

## Kann man die Qualität von Lehrveranstaltungen messen?

### Workshop

„Studentische Veranstaltungskritik: Ziele und Methoden“  
anlässlich des Fakultätentags Informatik an der TU Ilmenau  
23. November 2006

Materialien zum download unter: [www.hkromrey.de](http://www.hkromrey.de) >> downloads

## Das Vortragsthema im Kontext des Workshopblocks „Studentische Veranstaltungskritik“:



**Kann man die Qualität von Lehrveranstaltungen durch Befragung der teilnehmenden Studierenden messen?**

Nach  
der  
Evalu-  
ation:



Der Dozent hat gelernt!

## Die grundlegende Frage: Wer bewertet was nach welchen Kriterien?



**Nicht jede Bewertung ist – im wissenschaftlichen Sinne – „Evaluation“. Und nicht jede Evaluation ist „Qualitätsmessung“.**

## Exkurs: Der Begriff „Evaluation“ [1]

### Begriffsverwendungen:

- **alltagssprachlich**
- **"politisch"**
- **empirisch/methodisch:**
- **in der Umfrageforschung**

### Bedeutung:

- **Irgend jemand BEWERTET irgend etwas irgendwie unter irgend welchen Gesichtspunkten**
- **Überprüfung eines Programms / einer Einrichtung durch ein zu diesem Zweck berufenes Expertengremium (Resultat: Evaluationsgutachten mit Empfehlungen)**
- **Design für einen spezifischen Typ empirischer Sozialforschung (experimentelle / quasi-experimentelle Programmforschung, Implementations- und Begleitforschung, Erfolgskontrolle, formative Evaluation)**
- **Erhebung und Auswertung bewertender (=“evaluierender“) Aussagen; Analogie zur Meinungsforschung (s. Politbarometer, Teilnehmerbefragungen in Seminaren)**

## Exkurs: Der Begriff „Evaluation“ [2]

### alltäglicher Sprachgebrauch

- Irgend etwas  
wird
- von irgend jemandem
- nach irgendwelchen  
Kriterien
- in irgend einer Weise  
bewertet

### wissenschaftlicher Sprachgebrauch

- Programme, Maßnahmen, Organisationen  
werden
- durch Personen, die zur Bewertung besonders  
befähigt erscheinen: Sachverständige, Experten,  
Betroffene etc.
- nach explizit auf den Sachverhalt bezogenen und  
begründeten Beurteilungskriterien und ggf.  
Standards
- in einem objektivierten Verfahren, im Rahmen  
eines im Detail geplanten Evaluationsdesigns  
bewertet

### Angestrebtes Ergebnis:

eine datengestützte, in methodisch „objektiverer“ Weise zustande  
gekommene Bewertung mit intersubjektivem Geltungsanspruch.

## Exkurs [3]: Typen von „Evaluation“

### Bezeichnung:

- "Forschungsparadigma"
- "Kontrollparadigma"
- "Entwicklungsparadigma"

### Funktionen/Ziele/Prioritäten:

- **Evaluation als Chance für die Gewinnung anwendungsorientierten und transferfähigen Grundlagenwissens, Bindeglied zwischen Theorie und Praxis; wissenschaftliche und methodische Kriterien stehen im Vordergrund**
- **Evaluation als Beitrag zur Planungsrationalität durch Erfolgskontrolle; weitere Kontrollform neben Rechtmäßigkeitskontrolle (Gerichte), politischer Kontrolle (Parlamente) und Wirtschaftlichkeitskontrolle (Rechnungshöfe); Priorität hat das Programm, methodische Ansprüche müssen ggf. zurückstehen**
- **Prozeßorientierung, Beitrag zur Programm- oder Organisationsentwicklung; Evaluation ist neben Forschung auch Beratung; Einbezug der Betroffenen und Beteiligten ("stakeholders"); flexibles, situationsoffenes Design**

## Zurück zur grundlegenden Frage: Wer bewertet was nach welchen Kriterien?



Sowie zur spezifischen Frage, dieses Vortrags:  
Kann man die Qualität von Lehrveranstaltungen messen?

Die klare Antwort lautet: „Es kommt darauf an“, nämlich auf unseren Begriff von „Qualität“. Denn:

- Das Konstrukt „Qualität“ müsste so definierbar sein, dass seine Dimensionen als Merkmale des Gegenstands „Lehrveranstaltung“ zu verstehen sind.
- Methodologisch wäre dann „nur“ noch die Aufgabe zu lösen, diese Qualitätsdimensionen so durch Indikatoren zu operationalisieren, dass sie als situationsunabhängige „Qualitäts-Messwerte“ gelten können.

*Allerdings: Über einen solchen Qualitätsbegriff, dessen Semantik einigermaßen unbestritten und der zugleich so konkret wäre, dass er operationalisierbar ist, verfügen wir nicht.*

Noch grundlegender ist das Problem, dass ein Konstrukt „Qualität“, verstanden als Merkmal des zu bewertenden Objekts selbst, sachlich unangemessen wäre:

- Qualität ist ein relationales Merkmal: Eignung, Brauchbarkeit, Güte *in Bezug auf* bestimmte Ziele und Zwecke sowie auf bestimmte Nutzer- und Klientengruppen.
- Als Relation zwischen Gegenstand (hier: Lehrveranstaltung) und Nutzer (oder andere stakeholder) sowie Nutzung (hier z.B.: Teilnehmer/innen und ihre aktuellen Ziele und Lernstile; die wissenschaftliche Profession und ihr „state of the art“; Absolventen und ihre beruflichen Qualifikationsanforderungen) lassen jedoch sich beliebig viele „Qualitäten“ herleiten und plausibel begründen. Welches Konstrukt soll gelten?

Exkurs: Es ist alles schon mal dagewesen!  
Die „Sozialindikatorbewegung“: Indikatoren zur  
Messung von „Lebensqualität“

- **Erfüllung vielfältiger Funktionen:**
  - a) Beschreibung gesellschaftlicher Strukturen, Ziele, Leistungen, Wandlungsprozesse (**Messfunktion**),
  - b) Evaluierung von a) (**Bewertungsfunktion**),
  - c) Analyse der Wechselwirkungen zwischen bereichs-spezifischen Strukturen und Leistungen, z.B. durch Input-Output-Tabellen (**Gesamtrechnungsfunktion**; Analogie zur VGR),
  - d) Verbreitung erklärender neuer Informationen über Kausalzusammenhänge (**Erklärungsfunktion**),
  - e) Institutionalisierung neuer Verfahren und Instrumente in der Planung (**Innovationsfunktion**).

## Exkurs: „Sozialindikatorenbewegung“ [2] Nicht einlösbare Anforderungen an Indikatoren

- a) Die einzubeziehenden Indikatoren müssen in einem **sachbezogenen Zusammenhang zu dem Problembereich** stehen, für den Aussagen getroffen werden sollen (**Gültigkeit**).
- b) Die vorherige Aufstellung von Hypothesen soll gewährleisten, dass die Zusammenhänge zwischen dem einzelnen Indikator und anderen Indikatoren bzw. dem Indikatorsystem sichtbar werden. Im Idealfall wären die Indikatoren aus einem **theoretischen Wirkungsmodell** abzuleiten. (**Theoriebezug**).
- c) Es ist zu begründen, **warum** eine **bestimmte Maßzahl** zur Präsentation ausgewählt wurde (**Offenlegung des Wertbezugs**).
- d) Vor allem solche Indikatoren sind heranzuziehen, die für Gegenwart und Zukunft – möglichst auch für die Vergangenheit – erreichbar sind (**Zeitbezug, Dauerhaftigkeit**: Planung ist auf Dauer angelegt).
- e) Die Indikatoren sollten sich auf solche **Sachverhalte** beziehen lassen, die durch politische Maßnahmen **beeinflussbar** sind („**policy variables**“; **Praxisbezug, Instrumentenbezug**).
- f) Die Indikatoren müssen sinnvoll **aggregierbar und disaggregierbar** sein.
- f) **Gültigkeit und Zuverlässigkeit** müssen **abschätzbar** sein.

- **Probleme bei der analogen Anwendung auf die Aufgabe „Evaluation durch Qualitätsmessung“:**
- Was ist Qualität? (Wertproblem)
- Wer entscheidet über Qualitätskriterien und Standards? (Legitimierungsproblem)
- Können die relevanten Dimensionen durch Indikatoren gültig und zuverlässig gemessen werden? Oder richtet sich die Auswahl der Indikatoren nach ihrer leichten Messbarkeit? (ernsthafte Forschung oder „quick and dirty“?)
- Tragen die Messwerte zur Qualitätsverbesserung bei? Oder wird stattdessen die Optimierung von Indikatorwerten betrieben?
- **Fazit: *Qualität* definiert, wer die Indikatoren festlegt.**

## **Wozu** eigentlich Lehrveranstaltungs-Qualität **messen?** ▶▶ für die Ziele des Kontrollparadigmas:

### Bezeichnung:

- "Forschungsparadigma"
- "Kontrollparadigma"
- "Entwicklungsparadigma"

### Funktionen/Ziele/Prioritäten:

- Evaluation als Chance für die Gewinnung anwendungsorientierten und transferfähigen Grundlagenwissens, Bindeglied zwischen Theorie und Praxis; wissenschaftliche und methodische Kriterien stehen im Vordergrund
- Evaluation als Beitrag zur Planungsrationalität durch Erfolgskontrolle; weitere Kontrollform neben Rechtmäßigkeitskontrolle (Gerichte), politischer Kontrolle (Parlamente) und Wirtschaftlichkeitskontrolle (Rechnungshöfe); **Priorität hat das Programm, methodische Ansprüche müssen ggf. zurückstehen**
- **Prozeßorientierung, Beitrag zur Programm- oder Organisationsentwicklung; Evaluation ist neben Forschung auch Beratung; Einbezug der Betroffenen und Beteiligten ("stakeholders"); flexibles, situationsoffenes Design**

Sollten wir nicht lieber fragen: Kann man die Qualität von Lehrveranstaltungen verbessern? (Kontroll- zum Entwicklungsparadigma)



Im Kontext dieses Workshops also:

Kann man die Qualität von Lehrveranstaltungen durch Befragung der teilnehmenden Studierenden verbessern?

**Dafür** können Befragungen der Studierenden viele wertvolle Informationen liefern:

- über die Studierenden selbst (über ihre Lernziele, Teilnahmegründe und Interessen, ihre Erwartungen an die Lehre, ihre Vorkenntnisse, Lern- und Arbeitsstile, ihr Zeitbudget u.v.m.)
- über die Nutzung der „Lernhilfen“ (Vorlesung, Übung, Bücher etc.) und deren Eignung für das je individuelle Lernverhalten (also hier: die „Lehrevaluation“)
- über die soziale Lernumgebung (Beratung, Lern-“Klima“)

**Auf der Basis solcher Informationen kann zwar auch nicht Qualität „gemessen“ werden, jedoch das jeweils erreichbare Optimum an Qualität zwischen den Beteiligten ausgehandelt werden.**

## Doch zurück zum Messproblem:

- In einem Dozentenkurs des HDZ Essen von 1980 hieß es optimistisch: ***"Urteile (Schätzungen) von Studenten über die Lehre ... sind - wenn man etwa 20-30 Studenten urteilen lässt - zuverlässig wie professionelle Testverfahren"***.

Und: ***"Sie sind von anderen Merkmalen der Studenten selbst und der Dozenten wenig beeinflusst."*** (Schmidt, J.: Evaluation. I. Evaluation als Diagnose, Essen 1980, 51 f.)

Sofern dies zuträfe, wäre die Qualitätsmessung per Befragung der Königsweg zur Lösung aller Probleme: Die unmittelbar Betroffenen sind schließlich – so kann dann argumentiert werden – die besten Experten; sie wissen aus eigener Erfahrung, wie gut oder wie schlecht die Lehrveranstaltung ist, der sie ausgesetzt sind.

- **Leider alles falsch! Nicht nur aus Gründen der methodischen Schwierigkeit:**
  - ***Standardisiertes Befragen*** ist – entgegen einem gängigen Vorurteil – ein sehr problematisches Instrument der Informationsgewinnung; es müssen hinsichtlich des Erhebungsgegenstandes und der Forschungskontakt-situation sehr anspruchsvolle Voraussetzungen erfüllt sein, die fast immer nicht erfüllt sind.
  - Noch problematischer wird es, wenn ***Befragung als Instrument des „Messens“*** eingesetzt werden soll, und am schwierigsten, wenn ***Befragung als Methode des intersubjektiven Bewertens*** (d.h. des Messens von Qualität) dienen soll.

## Exkurs: generelle methodische Anforderungen an Messung durch Befragung

### formale Voraussetzungen, die für eine „Messung“ erfüllt sein müssen:

- der „Gegenstand“ (das Objekt) der Beurteilung ist eindeutig definiert
- das zu messende „Merkmal“ (Variable) ist eindeutig definiert und operationalisiert
- eine „Mess-Skala“ (Vergleichsmaßstab) existiert und ist eindeutig definiert (incl. eindeutiger Skalen-Endpunkte sowie unterscheidbarer Abstufungen zwischen den Skalen-Endpunkten)
- Befragte sind in der Lage, den „Gegenstand“ intersubjektiv übereinstimmend zu identifizieren, das zu messende „Merkmal“ intersubjektiv übereinstimmend zu erkennen (identische Semantik + identische Perspektive) und die „Mess-Skala“ in intersubjektiv übereinstimmender Weise darauf anzuwenden (= Messung durch Vergleich der Merkmalsausprägung eines Objekts mit einem Vergleichsmaßstab, Übersetzung dieses Vergleichs in Skalenpunkte)

## Exkurs [2]: zusätzliche Anforderungen an „evaluierendes“ Messen durch Befragung

Erfüllt sein müssen zunächst die formalen Voraussetzungen für eine „Messung“; hinzu kommen als weitere Anforderungen:

- Neben dem „*Kriterium*“ der Beurteilung („Qualitäts-Merkmal“) ist ein „*Standard*“ der Beurteilung zu definieren: Wann ist etwas „gut“, wann „schlecht“, wann „mittelmäßig“? etc.
- Dieser „*Standard*“ ist von allen Evaluierenden in intersubjektiv übereinstimmender Weise anzuwenden.
- Und schließlich ist auch noch die grundlegende Entscheidung zu treffen (und dem Befragten erkennbar zu machen):
  - Soll die befragte Person Auskunft über sich selbst und ihren Bezug zum Gegenstand der Bewertung geben? (“Akzeptanzaussage”)
  - Oder soll sie eine “quasi-objektive” Beurteilung des zu bewertenden Gegenstands vornehmen? (“Evaluationsaussage”)

- **Irrig ist die Vorstellung einer „intersubjektiven Messung“ von Qualität vor allem aus erkenntnistheoretischen Gründen.**

**Notwendige Bedingung wäre nämlich das Zutreffen eines der folgenden erkenntnistheoretischen Axiome:**

- **Entweder (wie schon angesprochen): Qualität ist ein „objektives“ Merkmal des Sachverhalts selbst; seine Ausprägung kann durch abbildende subjektive Wahrnehmung ohne systematische Verzerrung „gemessen“ werden (so dass der Erwartungswert bei hinreichend großer Zahl von Messungen dem „*wahren objektiven Wert*“ entspricht).**
- **Oder: Qualität ist ein dem Menschen gegebenes, intersubjektiv gültiges Konzept (bzw. anders formuliert: alle Menschen bewerten nach gleichen Kriterien und Standards in gleicher Weise). In konkreten Situationen auftretende Unterschiede zwischen Bewertern wären dann als Zufallsvarianz anzusehen (so dass der Erwartungswert bei hinreichend großer Zahl von Messungen dem „*wahren subjektiven Wert*“ entspricht).**

## aus dem Leben der Bewertungs-Forschung (Beispiele):



"Meinen Sie, daß Attila eine hervorragende, eine gute, eine mittelmäßige oder eine unfähige Führungspersönlichkeit ist?"

Tabelle: S04 Frage 4: Erwartungen erfüllt?  
by INTRINS Intrinsische Motivation<sup>2</sup>

		Count • INTRINS			
Col	Pct	•extrins.	intrins.	intrins.,	Row
		•	+ Pfl.	o.Pfl.	
		• -1	• 0	• +1	• Total
S04		.....			
	-	• 1735	• 373	• 102	• 2210
negative Urteile		• 30,7%	• 11,5%	• 7,7%	• 21,6%
		.....			
	0	• 1551	• 619	• 195	• 2365
		• 27,4%	• 19,0%	• 14,6%	• 23,1%
		.....			
	+	• 2370	• 2262	• 1036	• 5668
positive Urteile		• 41,9%	• 69,5%	• 77,7%	• 55,3%
		.....			
	Column	5656	3254	1333	10243
	Total	55,2%	31,8%	13,0%	100,0%

**Fazit: Eigenes Interesse stimmt gnädig!**

Tabelle 2: S04 Frage 4: Erwartungen erfüllt?  
by TEILNMOT Teilnehmermotivation in der Vorles

		TEILNMOT					
Count		> 90 %	> 80 %	gemischt	> 70 %	> 75 %	
Col	Pct	extrins.	extrins.		intrins.	intrins.	Row
		-2	-1	0	+1	+2	Total
S04							
Erwartungen	-2	151	257	135	20	1	564
überh.nicht erf.		16,1	9,4	3,0	1,3	,2	5,5
	-1	266	658	604	107	11	1646
		28,3	24,0	13,4	6,9	2,2	16,1
	0	266	787	1033	238	41	2365
		28,3	28,7	22,9	15,4	8,2	23,1
	+1	216	860	2006	720	172	3974
		23,0	31,4	44,4	46,5	34,5	38,8
Erwartungen	+2	40	177	739	464	274	1694
voll erfüllt		4,3	6,5	16,4	30,0	54,9	16,5
Column		939	2739	4517	1549	499	10243
Total		9,2	26,7	44,1	15,1	4,9	100,0

**Fazit: Die peer-group fällt das Urteil!**

Tab.: Förderung von Interesse und Teilnahmemotivation

		. TEILNAHMEMOTIVATION			
		. extrins.	. intrins.	. intrins.,	
			+ Pfl.	o. Pfl.	Zeile
INTERESSE	GEFÖRDERT?	-1	0	1	Total
	-2	1459	168	57	1684
	überhaupt nicht	25,1%	5,1%	4,2%	16,1%
	-1	1213	260	99	1572
		20,9%	7,9%	7,4%	15,1%
	0	1537	531	200	2268
	teils/teils	26,5%	16,1%	14,9%	21,7%
	+1	1464	1725	445	3634
		25,2%	52,4%	33,1%	34,8%
	+2	130	608	544	1282
	ja, sehr	2,2%	18,5%	40,4%	12,3%
Spalte		5803	3292	1345	10440
Total		55,6%	31,5%	12,9%	100 %

**Fazit: Interesse, das nicht schon vorhanden ist, kann kaum „gefördert“ werden kann!**

Tab 2: Förderung von Interesse und vorherrschende Teilnehmermotivativ der Vorlesung (alle Befragten)<sup>4</sup>

		VORHERRSCHENDE TEILNEHMERMOTIVATION <sup>5</sup>					Zeile Total
		> 90 % extrins.	> 80 % extrins.	gemischt	> 70 % intrins.	> 75 % intrins.	
<b>INTERESSE</b>		-2	-1	0	1	2	
<b>GEFÖRDERT?</b>							
	-2	383	721	509	69	2	1684
	<b>überhaupt nicht</b>	39,8%	25,8%	11,0%	4,4%	0,4%	16,1%
	-1	228	520	676	130	18	1572
		23,7%	18,6%	14,6%	8,4%	3,6%	15,1%
	0	190	607	1126	277	68	2268
	teils/teils	19,7%	21,7%	24,4%	17,8%	13,6%	21,7%
	+1	138	783	1794	732	187	3634
		14,3%	28,0%	38,8%	47,0%	37,5%	34,8%
	+2	24	168	518	348	224	1282
	<b>ja, sehr</b>	2,5%	6,0%	11,2%	22,4%	44,9%	12,3%
	Spalte Total	963	2799	4623	1556	499	10440
		9,2%	26,8%	44,3%	14,9%	4,8%	100 %

**Fazit: Die peer-group ist der „Chefmotivator“ (bzw. –demotivator)!**