<b>158</b>	<b>19,6</b>
<b>4: Themen und Inhalte</b>	<b>391</b>	<b>48,0</b>	<b>247</b>	<b>30,6</b>

<b>5: Medieneinsatz, Bild –und Arbeitsmaterialien</b>				
+5.1.	10	<b>1,2</b>	4	0,5
+5.2.	9	1,1	2	0,2
+5.3.	13	1,6	33	<b>4,1</b>
<b>Summe</b>	<b>32</b>	<b>3,9</b>	<b>39</b>	<b>4,8</b>
<b>Gesamt</b>	<b>814</b>	<b>100</b>	<b>806</b>	<b>100</b>

Legende: Kodierungen absolut = absolute Häufigkeit der Zuordnung einer Subkategorie; Kodierungen in % = relative Häufigkeit der Zuordnung einer Subkategorie bezogen auf die Gesamtheit aller positiven Nennungen; Summe = Häufigkeit der Zuordnung für die Hauptkategorie; Gesamt = Gesamthäufigkeit der Kodierungen für positive Meinungsäußerungen.

**Tabelle 17: Frequenzanalyse der Gruppen; negative Subkategorien**

Negative Subkategorien	Kodierungen Vorklinik		Kodierungen Klinik	
	absolut	in %	absolut	in %
<b>1: Organisation des Unterrichts</b>				
-1.1.	8	0,7	15	1,6
-1.2.1.	16	1,3	53	5,7
-1.2.2.	14	1,2	25	2,7
-1.2.3.	10	0,8	37	4,0
-1.3.	9	0,7	22	2,4
-1.4.	7	0,6	23	2,5
-1.5.	55	4,5	84	<b>9,1</b>
-1.6.1.	82	<b>6,8</b>	14	1,5
-1.6.2.	21	1,7	53	5,7
-1.6.3.	47	3,9	29	3,1
<b>Summe</b>	<b>269</b>	<b>22,3</b>	<b>355</b>	<b>38,3</b>
<b>2: Dozent als Unterrichtender und als Person</b>				
-2.1.1.	28	2,3	18	1,9
-2.1.2.	14	1,2	11	1,2
-2.1.3.	71	<b>5,8</b>	40	4,3
-2.1.4.	14	1,2	26	2,8
-2.2.1.	27	2,2	45	<b>4,9</b>
-2.2.2.	34	2,8	35	3,8
<b>Summe</b>	<b>188</b>	<b>15,6</b>	<b>175</b>	<b>18,9</b>
<b>3: Didaktik</b>				

-3.1.	126	<b>10,4</b>	95	<b>10,3</b>
-3.2.	59	4,9	33	3,6
-3.3.1.	3	0,3	15	1,6
-3.3.2.	-	-	7	0,8
-3.4.	54	4,5	14	1,5
-3.5.1.	41	3,4	16	1,7
-3.5.2.	44	3,6	3	0,3
-3.5.3.	2	0,2	2	0,2
-3.6.	37	3,1	15	1,6
-3.7.	83	6,9	37	4,0
-3.8.	44	3,6	16	1,7
-3.9.	55	4,5	5	0,5
<b>Summe</b>	<b>548</b>	<b>45,4</b>	<b>258</b>	<b>27,9</b>
<b>4: Themen und Inhalte</b>	<b>142</b>	<b>11,7</b>	<b>83</b>	<b>9,0</b>
<b>5: Medieneinsatz, Bild – und Arbeitsmaterialien</b>				
-5.1.	12	1,0	8	0,9
-5.2.1.	5	0,4	8	0,9
-5.2.2.	1	0,1	8	0,9
-5.3.1.	17	<b>1,4</b>	16	<b>1,7</b>
-5.3.2.	17	<b>1,4</b>	10	1,0
-5.4.	9	0,7	5	0,5
<b>Summe</b>	<b>61</b>	<b>5,0</b>	<b>55</b>	<b>5,9</b>
<b>Gesamt</b>	<b>1208</b>	<b>100</b>	<b>926</b>	<b>100</b>

Legende: Kodierungen absolut = absolute Häufigkeit der Zuordnung einer Subkategorie; Kodierungen in % = relative Häufigkeit der Zuordnung einer Subkategorie bezogen auf die Gesamtheit aller negativen Nennungen; Summe = Häufigkeit der Zuordnung für die Hauptkategorie; Gesamt = Gesamthäufigkeit der Kodierungen für negative Meinungsäußerungen.

## **Anhang H**

### **Verzeichnis der akademischen Lehrer**

Meine akademischen Lehrer an der Philipps-Universität waren die Herren und Damen Professoren und Professorinnen:

Arnold, Aumüller, Bach, Basler, Baum, Barth, Baum, Becker, Bertalanffy, Christ, Christiansen, Czubayko, Daut, Eilers, Feuser, Geus, Görg, Gotzen, Grimm, Griss, Gudermann, Happle, Hellinger, Hesse, Hinrichs, Hofmann, Jungclas, Kern, Klaus, Klenk, Klose, Koolman, Köhler, Kretschmer, Krieg, Kroll, Kuhlmann, Lang, Lammel, Maier, Maisch, Moll, Moosdorf, Mutters, Neubauer, Oertel, Renschmidt, Renz, Rothmund, Schäfer, Schmidt, Schnabel, Schneider, Seitz, Seyberth, Suske, Vogelmeier, Wagner, Weihe, Werner, Westermann, Wilke, Wulf, Zielke.



**Danksagung**

Mein Dank gilt an erster Stelle meinem wissenschaftlichen Betreuer Herrn Prof. Dr. Dr. Basler für die Überlassung des Themas sowie für die Unterstützung während Erstellung der Arbeit sowie deren kritische Durchsicht.

Besonders danken möchte ich Herrn Dipl.-Kfm. Thiemo Kohlsdorf, Universität Mainz, für die Zweitcodierung der Fragebögen sowie die wissenschaftliche Beratung bei Erstellung des wirtschaftswissenschaftlichen Teils der Arbeit.

Abschließend möchte ich meinen Eltern für die Unterstützung während der Zeit meines Medizinstudiums danken.