

»» Lehrevaluation ist mehr als Marketing – Ein Kommentar zur Postwurfsendung der Fa. ScienceConsult

C. Stosch¹, V. Fischer²

¹ Studiendekanat der Medizinischen Fakultät der Universität zu Köln

² Rektorat der Medizinischen Hochschule Hannover

Zusammenfassung: Die deutschen Fakultäten werden von legislativer Seite seit einigen Jahren zunehmend mit Anforderungen zur Evaluation von Studium und Lehre (häufig im Sinne studentischer Veranstaltungskritik) konfrontiert. Die Diskussion hat sich von anfänglich undifferenzierter Ablehnung in Richtung einer Methodenkritik solcher Verfahren entwickelt. In vielen Fakultäten wird überlegt, Ressourcen zur Implementierung von Evaluationsverfahren vorzuhalten. Die vermutete Dienstleistungslücke wird zunehmend auch von privaten Anbietern zu füllen versucht. Im Rahmen der Ressourcenoptimierung sollte eine Verlagerung der Evaluationen auf externe Anbieter zwar durchaus in Betracht gezogen werden. Die Qualität der Anbieter kann jedoch sehr unterschiedlich sein. Am Beispiel einer vor kurzem in allen medizinischen Fakultäten eingetroffenen Postwurfsendung der Fa. ScienceConsult werden mögliche Schwachstellen solcher Angebote aufgedeckt.

Schlüsselwörter: Evaluation – Methoden

Evaluation of Teaching and Learning is More than Marketing – Comment on Direct-Distribution Advertising by ScienceConsult Company: For some years German faculties have been confronted with increasing governmental demand for the evaluation of teaching and learning. This discussion developed from undifferentiated refusal in the past into the present day methodological criticism of such procedures. Today, in many faculties, it is considered to be helpful to reserve funds for the implementation of such evaluation procedures. Evaluation services are increasingly offered by private vendors. Within the framework of an optimal resource allocation, a shift to external providers of evaluation procedures should be taken into account. The quality of the providers, however, can be very divergent. Taking the bulk mail from the ScienceConsult Company as an example – recently arrived in all Medical Faculties in Germany – potential deficiencies of such offers are critically discussed.

Key words: Evaluation – Methods

Mit Datum vom 2. Juli 2001 ist in den Dekanaten und Rektoraten der deutschen medizinischen Fakultäten ein Werbe-Rundschreiben der Fa. ScienceConsult – Evaluation GmbH, dem ein Sonderdruck einer Publikation in der Zeitschrift Wissensmanagement [2] beigelegt war, eingetroffen. Soweit dieses einzig als Werbesendung zu verstehen ist, muss hier nicht darauf eingegangen werden.

Hier soll der Artikel aber als wissenschaftliche Äußerung betrachtet und kommentiert werden: Positiv zu vermerken ist, dass Frau Gieselmann, die im Begleitschreiben noch einmal auf ihre langjährige Tätigkeit bei der Hochschulrektorenkonferenz und im BMFT verweist, sich klar zur Evaluation von Studium und Lehre bekennt. Dies kann das Klima an den medizinischen Fakultäten durchaus positiv beeinflussen. Die Evaluation von Studium und Lehre ist aber nicht nur eine Frage für das Hochschulmarketing, sondern hat seine Wurzeln in der Wissenschaftsforschung im Allgemeinen und der Didaktik des jeweiligen Faches im Besonderen [9]. Unter dem momentan stetig zunehmenden Druck der Regierungen auf die medizinischen Ausbildungsstätten, Evaluationen in größerem Umfang und unter anderen Voraussetzungen als bisher zu betreiben, werden deshalb einige inhaltliche Bemerkungen zum vorgestellten Verfahren notwendig. In Nordrhein-Westfalen etwa werden durch das neue Hochschulgesetz (HG NRW, in Kraft getreten am 1.4.2000) mit dem §6 unter der Überschrift „Evaluationen“ deutliche Anforderungen an die regelmäßige Qualitätsbeurteilung in den Fakultäten des Landes formuliert. So sind die Studierenden „(...) zu ihrer Einschätzung der Lehrveranstaltungen und Studiengänge (...)“ (§6 Abs. 1 Satz 2 HG) zu befragen und „(...) hochschulauswärtige Sachverständige (...)“ (§6 Abs. 1 Satz 3) an der Bewertung zu beteiligen. Das als Kabinettsentwurf vorliegende Niedersächsische Hochschulgesetz (NHG) sieht neben den hochschulinternen Lehrveranstaltungsevaluationen ein noch viel umfassenderes, mehrstufiges Evaluationsverfahren unter Einbeziehung einer externen, vom Land zu bestimmenden Evaluationsagentur vor, die auch für die regelmäßige Erneuerung der Akkreditierung von Studiengängen zuständig ist (ZEvA). Dieses in nahezu allen Studiengängen (Ausnahmen: Human-, Tier- und Zahnmedizin) in Niedersachsen schon etablierte Evaluationsmodell sieht eine Begehung aller einschlägigen Fachbereiche des Landes durch eine Expertengruppe vor. Dieses vom NHG präferierte Evaluationsmodell ist deckungsgleich mit den Vorschlägen der Hochschulrektorenkonferenz. Allgemein wird den Hochschulleitungen in diesem Zusammenhang vom NHG ein umfangreiches

Befragungsrecht mit Auskunftspflicht seitens aller Hochschulmitglieder zugestanden.

Die unter diesen Vorzeichen den Fakultäten abverlangten Lehrveranstaltungsevaluationen werden durch das präsentierte Arrangement der Fa. ScienceConsult augenscheinlich formal erfüllt. Erst bei näherer Betrachtung stellen sich methodische Fragen an das Verfahren, die durch die Publikation nicht bzw. nicht klar beantwortet werden und die hier beschrieben werden sollen, da ihre wissenschaftlich befriedigende Beantwortung im Interesse jeder medizinischen Fakultät liegen sollte, die sich die Dienstleistungen von ScienceConsult einkauft.

Geht man von der Publikation von Gieselmann (2000, [2]) aus, so ist grundsätzlich anzumerken, dass sich diese Publikation primär mit der allgemeinen Darstellung des Evaluationsverfahrens begnügt und dabei verständlicherweise nicht auf alle Details eingehen kann. Grundsätzlich aber fehlt eine Darstellung der Verfahrensentwicklung und Angaben zur Testgüte, obwohl dies eine Grundvoraussetzung für ein wissenschaftliches Verfahren sein sollte [5]. Eine Auseinandersetzung mit den Genauigkeitsstandards des Joint Committees [10] sucht man vergebens. Von daher ist es konsequent, wenn auf eine Diskussion der Problematik von Onlinebefragungen, die den methodischen Dreh- und Angelpunkt dieses Verfahrens darstellen, ebenfalls verzichtet wird [1].

Bezüglich der dargestellten Items in Abb. 1 (Seite 41) und hier insbesondere die Veranstaltungsbewertung durch

- „b) Vermittlung von neuen Erkenntnissen der Wissenschaft“,
- „c) ausgewogenes Verhältnis von Basis- und Spezialwissen“ oder
- „d) Vermittlung von Themen des Gegenstandskataloges“

bleiben einige Fragen offen: So lassen sich im Zusammenhang mit der Itemskalierung (negativ – positiv), polyvalente Attributierungen durch den Beantwortenden denken, die als Konstruktäquivalent zu der verwendeten Skala gelten können. Im Falle des Items b) sind z.B. die Endpunkte „Vermittlung nicht geglückt bis geglückt“ aber auch „neue Erkenntnisse nicht erwähnt bis ausreichend bedacht“ möglich, um nur einige denkbare Varianten zu nennen. Darüber hinaus scheint es fraglich, wie Studierende diese Frage überhaupt beantworten können. Aus juristischer Sicht ist zumindest diskutabel, inwieweit eine Beurteilung der Inhalte einer Veranstaltung von zentraler Stelle mit der Freiheit der Lehre zu vereinbaren ist.

Es mutet merkwürdig an, dass im Text von bis zu 10-stufigen Skalen die Rede ist, alle Beispielfragen aber nur drei Antwortstufen aufweisen. Auf den Internetseiten von ScienceConsult (www.scienceconsult.de) wird dem Leser entweder eine grafische Skalierung der Fragen mit fünf optischen Hervorhebungen oder eine 6-stufige Skalierung angeboten. Dies alles spricht für die Flexibilität des Konzeptes, aber gegen die Stringenz der Testkonstruktion, zumal sie nicht mit allen Browsern auf allen Betriebssystemen einwandfrei funktioniert.

Die Gesamtbewertungen der Fächer (s. Abb. 2, Seite 4), über deren Entstehung der Leser nur spekulieren kann, werden dann einfach gegenübergestellt, so dass sich, den methodischen Einlassungen von Kromrey [5] folgend, hier ein Konvolut aus Einschätzungen zum Fachgebiet, persönlichen Vorlieben, Teilen einer didaktischen Qualität der Veranstaltung und der

aus unserer Sicht nicht statthaften Bewertung von Inhalten durch die Studierenden widerspiegelt. Damit erweist Gieselmann der Lehrveranstaltungsevaluation aber keinen guten Dienst. Wenn Befürworter der Lehrveranstaltungsevaluation wie Gieselmann die größten Vorwürfe von Kritikern wie Kromrey durch ihre Evaluationspraxis bestätigen, so ist dies bedauerlich. Dies gilt umso mehr, als gleichzeitig die nationale und internationale Forschung zur Evaluation von Lehrveranstaltungen die Verwendung eindimensionaler Veranstaltungsbeurteilungen als inadäquat nachgewiesen hat [6,7,9]. Und wenn sich, wie auf Seite 5 berichtet, tatsächlich ein signifikanter Zusammenhang zwischen den Detailanalysen und der Gesamtbewertung ergibt, kann dies auch ein Hinweis auf die Differenziertheit der Detailanalyse sein. Hierfür wäre es sicher hilfreicher gewesen, die Konstruktion und Nachbearbeitung des Fragebogens vorzustellen (ob validiert, reliabel, faktorenanalytisch entwickelt oder aus bestehenden Modellen zusammengefügt). Zu bemängeln ist auch, dass Gieselmann [2] zu ihrer Datenbasis lediglich die Angabe macht, dass ihr Verfahren seit dem Wintersemester 1999/2000 an der Medizinischen Fakultät in Heidelberg eingesetzt wird. Wie groß ist die Zahl der befragten Studierenden, der evaluierten Lehrveranstaltungen und der betroffenen Dozenten? Hierzu finden sich keine Angaben.

Ein Vorteil des Evaluationsbogens von Gieselmann ist, dass alle Befragungsteilnehmer ihre Matrikelnummer und ihr Abschneiden in den letzten staatlichen Prüfungen angeben müssen. Dies ermöglicht umfangreiche Auswertungsstrategien, zumal zumindest die Demoversion im Internet die Beantwortung jeder Einzelfrage verlangt. Es mag sein, dass für privatwirtschaftliche Unternehmen, die im Auftrag einer Hochschule tätig werden, das Bundesdatenschutzgesetz nicht gilt. Die nur durch Gesetz einschränkbare Freiheit des Einzelnen, Fragebogen nicht bzw. unvollständig ausfüllen zu dürfen oder auch eine geschützte Datenübertragung angeboten zu bekommen, wird hier aber allen Beteuerungen im Artikel zum Trotz nicht beachtet.

Natürlich beruhigt es den Befürworter von Lehrveranstaltungsevaluationen, wenn sich nach den Erfahrungen der Firma ScienceConsult feststellen lässt, „dass die Studierenden nie beliebig, sondern differenziert und kohärent urteilen“, also die gesetzlich vorgeschriebene Lehrveranstaltungsevaluation sinnvoll ist (Seite 5). Diese Erkenntnis ist nicht neu. So kommt Marsh 1984 auch zu diesem Ergebnis, wobei er bereits damals auf eine um den Faktor 10 größere Stichprobe von Studierenden zurückgreifen konnte, die mit einem nach wissenschaftlichen Kriterien konstruierten Fragebogen befragt wurde. Die Validität studentischer Lehrveranstaltungsbeurteilungen ist in der wissenschaftlichen Diskussion keineswegs mehr umstritten, Gieselmann bräuchte diese „Erfahrung“ also nicht zu betonen. Umstritten hingegen ist, wie stark die Validität studentischer Evaluation von Hochschulmitgliedern z.B. durch nachsichtige Benotung gefährdet wird (vgl. z.B. [3,7]).

Wenn, wie auf Seite 5 beschrieben, aus methodischen Gründen keine Veranstaltungen ohne Teilnahmepflicht in die summative Evaluation aufgenommen werden können, stellt sich uns die Frage nach der Verwendbarkeit der Ergebnisse (insbesondere im Rahmen des Konzepts der auf Seite 6 vorgestellten Meilensteinevaluation), da im Regelfall das Curriculum der Semester 1 – 10 in Deutschland zur Hälfte aus eben solchen Ver-

anstaltungen ohne Teilnahmepflicht (vornehmlich Vorlesungen) besteht und damit ein ganz wesentlicher Anteil des Unterrichts somit nicht abgebildet ist. Eine Meilensteinevaluation bietet in der Tat eine Reihe von Vorteilen gegenüber der jeweils separaten Bewertung einzelner Lehrveranstaltungen. Dem stehen aber auch eine Reihe von Nachteilen gegenüber [11], auf die man in einer wissenschaftlichen Arbeit zumindest hinweisen sollte. Zumindest die naiv-psychologischen Aussagen zur stärkeren Berücksichtigung nur jener Einflüsse auf die Bewertung, die „tatsächlich nachhaltig auf die Lehrqualität eingewirkt haben“ ([2], S. 6), und die Behauptung, dass die Einschätzung eines Qualitätsmerkmals umso stärker variiert je nachrangiger es ist, hätten eines minimalen empirischen Belegs bedurft.

Nebenbei wäre es interessant zu erfahren, ob die Verknüpfung zwischen den Prüfungs- und Evaluationsergebnissen und der berichtete statistische Zusammenhang zwischen beiden Variablen nur deshalb in „(...) einer der Evaluationsrunden für die klinisch-theoretisch ausgerichteten Fächer (...)“ ([2], S. 7, Hervorhebung durch die Autoren) nachzuweisen war, weil eben nur ein singulärer Messzeitpunkt vorliegt oder weil die anderen Messzeitpunkte nicht signifikante Ergebnisse erbrachten. Auch hier bleiben die näheren Umstände des Verfahrens im Dunklen.

Dieses aus unserer Sicht insgesamt methodisch problematische Evaluationsverfahren mündet abschließend in dem Kommentar, dass, würde das Verfahren an mehreren Standorten in gleicher Weise durchgeführt, ein Qualitätsvergleich auf Fächer- und sogar Standortebene möglich wäre ([2], S. 7). Dies lässt aufhorchen, verspricht die Firma ScienceConsult doch damit den Kindheitstraum der neueren Bildungspolitik zu erfüllen und eine scheinbar versierte Grundlage für Steuerungsverfahren (qualitätsorientierte Mittelvergabe) zu produzieren. Dazu und abschließend zwei Bemerkungen: Ein Ergebnis ist sicher nicht deshalb schon vergleichbar, weil die angewandte Methode dieselbe ist. Hier sei auf die Publikation des IMPP hingewiesen [8], die schon den interfakultären Vergleich der Fächer anhand ihrer eigenen Examensergebnisse (nach immerhin über 25 Jahren Erfahrung mit ihrem Messinstrument) aufgrund der dokumentierten Populationsunterschiede für unzulässig hält (zur Problematik von Rankings vgl. in diesem Zusammenhang auch Süllwold 1997 [12]). Zweitens, und das erscheint uns noch wichtiger, haben die politischen Entscheidungsträger zumindest in NRW nicht davon abgesehen, die Ergebnisse des IMPP (ein mancher mag meinen wider besseren Wissens) steuerungspolitisch zu verwenden. Dabei war sicherlich die Praktikabilität (Einfachheit der Datenaquisition und deren Verarbeitung) der ausschlaggebende Grund. Dies insgesamt ist nicht grundsätzlich verwerflich, zumal das IMPP für die von ihm entwickelten Prüfungsfragen regelmäßig die Testgütekriterien bestimmt. Im Hinblick auf die Entwicklung von adäquaten Verfahren zur Messung der Qualität der Lehre an den Fakultäten sind die Implikationen der Dichotomie „einfach = (gerecht) durchführbar“ versus „komplex = (unnötig) kompliziert“ aber eher hinderlich. Oder haben sich durch das verwendete Verfahren zur qualitätsorientierten Mittelvergabe in NRW deren Fakultäten in den Examensnoten nachhaltig verändert?

Die Förderung von Studium und Lehre ist auch eine Aufgabe für das Hochschulmarketing. Dazu kann die Evaluation von

Studium und Lehre einen wesentlichen Beitrag leisten. Eine rege, die Probleme beim Namen nennende wissenschaftliche Diskussion der verschiedenen Evaluationsstrategien ist hierfür eine unabdingbare Voraussetzung. Der Artikel von Gieselmann trägt zu dieser Diskussion leider nicht viel bei.

¹ Soweit nicht anders gekennzeichnet, beziehen sich die Seitenangaben auf den Sonderdruck: Gieselmann, 2000

Literatur

- ¹ Batinic B, Werner A, Gräf L, Bandilla W (Hrsg). Online-Research. Göttingen: Hogrefe, 1999
- ² Gieselmann G. Online-Befragung von Studierenden – Zuverlässiges Instrument für die Evaluation der Lehre. Wissenschaftsmanagement 2000; (Sonderdruck)
- ³ Greenwald AG, Gillmore GM. No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology* 1997; 89: 743 – 751
- ⁴ Kromrey H. Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In: Mohler P (Hrsg): *Universität und Lehre: ihre Herausforderung an die empirische Sozialforschung*. Münster, New York: Waxmann, 1994: 91 – 114
- ⁵ Lienert GA. Testaufbau und Testanalyse. München: Psychologie Verlags Union, 1989 (4. neu ausgest. Aufl.)
- ⁶ Marsh HW. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 1984; 76: 707 – 754
- ⁷ Marsh HW, Roche IA. Making students' evaluations of teaching effectiveness effective. *American Psychologist* 1997; 52: 1187 – 1197
- ⁸ Neumann D, Voigtmann K. Hochschulen im Vergleich – Eine Analyse der schriftlichen Ergebnisse der ärztlichen Vorprüfung 1984 bis 1991. *Dtsch Ärztebl* 1995; 92: A-256 – A-261
- ⁹ Rindermann H. Die studentische Beurteilung von Lehrveranstaltungen – Forschungsstand und Implikationen. In: Spiel C (Hrsg): *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck*. Münster: Waxmann, (in Druck)
- ¹⁰ Sanders JR (Hrsg). *Handbuch der Evaluationsstandards – Die Standards des „Joint Committee on Standards for Educational Evaluation“*. Opladen: Leske & Buderich, 1999
- ¹¹ Stein M. Vorlesungs-Psychogramme. Eine Methode zur Evaluation großer Vorlesungen. *Forschung & Lehre* 1996; 3: 256 – 257
- ¹² Süllwold F. Ranking ist oft ein Synonym für Unsinn. Hinweis auf ein messmethodisches Grundproblem. *Forschung & Lehre* 1997; 4: 578 – 579

Dr. h. c. (RUS) Christoph Stosch

Studiendekanat der Medizinischen Fakultät
der Universität zu Köln
Joseph-Stelzmann-Straße 9
50924 Köln

E-mail: christoph.stosch@medizin.uni-koeln.de