

Struktur und Prozess psychiatrischer Hochschullehre

Ein Anwendungsgebiet für Qualitätssicherungs- und Evaluationsmethoden

Structure and process of university teaching in psychiatry

A field for methods of quality assurance and evaluation

• Claus Barkmann¹ • Katja Weidtmann² • Michael Schulte-Markwort¹

Zusammenfassung:

Zielsetzung: Die besondere Arbeitsbelastung an einer psychiatrischen Universitätsklinik sowie das bislang geltende Wertesystem mit der Betonung von Klinik und Wissenschaft führen zu einer systematischen Vernachlässigung der Lehre.

Methodik: Mit Hilfe evaluativer Methoden wurden Vorlesungen und Seminare eines Semesters an der Klinik für Kinder- und Jugendpsychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf aus Sicht von Studierenden und Dozenten getrennt für die Inhaltsbereiche Form, Inhalt, Dozenten und Gesamtbewertung per Fragebogen beurteilt.

Ergebnisse: Trotz vorhandener organisatorischer Mängel bewerteten die Studierenden die Veranstaltungen hinsichtlich aller vier Inhaltsbereiche im Mittel als gut. Ein bivariates Vorhersagemodell erklärte 46% der Varianz der Gesamtbeurteilung. Überraschend ist die hohe Übereinstimmung von Studierenden- und Dozentenurteilen.

Schlussfolgerung: Die fortlaufende und systematische Evaluation von Lehrveranstaltungen dient der Qualitätssicherung und -verbesserung gegenwärtiger und zukünftiger Lehrmethoden.

Schlüsselwörter: Lehrveranstaltungen, Hochschullehre, Evaluation, Qualitätssicherung, Kinder- und Jugendpsychiatrie

Abstract:

Objective: Given the exceptional workload at a university psychiatric hospital and the current emphasis on clinical medicine and science, teaching is systematically being neglected.

Methods: With the help of evaluation methods involving the completion of a questionnaire, lectures and seminars held during one semester at the Department for Child and Adolescent Psychiatry and Psychotherapy of the University Hospital Hamburg-Eppendorf were assessed separately by students and lecturers in terms of form, content, lecturers, and overall assessment.

Results: Despite organizational shortcomings, the lectures and seminars were rated on average as good in all four assessment areas. Using a bivariate prediction model, it was possible to explain 46% of the variance in overall assessment. A surprisingly high concordance was found between the assessments by students and lecturers.

Conclusion: Continuous and systematic evaluation of lectures and seminars ensures and improves the quality of current and future teaching methods.

Keywords: lectures, university teaching, evaluation, quality assurance, child and adolescent psychiatry

Einleitung

• 1. Problemstellung

Universitäre Psychiatrie besteht, glaubt man ihren Publikationsorganen, ausschließlich aus klinischen und wissenschaftlichen Arbeitsfeldern. Die Hochschullehre ist nicht im Blick der Profession. Dies hat dieses Fachgebiet gemein mit allen anderen medizinischen Disziplinen und vielen anderen Fachbereichen. Zu den Aufgaben eines Universitätsklinikums zählt neben der klinischen Versorgung und der Forschung aber auch die Lehre. Die besondere Arbeitsbelastung durch den alltäglichen klinischen Versorgungsdruck sowie das bislang geltende Wertesystem mit der Betonung von Klinik und Wissenschaft führen allerdings zu einer Vernachlässigung derselben. Dabei ist die Lehre von besonderer, auf keinen Fall geringerer Bedeutung für das Fach. Schließlich haben Studierende Anspruch auf eine gute Ausbildung, Patienten haben Anspruch

auf gut ausgebildetes Personal, die Gesellschaft hat Anspruch auf die sinnvolle Verwendung von Steuergeldern und das Fach will sich im Reigen konkurrierender Disziplinen weiter etablieren und den eigenen Nachwuchs sichern.

• 2. Evaluationskonzepte

Universitäre Lehre ist auf drei verschiedenen Ebenen evaluierbar; auf der Ebene der gesamten Hochschule als Institution, auf Fachbereichsebene sowie auf der Stufe einzelner Lehrveranstaltungen und Lehrenden [1]. Nach dem Kriterium der Zielorientierung lassen sich fünf verschiedene Modelle der Lehrevaluation unterscheiden [2]: das Qualifikationsmodell (Optimierung der Lehrqualität des Dozenten und der Veranstaltung), das Transparenzmodell (Aufzeigen von Stärken und Schwächen), das Kommunikationsmodell (Anregung der Diskussion), das Steuerungsmodell (Instrument bei Entscheidungen über Mittelvergabe, Ausstattung etc.) sowie das Forschungsmodell (Untersuchung von Lehr- und Lernprozessen).

¹ Zentrum für Frauen-, Kinder- und Jugendmedizin, Klinik und Poliklinik für Kinder- und Jugendpsychosomatik, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Deutschland

² Studiendekanat, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Deutschland

sen). Die für eine Designplanung wichtigste Unterscheidung ist die von formativer und summativer Evaluation bzw. von Prozess- und Ergebnisaspekten, denn sie legt Ziele, Zeitpunkte und Ausgestaltung der Evaluation fest: Während formativ fortlaufende rekursive Veränderung und Überprüfung der Veränderung bedeutet, wird summativ eine bereits abgeschlossene Veranstaltung bewertet [3]. In der Literatur sind vor allem studentische Gruppen, einige Printmedien und manchmal auch Hochschulinstitutionen selbst als Evaluatoren zu finden [4]. Träger systematischer studentischer Lehrevaluationen sind meist Asten, studentische Evaluationszentren oder Fachschaften [5]. Die angewandten Methoden entstammen den klassischen Evaluationsmethoden, zeigen aber charakteristische Häufungen bei mündlichem Feedback (Gruppendiskussionen, Interviews, Blitzlichter und Abschlussbefragungen) und schriftlichem Feedback (Fragebögen mit offenen oder geschlossenen Items).

• 3. Empirische Ergebnisse

Der aktuelle Forschungsschwerpunkt zur Qualität der Hochschullehre im deutschen Sprachraum liegt auf den Vor- und Nachteilen der Veranstaltungskritik durch Studierende [6]. Rindermann [7] belegt anhand korrelativer und varianzanalytischer Vergleiche studentischer Einschätzungen mit Urteilen von Lehrenden und Fremdbeurteilern die Validität der Studierendenurteile. Schweer und Rosemann [8] zeigen, dass Studenten in ihren Urteilen keinen einheitlichen Maßstab zugrunde legen, sondern unterschiedliche Aspekte der Lehre für deren Qualität als ausschlaggebend anzusehen scheinen. Verschiedene Autoren haben in voneinander unabhängigen Studien keine Hinweise für Verzerrungen von Urteilen durch Geschlecht, Alter oder universitäre Vorbildungen gefunden [9], [5], [8].

Zur Zeit werden an 84% der medizinischen Fachbereiche in Deutschland Evaluationen durchgeführt, die sich allerdings erheblich in den Zielen, Methoden und im Umfang unterscheiden [10]. Bislang sind sieben Arbeiten medizinischer Fakultäten mit Fokussierung auf psychosoziale Fachabteilungen publiziert worden [11], [12], [13], [14], [15], [16], [17], [18]. Deren Ergebnisse zeigen, dass die Beurteilungen der Studierenden durchweg besser ausfallen als erwartet und eine curriculare Stärkung der psychosozialen Fähigkeiten befürwortet wird.

• 4. Fragestellung der Untersuchung

Hauptfragestellung war die Analyse von Struktur und Qualität der Lehrveranstaltungen an der Klinik für Kinder- und Jugendpsychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf. Einen besonderen Schwerpunkt bildete die Identifikation von Prädiktoren einer positiven Gesamtbewertung der Lehrveranstaltungen. Darüber hinaus wurde nach systematischen Urteilsunterschieden zwischen Studierenden und Dozenten gefragt. Aus den Ergebnissen sollten anschließend Maßnahmen zur Qualitätssicherung und -verbesserung abgeleitet werden.

Methode

• 1. Design

Gegenstand der Evaluation waren die Lehrveranstaltungen der Klinik für Kinder- und Jugendpsychiatrie und Psychotherapie des Universitätsklinikums Hamburg-Eppendorf. Unter Verwendung formativer Evaluationskonzepte wurde eine empirische Quer-

schnittsbefragung an Hand eines selbstentwickelten Fragebogens durchgeführt. Als Urteiler wurden die direkt an der Lehre Beteiligten, also Studierende und Lehrende gewählt. Auf eine vollständige Spiegevaluation (Studierende und Dozenten bearbeiten dieselben Items) wurde verzichtet, da einige der Fragen nicht gleichermaßen von beiden Urteilergruppen beantwortbar sind. Schließlich wurden die Ergebnisse in konkrete Maßnahmen zur Verbesserung der Lehre umgesetzt und ein Konzept zur fortlaufenden Evaluierung der Lehre erarbeitet (Metaevaluation).

• 2. Operationalisierung

Die angebotenen Lehrveranstaltungen ließen sich sowohl in organisatorischer als auch in inhaltlich-konzeptioneller Hinsicht in zwei Gruppen gliedern, dem Psychiatrischen Praktikum auf der einen sowie Vorlesungen und Seminare auf der anderen Seite: Das Psychiatrische Praktikum ist als Pflichtkurs ausschließlich für Studierende der Medizin vorgesehen, die insgesamt jeweils zwei Sitzungen von vier Unterrichtsstunden mit schwerpunktmäßiger Fallbearbeitung absolvieren ("bedside teaching"). Dagegen finden alle Vorlesungen und Seminare auf freiwilliger Basis in wöchentlichen Sitzungen statt und sind Studierenden aller Fachbereiche zugänglich. Dieser Unterscheidung wurde sowohl bei der Konzeption der Fragebögen und Durchführung der Datenerhebung als auch bei der statistischen Auswertung Rechnung getragen. Die Auswahl der zu erfassenden Variablen erfolgte gemäß dem Münchener multifaktoriellen Modell der Lehrveranstaltungsqualität von Rindermann [19], das neben formalen auch inhaltliche Merkmale der Veranstaltungen erfasst. Die inhaltliche Ausrichtung ist in einer alle Veranstaltungen einer Klinik umfassenden Evaluation allerdings dadurch begrenzt, dass unterschiedliche Veranstaltungsinhalte vergleichbar erfasst werden müssen. Für alle Veranstaltungen wurden außerdem die Regelmäßigkeit der Termine und die fortlaufende Entwicklung der Teilnehmerzahlen überprüft. Entsprechend diesen Ausgangsbedingungen wurden insgesamt vier unterschiedliche, siebenseitige Fragebogenversionen erstellt (2 Urteilergruppen x 2 Veranstaltungsarten). Die Items wurden als Aussagesätze formuliert (Beispiel: "Das räumliche Umfeld ist angenehm", "Der Inhalt ist praxisbezogen"). Zur Bewertung wurde eine sechsstufige Ratingskala mit zusätzlich verbaler Verankerung ("gar nicht" bis "absolut zutreffend") gewählt. Als jeweils zusammenfassende Bewertung einer Qualitätskomponente wurde ein Item mit Schulnotenbewertungssystem eingesetzt.

• 3. Stichprobe

Es wurden n=146 Teilnehmerinnen des Psychiatrischen Praktikums, n=58 Teilnehmerinnen der Vorlesungen und Seminare (zusammen n=204 Studierende) sowie n=24 Dozenten des Sommersemesters 2001 befragt. Die Stichprobenbeschreibung wird im Ergebnisteil den jeweiligen Daten vorangestellt. Die Größe der Urteilerstichprobe wurde aufgrund des explorativen Vorgehens nicht a-priori festgelegt, sondern ergab sich durch die Klumpenauswahl. Dieses Semester ist repräsentativ für die Semester der letzten drei bis vier Jahre, da sich in dieser Zeit weder beim Lehrplan noch bei den Dozenten Veränderungen ergeben haben.

Erhebungszeitpunkt für Vorlesungen und Seminare war die sechste Semesterwoche. Dadurch sollte einerseits die Gefahr einer Datenverzerrung in Richtung auf zu positive Ergebnisse durch systematische Drop-Outs von unzufriedenen Veranstaltungsabnehmern minimiert, andererseits eine ausreichende Urteilsbasis zur

Veranstaltungsbewertung für die Studierenden ermöglicht werden. Die Versuchsleiterin suchte diese Veranstaltungen persönlich auf, verteilte die Fragebögen am Ende der Sitzung und sammelte sie dort auch wieder ein. Kompaktseminare wurden ebenfalls von der Versuchsleiterin persönlich am Ende der Veranstaltung aufgesucht. Von insgesamt 67 Studierenden konnten 58 vollständig ausgefüllte Fragebögen erfasst werden. Kein Student besuchte mehr als eine Veranstaltung der Klinik.

• 4. Analysen

Die Daten wurden mit Hilfe von SPSS 10.0 für Windows analysiert. Zur Beantwortung der quantitativen Fragestellungen wurden Tabellen, Grafiken und statistische Kennwerte verwendet. Die Darstellung der Einzelitems mit Hilfe von Boxplots dient dem Vergleich zwischen den einzelnen Merkmalen sowohl hinsichtlich der zentralen Tendenz als auch dem Ausmaß der Unterschiedlichkeit der Beurteilungen. Zur Beantwortung der Prädiktionsfragestellung wurde neben bivariaten Pearson-Korrelationen eine multiple lineare Regressionsanalyse durchgeführt (schrittweise Methode, Einschlusskriterium $p(F) \leq 5\%$). Die Urteilerübereinstimmung wurde mit Hilfe der Intraklassenkorrelationskoeffizienten und standardisierten Mittelwertsabständen (d , Cohen, 1988) bestimmt. Das Alphafehlerniveau wurde auf 5% festgesetzt, die Testungen erfolgten zweiseitig.

Ergebnisse

Wegen der Fülle der Daten beschränkt sich nachfolgende Darstellung auf die Vorlesungen und Seminare (Ergebnisse zum Psychiatrischen Praktikum siehe Barkmann et al [20]). Zu Beginn wird die Beurteilung aus Sicht der Studierenden präsentiert, dann die Studierenden- und Dozentenurteile einander gegenübergestellt. Abschließend wird die Umsetzung der Ergebnisse skizziert. Die Bewertungsachsen aller Abbildungen werden wegen der besseren Übersichtlichkeit einheitlich als Schulnotenskala dargestellt.

• 1. Lehrveranstaltungen an der Klinik

Tabelle 1 zeigt die in dieser Studie untersuchten Lehrveranstaltungen, die veranstaltungsspezifischen Teilnehmerzahlen im Laufe des Semesters sowie Angaben zur Regelmäßigkeit der Sitzungen. Mit Ausnahme des Seminars "Filme zur Kinder- und Jugendpsychiatrie" sind keine substantiellen Veränderungen der Teilnehmerzahlen über die Dauer des Semesters zu beobachten. Auffällig ist, dass von insgesamt neun im Vorlesungsverzeichnis angekündigten Veranstaltungen immerhin drei aufgrund mangelnder Teilnehmerzahlen komplett ausfielen. Die Prüfung der Regelmäßigkeit des Psychiatrischen Praktikums offenbart bei 17 von 20 Kursgruppen unvorhergesehene organisatorische Veränderungen (meist Dozententausch oder Einzelleitung statt der vorgesehenen Doppelleitung).

Tabelle 1: Lehrveranstaltungen, Teilnehmerzahlen und Regelmäßigkeiten

Nr.	Titel der Veranstaltung	Typ	Anzahl der Teilnehmerinnen (n)				
			SW 1	SW 3	SW 6	SW 9	SW 12
1	Ausgewählte Kapitel der KJP	V	27	32	15	20	15
2	Einführung in die KJP II	V	37	31	25	31	20
3	Einführung in SPSS (kompakt)	S			---		
4	Familientherapie (zweiwöchig)	S	---	---	---	---	---
5	Filme zur KJP	S	11	8	6	5	2
6	Forschungsmethoden der KJP	S	8	6	10	---	7
7	Psychotraumatologie (kompakt)	S	11 (*6.SW); 10 (*10.SW); 8 (*15.SW)				
8	Schizophrenie (kompakt)	S			---		
9	Psychiatrisches Praktikum	P	Zählungen jeweils zur 2.Sitzung				

V/S/P=Vorlesung/Seminar/Praktikum; SW=Semesterwoche; --- = Ausfall;

* =Datum der Zählung

• 2. Teilnehmerstichprobe

Die Teilnehmerstichprobe von $n=58$ Studierenden der Vorlesungen und Seminare besteht zu 86% aus Frauen, das Alter ist linkssteil verteilt mit $M=27$ Jahren ($SD=6.0$), 93% sind deutscher Nationalität. Der größte Teil der Studierenden stammt aus den Fachbereichen Medizin und Sonderpädagogik (jeweils 16%) sowie der Psychologie (12%). In geringer Zahl nehmen Studierende sonstiger Fachrichtungen (wie z. B. Theologie oder Sportwissenschaften) sowie Gasthörer teil. Die Anzahl der Fachsemester schwankt deutlich ($M = 13$; $SD = 23.80$). Frauen urteilen positiver als Männer ($d=.15$), Medizinstudenten urteilen besser als Studierende anderer Fachrichtungen ($d=.58$).

• 3. Bewertung aus Sicht der Studierenden

Abbildung 1 zeigt, dass die Befragten mit den verschiedenen Formmerkmalen insgesamt zufrieden sind. Der Informationsfluss und das räumliche Umfeld wurden relativ am schlechtesten bewertet. Bei der Differenzierung einer eventuellen Unzufriedenheit mit der Gruppengröße beurteilten 12% der Befragten sie als zu groß (v. a. Teilnehmer der beiden Vorlesungen) und 9% als zu klein. Nur in zwei Seminaren wurden Unterrichtsmaterialien ausgehändigt. In Abhängigkeit von der jeweiligen Veranstaltung wünschten sich die Studierenden mehr Materialien mit mehr Beispielen, die zu Beginn der Veranstaltung ausgeteilt werden sollten. Die Form der Vorlesungen und Seminare wurde insgesamt mit der Schulnote $M=2.2$ ($SD=0.56$) bewertet.

Abbildung 1 zeigt auch die Bewertungen der verschiedenen inhaltlichen Merkmale. Während die Interessantheit des Inhaltes und die Praxisbezogenheit besonders gut bewertet werden, wird die Vertiefung des Stoffes relativ am schlechtesten beurteilt. Bei der Differenzierung einer möglichen Unzufriedenheit hinsichtlich des Tempos zeigt sich, dass jeweils 14% der Befragten es als zu schnell bzw. als zu langsam empfanden. Beim Umfang zeichnet sich ein deutlicheres Bild ab: 21% beurteilten ihn als zu klein, 7% als zu groß. Der Inhalt der Vorlesungen und Seminare wird in der zusammenfassenden Beurteilung mit der durchschnittlichen Schulnote von $M=2.0$ ($SD=0.66$) bewertet.

Bei der Beurteilung der Dozentenleistungen gingen die Urteile von $n=15$ Studierenden aus "Forschungsmethoden der KJP" und

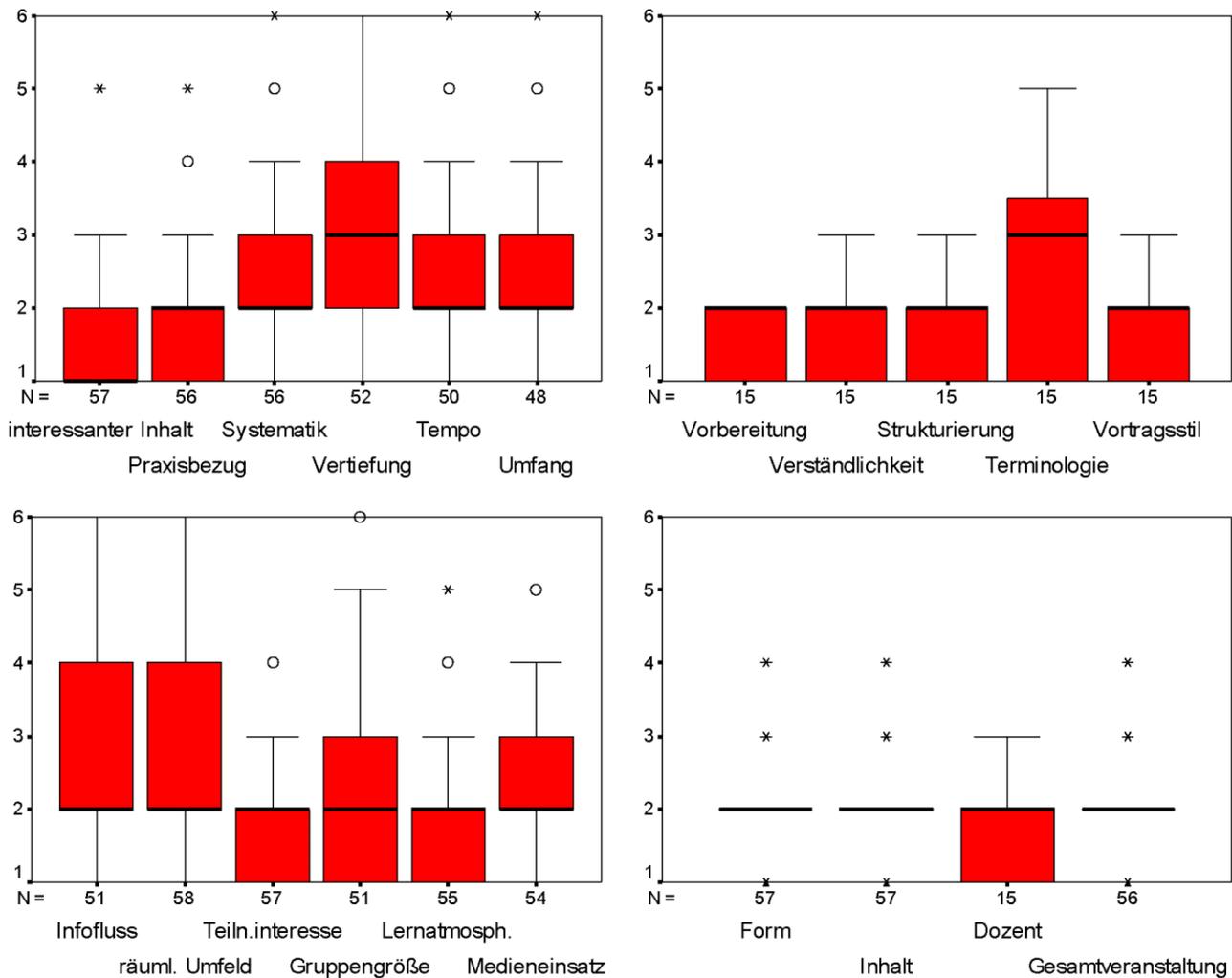


Abbildung 1: Gesamtbewertung von Form, Inhalt, Dozent sowie Globalbewertung im Urteil der Studierenden (n=58, Schulnotenskala)

aus "Ausgewählten Kapitel der KJP" in die Berechnungen ein. Die übrigen Veranstaltungen wurden nicht berücksichtigt, da in diesen jeweils mehrere verschiedene Dozenten zu bewerten gewesen wären. Die fachliche Kompetenz der Lehrenden wird hinsichtlich der Merkmale Fachwissen, logische Konsistenz, Eingehen auf Fragen und Sicherheit im Patientengespräch mit sehr gut bis gut bewertet (Durchschnittsnote M=1.6). Das gleiche gilt hinsichtlich der Merkmale sozialer Kompetenz (Pünktlichkeit, Freundlichkeit, Motivierung der Teilnehmer, Offenheit für Fragen sowie Interesse am Lernerfolg der Teilnehmer; Durchschnittsnote M=1.3). Bei der didaktischen Kompetenz der Lehrenden wird vor allem die Einführung in die Terminologie deutlich kritischer beurteilt (vgl. Abbildung 1). Die Gesamtbewertung der Teilnehmer der Vorlesungen und Seminare fällt für den Dozentenblock mit M=1.6 und SD=.06 besser aus als die für die Form- und Inhaltsblöcke.

Die Globalbewertung der Vorlesungen und Seminare als Schulnote fällt mit M=2.0 (SD=.07) gut aus und entspricht damit den Einzelbenotungen für Form und Inhalt.

• 4. Wovon hängt die Gesamtbeurteilung ab?

Zur Identifikation derjenigen Variablen, die den höchsten Erklärungswert für die Gesamtbewertung der Vorlesungen und Seminare

bieten, wurde eine multiple lineare Regressionsanalyse durchgeführt. In die Berechnung gingen alle Variablen ein, die spezifische Teilaspekte der Lehre erfassen und von allen Teilnehmern bewertet wurden. Tabelle 2 liefert eine Übersicht der höchsten Korrelationen zwischen diesen spezifischen Variablen und der Gesamtbewertung der Vorlesungen und Seminare. Die folgende Gleichung zeigt das errechnete Modell, das den Anteil einzelner Variablen an der Gesamtnote erklärt: $GES = 2.78 + .26 VOR + .28 INT$. Es berücksichtigt die Variablen „Abstimmung auf Voraussetzungen der Teilnehmer" sowie „interessanter Inhalt" und klärt mit 46% maximale Varianz auf (standardisierte Partialregressionskoeffizienten: $\beta_{Vor}=.47, \beta_{Int}=.34; R=.67; R^2=.46$). Alle übrigen in Tabelle 2 dargestellten Variablen wurden ausgeschlossen, da sie keinen darüber hinausgehenden Erklärungswert besitzen. Suppressorvariablen konnten nicht gefunden werden.

Tabelle 2: Maximale Korrelationen spezifischer Bewertungsaspekte mit Gesamtnote

Bewertungsaspekte	Gesamtnote
Umfang	.58**
Abstimmung auf die Teilnehmer	.56**
Vertiefung	.52**
Tempo	.50**
Interessantheit	.48**
Praxisbezogenheit	.45**
Einblick in die Thematik	.43**
Systematik	.38**
Qualität der Lernatmosphäre	.35**
räumliches Umfeld	.29*

n=58; Pearson-Korrelation, *:p≤.05, **:p≤.01

• 5. Vergleich der Studierenden- und Dozentenurteile

Statt der itemweisen Ergebnisdarstellung der Dozentenbeurteilung wird hier die Übereinstimmung von Dozenten- und Studierendenurteil in den von beiden Urteilergruppen gleichermaßen bearbeiteten Items über alle Veranstaltungen präsentiert (Tabelle 3). Die Urteilerstichprobe der Dozenten besteht aus n=7 wissenschaftlichen Mitarbeiterinnen der Hamburger Kinder- und Jugendpsychiatrie, davon zwei Professoren, zwei Fachärzte, zwei Assistenzärzte und ein Psychologe. Für die Vorlesungen und Seminare ergibt der Vergleich der durchschnittlichen Urteile eine 83%ige Übereinstimmung der absoluten Werte (Intraklassenkorrelationskoeffizient $r_{icc}=.91$). Obwohl über alle Merkmale hinweg eine gute Urteilerkonkordanz besteht, gibt es hinsichtlich bestimmter Variablen größere Abweichungen: Die Dozenten bewerten insbesondere den Informationsfluss und das Ausmaß, in dem die Veranstaltung zur eigenen Beschäftigung mit der Thematik motiviert, besser als die Studierenden. Letztere bewerten das Interesse der Teilnehmer höher als Dozenten. Dieses Ergebnis wird durch den Vergleich der Globalbeurteilungen hinsichtlich Form, Inhalt und Gesamtbewertung bestätigt (Abbildung 2). Daneben beurteilten die Dozenten die Teilnehmer ihrer Sitzungen hinsichtlich Merkmalen wie Pünktlichkeit (M=2.3, SD=.49), Interesse an der Thematik (M=1.4, SD=.54), Freundlichkeit (M=1.4, SD=.54), Respekt (M=1.9, SD=.38) und einer angemessenen Distanz gegenüber dem Dozenten (M=2.0, SD=.58). Die Bewertungen vielen durchweg positiv aus.

Tabelle 3: Vergleich beider Urteilergruppen in den Spiegelitems über alle Veranstaltungen

Variable	Urteiler				d
	Studierende n=58		Dozenten n=7		
	M	SD	M	SD	
Informationsfluss	2.8	1.43	4.0	1.73	0.58
zur Eigenbeschäftigung motivierend	1.9	0.92	2.4	0.92	0.53
Lernatmosphäre	2.0	0.87	2.5	0.71	0.36
Umfang des Inhalts	2.6	1.13	3.0	1.29	0.34
Gruppengröße	2.0	1.25	2.4	0.98	0.32
Gliederung des Inhalts	2.7	1.22	3.0	0.93	0.27
Abstimmung auf Teilnehmer	2.4	1.19	2.7	0.96	0.20
Vertiefung des Inhalts	2.9	1.18	2.8	1.04	-0.12
Einblick in Thematik	2.2	0.77	2.0	0.82	-0.20
Tempo des Inhalts	2.5	1.28	2.1	0.69	-0.25
Interesse der Teilnehmer	1.8	0.74	1.4	0.54	-0.54

Schulnotenskala; Intraklassenkorrelation $r_{icc}=.91$; d=standardisierte Mittelwertsdifferenz nach Cohen (1988).

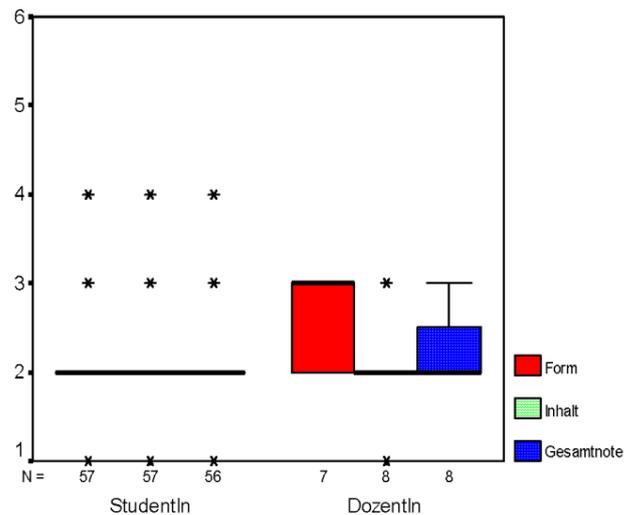


Abbildung 2: Vergleich der Studierenden- und Dozentenurteile hinsichtlich der Globalitems

• 6. Ergebnissumsetzung

Die Ergebnisse der Studie wurden in einem klinikinternen wissenschaftlichen Kolloquium allen Dozenten vorgestellt, diskutiert und Beschlüsse über das weitere Vorgehen gefasst. Außerdem wurde ein Lehrpreis für das beste Dozentenpaar im Psychiatrischen Praktikum überreicht. Diese zeremonielle Ehrung der beiden besten Dozenten sollte die Bedeutung der Lehrleistung betonen und als positiver Verstärker wirken. Zu den Beschlüssen zählen u. a. die Benennung eines Lehrkoordinators, die Einrichtung regelmäßiger Koordinationstreffen und eines Qualitätssicherungsordners für Dozenten zur Durchführung von Lehrveranstaltungen. Für jeden

Dozenten wurden Feedbackbögen zusammengestellt, in denen die Ergebnisse der Evaluation seiner Veranstaltungen und insbesondere seiner eigenen Lehrleistungen dargestellt und erläutert wurden. Darin sind auch Vergleiche zu Lehrleistungen anderer Dozenten sowie dozenten-spezifische Empfehlungen zur Optimierung des Lehrverhaltens enthalten.

Diskussion

• 1. Hauptergebnisse

- Die Strukturanalyse ergab unerwartete Unregelmäßigkeiten im Semesterablauf und hohe Ausfallquoten einzelner Veranstaltungen. Trotzdem bewerteten die Studierenden die Veranstaltungen in Form, Inhalt, hinsichtlich der Dozenten und in der Gesamtbeurteilung im Mittel mit "gut".
- Die beiden Prädiktoren "Abstimmung auf Voraussetzungen der Teilnehmer" und "interessanter Inhalt" erklärten gemeinsam 46% der Varianz der Gesamtbeurteilung.
- Obwohl in einigen wenigen Bewertungskriterien deutliche Unterschiede auftraten, beträgt die durchschnittliche Übereinstimmung von Studierenden- und Dozentenurteilen insgesamt 83%.

• 2. Methodische Kritik

- Während sich bei Analysen über alle Veranstaltungen bzw. der Gegenüberstellung der beiden Veranstaltungsarten genügend große Fallzahlen für präzise und aussagekräftige statistische Testung bildeten, reichten diese bei veranstaltungsspezifischen Analysen nicht aus.
- Inwieweit die Ergebnisse durch eine Tendenz zur Beantwortung nach sozialer Erwünschtheit in eine zu positive Richtung verfälscht sind, ist unklar. Alle Befragten wurden aber über die Anonymität der Befragung aufgeklärt und darum gebeten, ehrlich zu antworten.
- Außerdem muss berücksichtigt werden, dass studentische Beurteilungen nicht in jedem Fall etwas über die Qualität der Lehre aussagen. Es kann sich auch um Maß für studentische Akzeptanz oder Vorlieben handeln [6].

• 3. Diskussion einzelner Ergebnisse

Die objektive und standardisierte Kontrolle der Regelmäßigkeit von Veranstaltungen und ihrer Teilnehmerzahlen stellt ein so einfaches wie effektives Maß der Qualitätssicherung in der Lehre dar. In der vorliegenden Untersuchung wurden so eine hohe Zahl ungeplanter organisatorischer Veränderungen und hohe Ausfallquoten nachgewiesen, die auf Nachlässigkeiten in der Umsetzung der Lehrpläne schließen lassen. Daniel [21] und Rindermann [19] zeigen, dass Studierende solche organisatorischen Mängel deutlich negativ bewerten und im Rahmen möglicher Verbesserungsvorschläge am häufigsten aufführen.

Die studentischen Bewertungen der verschiedenen Merkmale der Veranstaltungen fallen durchaus differenziert aus: Obwohl die Formmerkmale insgesamt als gut beurteilt wurden, wurden der schlechte Informationsfluss, das räumliche Umfeld und die mangelnde Stoffvertiefung beklagt, letzteres etwas deutlicher für die Vorlesungen als für die Seminare ($d = .14$). Offensichtlich erwarten

Studierende auch von einer Vorlesung ein gewisses Maß an inhaltlicher Vertiefung.

Das identifizierte optimale Vorhersagemodell mit den Variablen "Abstimmung auf Voraussetzungen der Teilnehmer" und „interessanter Inhalt“ klärt fast die Hälfte der Varianz in der Gesamtbeurteilung auf. Dieses Ergebnis geht konform mit Befunden anderer Autoren [22], [2], [9], dass die Beurteilung von Lehrveranstaltungen neben objektiven Rahmenbedingungen, inhaltlicher und didaktischer Qualität maßgeblich vom Interesse der Urteiler mitbestimmt wird. Dabei handelt es sich um in der Lehre leicht steuerbare Merkmale, weil beide durch einfache Rückfragen der Dozenten an die Teilnehmer fortlaufend erfasst werden können.

Die hohe Übereinstimmung von Studierenden- und Dozentenurteil ist bemerkenswert, in einzelnen Merkmalen tritt maximal eine Abweichung von einer halben Standardabweichungseinheit auf. In der Literatur korrelieren Studierenden- und Dozentenurteile meist weniger hoch [2]. Es wäre denkbar, dass die Dozenten in ihren Ansichten tatsächlich mit denen der studentischen Teilnehmer übereinstimmen oder aber sich gut in deren Lage hineinversetzen können und die Fragen entgegen der Instruktion aus ihrer Sicht beantworteten. Dieses hohe Maß an Übereinstimmung spricht in jedem Fall dafür, dass sich die Lehrenden berechtigterweise auf ihre Sicht der Lehre verlassen können, um im Sinne der Teilnehmer zu handeln.

• 4. Ausblick

Der formative Evaluationsansatz bietet den Vorteil, durch die Datenerhebung zur Mitte des Semesters sowie die schnelle Ergebnismeldung direkt wirksame Maßnahmen zur Lehrqualitätsverbesserung im nächsten Semester einleiten zu können. Die Wirkung der Ergebnisumsetzung kann durch eine Wiederholungsbefragung anhand derselben Instrumente im Folgesemester geprüft werden. Erst die fortlaufende und systematische Evaluation von Lehrveranstaltungen trägt der Bedeutung der Lehre für das Fach Rechnung und hilft diese weiterzuentwickeln und zu optimieren. Der Gesetzgeber hat dies bereits erkannt und die Durchführung von Lehrevaluationen unter Berücksichtigung studentischer Beurteilungen in der ab Oktober 2003 gültigen ärztlichen Approbationsordnung verankert [23]. Dies wird in Zukunft besonders wichtig sein, weil traditionelle Lernformen (Vorlesung, Lehrbuch, Lernen am Krankenbett) mit multimedialen Lernmethoden, webbasierten Lern- und Trainingssystemen und virtuellen Patienten konkurrieren müssen [24].

Korrespondenzadresse:

- Dr. phil. Dipl.-Psych. Claus Barkmann, Zentrum für Frauen-, Kinder- und Jugendmedizin, Klinik und Poliklinik für Kinder- und Jugendpsychosomatik, Universitätsklinikum Hamburg-Eppendorf, Martinistraße 52, 20146 Hamburg, Deutschland, Tel.: 040/42803-3697, Fax: 040/42803-5105
barkmann@uke.uni-hamburg.de

Literatur:

- [1] El Hage N. Lehrevaluation und studentische Veranstaltungskritik. Bonn: Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie; 1996.
- [2] Rindermann H. Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur

- Evaluation computerbasierten Unterrichts. Landau: Verlag Empirische Pädagogik; 2001.
- [3] Kromrey H. Evaluation von Lehre und Studium - Anforderungen an Methodik und Design. In: Spiel C. Evaluation universitärer Lehre - zwischen Qualitätsmanagement und Selbstzweck. Münster: Waxmann; 2001:S21-59.
- [4] Reissert R. Evaluation der Lehre - aktuelle Aktivitäten an deutschen Hochschulen. Hannover: Hochschul-Informations-System; 1992.
- [5] Krause G. Besuchsziele und Veranstaltungsmerkmale - relevante Biasvariablen bei der Lehrevaluation? In: Krampen G, Zayer H (Hrsg). Psychologiedidaktik und Evaluation I. Konzepte, Erfahrungsberichte und empirische Untersuchungsbefunde aus Anwendungsfeldern der Aus-, Fort- und Weiterbildung. Bonn: Deutscher Psychologen Verlag; 1998:S317-384.
- [6] Kromrey H. Studierendenbefragung als Evaluation der Lehre? Anforderungen an Methodik und Design. In: Engel U. Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre. Frankfurt a. M.: Campus; 2001:S1-47.
- [7] Rindermann H. Untersuchungen zur Brauchbarkeit studentischer Lehrevaluationen. Landau: Verlag Empirische Pädagogik; 1996.
- [8] Schweer M, Rosemann B. Qualität der Lehre: Bedingungsvariablen des studentischen Urteils. Z Entwicklungspsychol Padagog Psychol. 1995;9(3-4):189-196.
- [9] Spiel C, Goessler PM. Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende. Z Entwicklungspsychol Padagog Psychol. 2000;14(1):38-47.
- [10] Pabst R, Nave H, Rothkötter H, Tschernig T. Lehrevaluation in der Medizin. Dtsch Arztebl. 2001;98(12):598-600.
- [11] Allert G, Gommel M, Tamulionyte L, Appelt M, Zenz H, Kaechele H. Das interdisziplinäre Längsschnittcurriculum "Medizinische Psychologie, Psychotherapie und Psychosomatik" an der Universität Ulm. Verlauf und Evaluation des klinischen Abschnitts. Psychother Psychosom Med Psychol. 2002;52(8):355-362.
- [12] Guendel H, Wefelmaier T. Das psychiatrisch-psychotherapeutische Unterrichtsgespräch im Erleben von Patient und gesprächsführendem Studenten. Nervenarzt. 1998;69(9):806-810.
- [13] Schueppel R, Bayer A, Hrabal V, Hoelzer M, Allert G, Tiedemann G, Hochkirchen B, Stephanos S, Kaechele H, Zenz H. Fachübergreifendes Längsschnittcurriculum "Medizinische Psychologie, Psychotherapie und Psychosomatik" Erfahrungen aus dem vorklinischen Abschnitt. Psychother Psychosom Med Psychol. 1998;48(5):187-192.
- [14] Inselmann U, Faller H, Lang H. Unterricht im Fach Psychosomatische Medizin und Psychotherapie: Beurteilung durch die Studenten. Psychother Psychosom Med Psychol. 1998;48(2):63-69.
- [15] Doering S. Gute Noten für das Praktikum "Medizinische Psychologie". Psychol in der Med. 1997;8(3):21-25.
- [16] Obliers R, Heindrichs G, Koehle K. Konzeption und Evaluation eines problemorientierten Erstsemester-Tutoriums in der Medizinpsychologie. In: Senf W, Heuft G (Hrsg). Gesellschaftliche Umbrüche - individuelle Antworten. Frankfurt a. M.: Verlag für Akademische Schriften; 1995:S261-282.
- [17] Stoessel U, Braehler E, Gostomzyk JG, Kochen MM, von Troschke J, Wildgrube K, Wirsching M. Evaluation von Lehrveranstaltungen in den psychosozialen Fächern und der Allgemeinmedizin im Studiengang Humanmedizin. Gesundheitswesen. 1996;58(4):249-252.
- [18] Lempa W, Kuensebeck HW, Nordmeyer J, Davin U, Witzig L, Liedtke R, Freyberger H. Empirische Belege zur Effektivität des studentischen Psychosomatik-Wahlfach-Modells der Medizinischen Hochschule Hannover. Psychother Psychosom Med Psychol. 1985;35(5):127-132.
- [19] Rindermann H. Skalen der Lehrevaluation: Welche Aspekte sollen in universitären Lehrveranstaltungen beurteilt werden? In: Krampen G, Zayer H (Hrsg). Psychologiedidaktik und Evaluation I. Konzepte, Erfahrungsberichte und empirische Untersuchungsbefunde aus Anwendungsfeldern der Aus-, Fort- und Weiterbildung. Bonn: Deutscher Psychologen Verlag; 1998:S295-316.
- [20] Barkmann C, Weidtmann K, Schulte-Markwort M. Evaluation als Mittel der Qualitätssicherung in der kinder- und jugendpsychiatrischen Hochschullehre. Z Klin Psychol Psychiatr Psychother. 2004;2(52):118-136.
- [21] Daniel H. Bewertung der Lehre aus Sicht der Studierenden und Absolventen. In: Müller-Böling D. Qualitätssicherung in Hochschulen. Gütersloh: Verlag Bertelsmann Stiftung; 1995:S160-185.
- [22] Berger U, Schleussner C. Hängen Ergebnisse einer Lehrveranstaltungs-Evaluation von der Häufigkeit des Veranstaltungsbesuchs ab?. Z Entwicklungspsychol Padagog Psychol. 2003;7(2):125-131.
- [23] Krüger-Brand H. E-Learning in der Medizin: Vor dem Durchbruch. Dtsch Arztebl. 2002;99(22):1491-1493.
- [24] Berger U, Schlußner C, Strauß B. Umfassende Lehrevaluation in der Medizin - eine Aufgabe für die psychosozialen Fächer?. Psychother Psychosom Med Psychol. 2003;53:1-78.