

## TEILNEHMERBEWERTUNG VON VORTRÄGEN

G. P. Burkhard; H. E. Renschler, Bonn

Neben der Ausbildung kommt der Fortbildung zur Qualitätssicherung der ärztlichen Tätigkeit eine wichtige Bedeutung zu. Hier wie dort sollte es um die Feststellung sowie Steigerung der Effektivität von Lernprozessen, dementsprechend um die Optimierung des Lernverhaltens, nicht zuletzt aber auch um die des Lehrerverhaltens gehen. Gerade diese Frage steht im Gegensatz zu anderen Ländern bei uns stark im Hintergrund. Es wird fast immer darüber diskutiert, was vermittelt werden soll, jedoch nicht wie es vermittelt werden soll, um eine größtmögliche Effektivität zu erreichen.

Hier wird darauf vertraut, daß der Lehrer ankommt und die Fähigkeit, etwas weiterzugeben, beherrscht. Wenn man dem Lehrer klarmachen könnte, daß dieses Gottvertrauen unbegründet ist und in den Augen der Zielgruppe vielfach voll-

kommen fehl am Platz ist, wäre viel erreicht. Erst das Bewußtmachen bzw. -werden einer Situation macht ihre Veränderung möglich.

Ausgehend von dieser Grundsituation entstand die Frage, wie ein Evaluationsansatz zur Feststellung der Lehrerqualität durch die Teilnehmer aussehen könnte, der die Forderung der Reliabilität, Validität, Praktikabilität erfüllt und der zu Entscheidungen hinsichtlich dieses Lehrers bzw. Vortragenden führt. Im folgenden soll berichtet werden über einen entsprechenden Versuch, der seit 1980 im Rahmen eines von der Akademie für Ärztliche Fortbildung Hessen abgehaltenen Notfalldienstseminars durchgeführt wird.

### 1. Sinn und Möglichkeit der Unterrichtsevaluation

Bei der Planung und Implementierung von Unterrichtsaktivitäten ist die Evaluation eine notwendige Komponente in diesem Prozeß. Trotzdem wird die Evaluation oft bewußt oder nicht bewußt vernachlässigt. Die Gründe für das Unterlassen scheinen dabei vielfältiger Natur zu sein.

a) Zunächst wird der Prozeß der Evaluation oft nicht verstanden. Die umfangreiche Literatur und die Breite dieses Gebietes führen oft zur Verwirrung.

b) Die Wichtigkeit der Evaluation wird vielfach nicht richtig erkannt oder bei richtiger Einschätzung dieses Prozesses werden die Ergebnisse nicht benötigt. Damit erscheint die Evaluation als nutzloser Prozeß.

c) Der Prozeß der Evaluation ist oft recht arbeits- und damit zeitintensiv, wenn er nicht systematisch und regelmäßig durchgeführt wird.

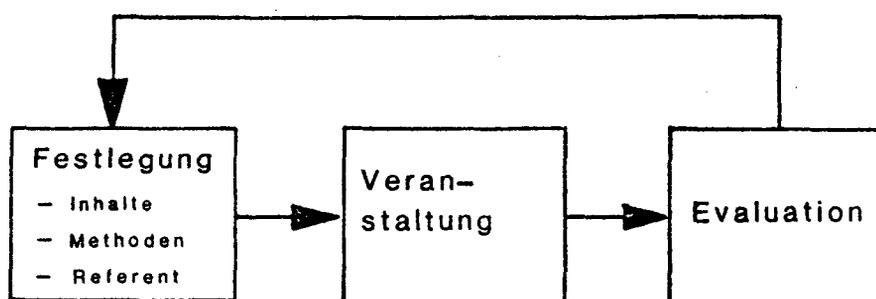
Drei Gründe scheinen jedoch für eine derartige Evaluation zu sprechen:

- a) Der Erfolg eines Unterrichtes kann erst hierdurch sichergestellt werden. Aspekte, die ineffizient sind, müssen bzw. können modifiziert werden.
- b) Der Aufwand für einen Unterricht bzw. Veranstaltung kann nachträglich gerechtfertigt werden.
- c) Es kann nach und nach der Standard erhöht werden.

Nach STUFFELBEAM 1971 bedeutet Evaluation "die Gewinnung von Informationen durch formale Mittel, wie Kriterien,

Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällen von Urteilen in Entscheidungssituationen zu erhalten".

Sie ist demnach nicht Selbstzweck, sondern soll helfen über Entscheidungen zu Verbesserungen zu gelangen. Nach Festlegung der Inhalte, Methoden, Referenten sollte eine Veranstaltung konzipiert werden und die Evaluation anschließend Auskunft geben über die erreichte Qualität. Diese Ergebnisse müssen im Sinne eines Kreisprozesses wieder zu Veränderungen führen (Abb. 1).



## **ABB.1: Bedeutung der Evaluation im Rahmen des Unterrichts**

### 2. Operationalisierung

Der vorgegebene Rahmen der Evaluation betraf das eingangs erwähnte Notfalldienstseminar der Landesärztekammer Hessen, das seit 1979 etwa 15 Mal durchgeführt wurde und weiter durchgeführt wird. Es sollte die daran teilnehmenden Ärzte durch 19 Referate zu 5 verschiedenen Notfallthemenbereichen für Notfallbehandlung kompetenter machen. Die Aufgabe bestand in einer Mitfestlegung von Inhalten aufgrund von Analysen von Notfallstatistiken und Befragung der Teilnehmer,

der Mitgestaltung der Veranstaltung, der Schulung der immer wieder eingesetzten Referenten und schließlich in der Evaluation der Veranstaltung. Dieser letzte Aspekt zielte einmal auf den Lerneffekt der Veranstaltung ab, zum anderen auf die Einschätzung der Referenten durch sogenannte Experten und durch Teilnehmer. Diese Einschätzungen wurden für Rückmeldungen an die Referenten herangezogen, damit zur eingangs geforderten Bewußtmachung der eigenen Lehrerkompetenz

und zur langfristigen Verbesserung der Qualität des Vortragsverhaltens. Damit hatte dieser Evaluationsansatz Entscheidungscharakter.

### 3. Evaluationskriterien der Teilnehmerbewertung

Eine Durchsicht der relevanten deutschen Literatur ergab kaum Hinweise auf Kriterien für eine Bewertung von Vortragenden. Im amerikanischen Bereich fanden sich verschiedene Bewertungssysteme. Beispielhaft seien die 24 Items umfassende Skala 'Rating your lecture' von FOLEY vom Center for Educational Development, University of Illinois, 1976 oder der 'Audience Response Survey' vom Department of Postgraduate Medical Studies mit 12 Items genannt. Oft sind diese Beurteilungsteile jedoch so komplex und differenziert, daß vor der Einschätzung ein sogenanntes Ratertraining erfolgen muß. Dies kann nicht das Ziel sein.

Wie Popham 1975 darstellt, hat es seit 1900 verschiedene Versuche gegeben, befriedigende Kriterien für Effektivität eines Lehrers bzw. Vortragenden zu finden. Bis heute gibt es jedoch keine einheitlichen Kriterien, die sowohl ihre Brauchbarkeit, als auch ihre Genauigkeit und Gültigkeit unter Beweis gestellt haben. Die an derartige Einschätzungs-

### 4. Stichprobe

Seit 1981 kommen die Einschätzungen nach Relevanz hier 'Notfallrelevanz' und 'Lernerfolg' zum Einsatz, ab 1983 das Kriterium des Gefallens. Die im folgenden dargestellten Ergebnisse basieren auf etwa 1400 teilneh-

Aus dieser gesamten Thematik soll entsprechend der Themenstellung nur die Teilnehmerbewertung weiterverfolgt werden.

kriterien gestellten Anforderungen waren:

- wenige Items
- für Rater gut verständliche Items
- quantitatives Urteil für
  - Feedback an Referenten
  - Rangbildung unter den Referenten.

Als Kriterien, die diese Forderung erfüllen sollten, wurden gewählt:

1. Wie relevant war das Referat, der Vortrag in Bezug auf das Unterrichtsziel?
2. Wie hoch war der Lernerfolg?
3. Wie hat der Vortrag/Vortragende gefallen?

Da eine Aussage bezüglich der Ausprägungsstärke der genannten Kriterien gemacht werden sollte, wurde eine ordinale 5-Punkte-Skala gewählt, die von 1 = sehr hoch bis 5 = überhaupt nicht reichte.

Diese Einschätzungen wurden ergänzt durch qualitative Aussagen, wie freie Angaben und Persönlichkeitsangaben.

menden Ärzten bei 7 Veranstaltungen mit jeweils 19 Referenten. Übergreifend betrug der Rücklauf etwa 76 %. Es kann somit von einer zufriedenstellenden Datenbasis ausgegangen werden.

## 5. Ergebnisse

Jedes Erfassungsinstrument muß, um die Bewertungsfunktion erfüllen zu können, eine gewisse Meßgenauigkeit und Gültigkeit unter Beweis stellen. Meßgenau ist ein Instrument bzw. Item dann, wenn es bei Wiederholung am gleichen Objekt zum gleichen Ergebnis kommt. Dies ist im vorliegenden Fall nur schwer machbar, da dieselben Teilnehmer denselben Referenten mit identischem Vortragsverhalten in der Regel nicht noch einmal sehen und bewerten. Ein Maß für die Meßgenauigkeit scheint im Grad der Obereinstimmung der Urteile hinsichtlich derselben Referenten zu liegen. Neben den Methoden, die von FERMAN, LANCRENON und GREMY (1979) referiert wurden und teilweise auf der Methode des  $\chi^2$ -Testes und teilweise auf Korrelationsberechnung basieren, gibt es ein von HORST (1949) vorgeschlagenes Verfahren zur Ermittlung der 'Reliabilität des Validitätskriteriums'. Dieses Verfahren macht eine Wiederholung der Beurteilung überflüssig, indem es sich auf die innere Konsistenz der subjektiven Schätzwerte stützt, wobei die Reliabilität der Durchschnittsbeurteilungen berechnet wird. Danach ergeben sich z.B. für die Veranstaltung im April 1984 folgende Koeffizienten:

Relevanz $r_{cc}$	=	0,95
Lernerfolg $r_{cc}$	=	0,97
Gefallen $r_{cc}$	=	0,99

Diese Ergebnisse müssen gemessen am Maximalwert von +1 als sehr hoch bezeichnet werden und lassen den Schluß auf eine sehr hohe Raterübereinstimmung, d.h. Meßgenauigkeit zu. Ähnlich hohe Werte ergeben sich für die anderen Veranstaltungen. Bei der Überprüfung der Validität ergeben sich in der Praxis noch sehr viel größere Schwierigkeiten. Als valide wird ein Instrument dann ange-

sehen, wenn es das mißt, was es zu messen beansprucht. Diese Überprüfung erfolgt an einem Außenkriterium, von dem angenommen wird, daß es Gültigkeit besitzt. Als Außenkriterium benutzt man Expertenurteile, die bei kritischer Betrachtung jedoch auch nicht valide sein müssen. Immerhin ist diese Methode ein operationaler Weg. In unserem Falle wurde ebenfalls auf eine Expertenbewertung zurückgegriffen. Bei der Frage des Zusammenhanges zwischen den Kriterien der Teilnehmerbewertung und der Expertenbewertung ergibt sich aus Daten von 7 Veranstaltungen ab 1981 folgendes Bild:

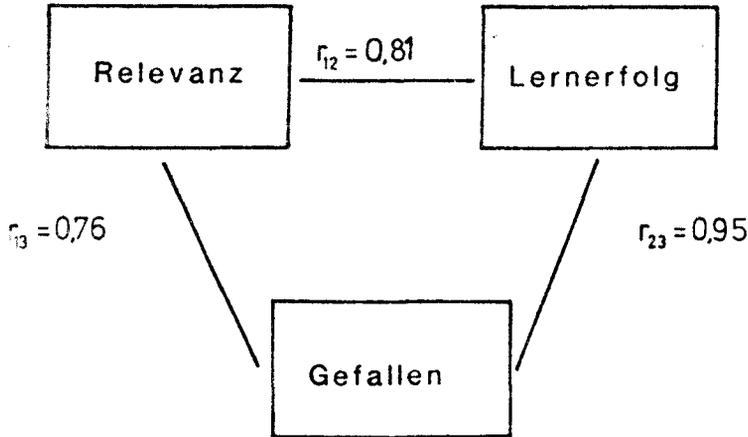
Relevanz $r_{TE}$	=	0,65
Lernerfolg $r_{TE}$	=	0,65
Gefallen $r_{TE}$	=	0,77

Alle Koeffizienten zeigen hochsignifikante Zusammenhänge. Dennoch scheint das Kriterium 'Gefallen' die für das Expertenurteil gültigste Größe darzustellen.

Eine weitere Frage betraf den Zusammenhang der drei Kriterien untereinander. Sollte sich eine hohe Interkorrelation herausstellen, so ist zu fragen, ob nicht nur ein Kriterium die Qualität des Vortrages bzw. Vortragenden erfassen konnte. Entsprechend der These, wonach im ganzheitlichen Erleben eine gewisse Obereinstimmung bei den Einzelgliedern vorzufinden ist, müßte sich ein hinter den Kriterien stehender ganzheitlicher Faktor finden lassen, der die Einzelkriterien determiniert. Viele vermuten einen sogenannten Happiness-Index (BERTRAM 1977) als gemeinsamen Faktor. Empirisch zeigen sich folgende Ergebnisse (Abb. 2).

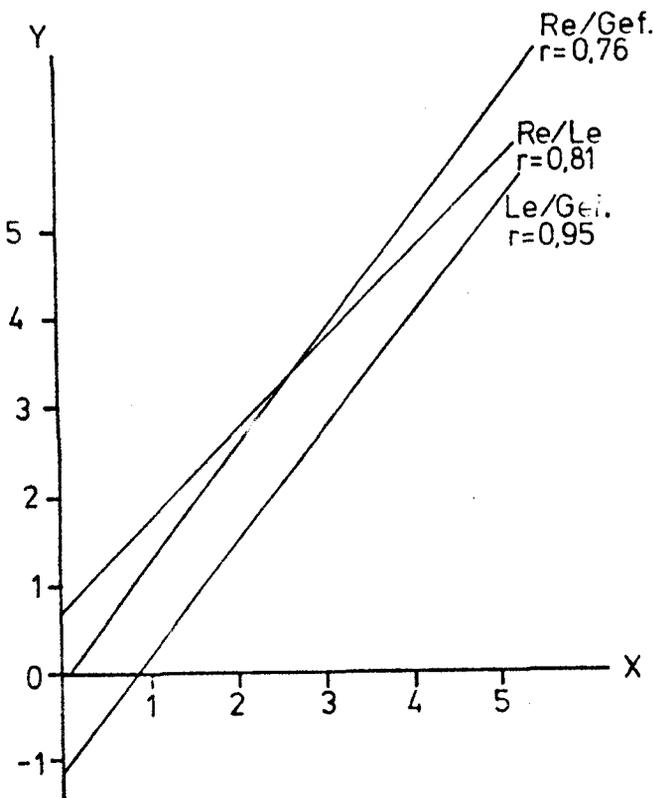
Folgendes Bild, in dem die Regressionsgeraden für alle drei Zusammenhänge dargestellt sind, mag diese veranschaulichen (Abb. 3).

**ABB.2: Interkorrelation zwischen Relevanz, Lernerfolg, Gefallen** **April 84**



**ABB.3**

Zusammenhang zwischen Relevanz/Gefallen  
Relevanz/Lernerfolg  
Lernerfolg/Gefallen  
April 1984



Obwohl alle drei Koeffizienten hochsignifikant sind, zeigt sich bei dem Zusammenhang Lernerfolg/Gefallen am ehesten ein gemeinsamer Faktor. Offensichtlich ist die Frage nach der Relevanz eines Themas unabhängig zu sehen. Inhaltlich ist dies plausibel. Die Relevanz eines Themas hat noch nichts mit dessen Lernerfolg zu tun. Andererseits wird ein Vortrag, bei dem viel gelernt wurde auch gefallen haben und umgekehrt.

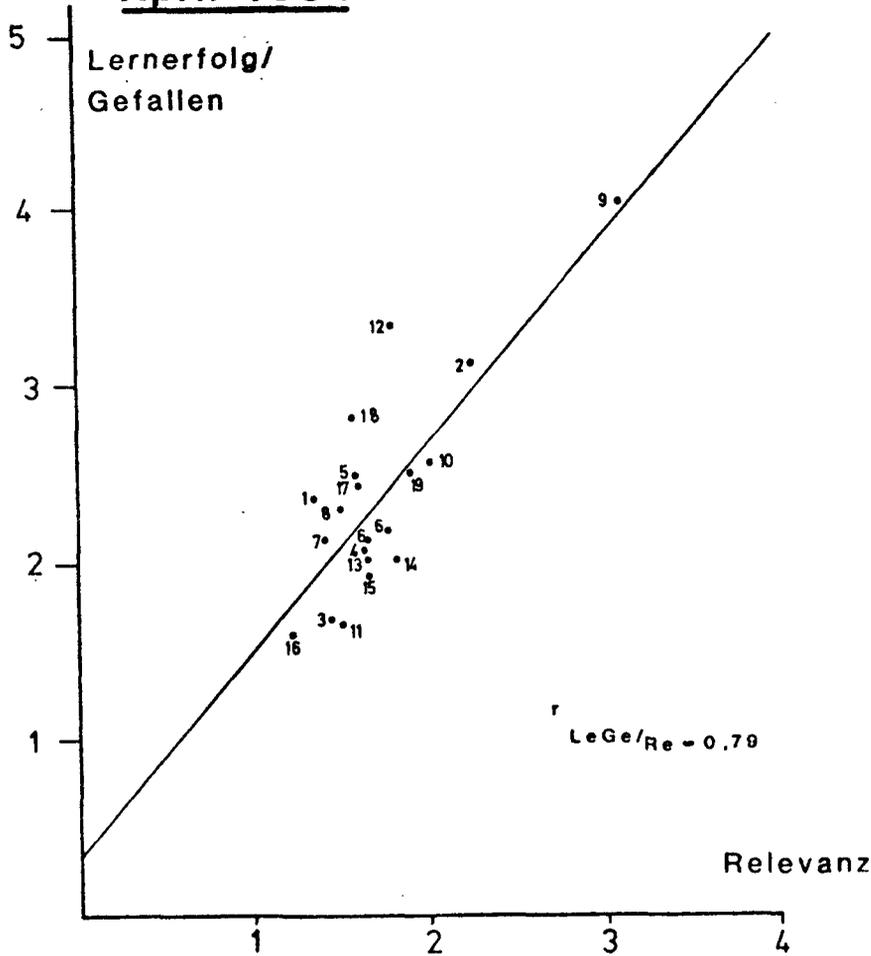
Insgesamt sprechen diese Ergebnisse für die Anwendbarkeit der vorliegenden Kriterien.

An einem konkreten Beispiel von April 1984 mag das verdeutlicht werden (Abb. 4).

ABB.4

Zusammenhang zwischen  
Lernerfolg/Gefallen und Relevanz

April 1984



Auf der y-Achse sind Gefallen und Lernerfolg rechnerisch über Mittelwert zusammengefaßt, auf der x-Achse die Relevanz.

Entsprechend der Aufgabe der Referentenevaluation wäre beispielsweise ein Referent entweder einer weiteren Analyse und Schulung zuzuführen, dessen Thema eine hohe Relevanzeinschätzung ergibt, der jedoch bei Lernerfolg und Gefallen schlecht abschneidet. Wie auf der Abbildung zu sehen ist, wären die Referenten 9, 12, 2, 18 zunächst einer weiteren Analyse zu unterwerfen, einer Schulung zuzuführen oder bei Mißerfolg dieser Schulung auszuschließen.

Damit wäre die Forderung erfüllt, die STUFFELBEAM in seiner Bestimmung der Evaluation genannt hat. Evaluation darf danach nicht Selbstzweck sein, sondern muß Informationen liefern, die Entscheidungen möglich machen.

Zusammenfassung

Ausgehend von der Bedeutung der Unterrichtsevaluation wurde als ein Bereich die Evaluation des Lehrers bzw. des Referenten ausgegliedert und eine Möglichkeit in der Teilnehmerbewertung gesehen.

Die von uns vorgeschlagenen Bewertungskriterien der Relevanz eines Themas gemessen am Ziel, des Lernerfolges und

Gefallens wurden anhand von empirischen Daten der Jahre 1981-1984 aus den Notfalldienstseminaren auf Reliabilität, Validität, Interkorrelation überprüft und für sehr brauchbar befunden. An einem konkreten Beispiel wurde ihre Funktion als Entscheidungsgrundlage bei der Evaluation aufgezeigt.

LITERATUR

1. ALPEROVITSCH, A., de BOMBAL, F.T., GREMY, F. (Ed.)  
Evaluation of Efficacy of Medical Action  
Amsterdam : North Holland Publishing Company, 1979
2. BERTRAM, D.A., BROOKS-BERTRAM, D.A.  
The Evaluation of Continuing Medical Education: A Literature Review  
Health Education Monographs 5, 1977
3. FERMANIAN, J., LANCRENON, S., GREMY, F.  
Companion of Several Coefficients of Agreement between Two Judges Used in Medical Decision.-  
In: Evaluation of Efficacy of Medical Action  
ALPEROVITSCH, A., de BOMBAL, F.T., GREMY, F. (Ed.)  
Amsterdam : North-Holland Publishing Company, 1979
4. FOLEY, R., SMILANSKY, J., BUCHMAN, E., SAJID, A.  
Improving Lecture Skills, Instructional Guide  
The Center for Educational Development, Univ. of Illinois at the Medical Center  
Chicago : Aldine Publishing Company, 1976
5. HORST, A.P.  
A Generalized Expression for the Reliability of Measures  
Psychom. 14 : 21
6. LIENERT, G.A.  
Testaufbau und Testanalyse  
Weinheim : Verlag Julius Beltz, 1969
7. POPHAM, W.J.  
Educational Evaluation  
New Jersey : Prentice Hall, 1975
8. SOMMER, J., PETERMANN, Fr.  
Deskriptive und präskriptive Aussagen in einem Fragebogen zur Beurteilung akademischer Lehrveranstaltungen  
Zeitschrift für Entwicklungspsychologie und Päd. Psychologie 1978, Band X, Heft 4, S 336-346
9. STUFFELBEAM, D.L., FOLEY, W.J., GEPHART, W.J., GUBA, E.G., HAMMOND, R.L., MERRMAN, H.O., PROUNS, M.M.  
Educational Evaluation and Decision Making  
Itasca, I 11.: F.E. Peacock, 1971
10. STUFFELBEAM, D.L.  
Alternative Approaches to Educational Evaluation  
In: Evaluation in Education: Current Application  
Berkeley, Calif. : Mc Cutchan Publishing , 1974
11. Audience Response Survey  
Department of Postgraduate Medical Studies  
Manchester, M60 7LP

Dr. G. B. Burkhard und Prof. Dr. H. E. Renschler, Institut für Didaktik der Medizin, Sigmund-Freud-Straße 25, D 5300 Bonn 1