

Fallstricke der Lehrevaluation

Ein Plädoyer für einen sachgemäßen Umgang mit studentischen Lehrveranstaltungsbewertungen

| **TOBIAS WOLBRING** | **Lehrveranstaltungsevaluationen messen nicht nur die Lehrqualität, sondern werden auch vom Veranstaltungskontext, Veranstaltungsverlauf und der Zusammensetzung der Teilnehmerschaft beeinflusst. Dies hat Implikationen für die Nutzung studentischer Bewertungen der Lehre.**

Die Bewertung der Qualität universitärer Lehre durch Studierende mittels standardisierter Fragebögen hat in Deutschland flächendeckende Verbreitung gefunden und ist auch in Gesetzen und hochschulpolitischen Empfehlungen fest verankert. Die Frage ist also nicht mehr, ob die Lehre durch die Studierenden beurteilt werden soll, sondern wie mit den Ergebnissen zu verfahren ist. Vermehrt wird in jüngster Zeit die Nutzung des Instruments zur Setzung von Anreizen für bessere Lehre, als Entscheidungsgrundlage bei Berufungsverfahren und auch zur Erstellung von Hochschulrankings diskutiert und teilweise bereits praktiziert. Eine entsprechende Verwendung studentischer Lehrbewertungen setzt aber voraus, dass mit dem Instrument tatsächlich nur die individuelle Lehrleistung (und nicht auch etwas Anderes) gemessen wird und dass die Ergebnisse über Veranstaltungen hinweg vergleichbar sind. Im Folgenden sollen Möglichkeiten und Grenzen der Nutzung von Lehrveranstaltungsbewertungen aufgezeigt werden. Der Beitrag speist sich dabei aus empirischen Analysen auf Grundlage studentischer Lehrveranstaltungsevaluationen an der Sozialwissenschaftlichen Fakultät der LMU München.

Entscheidend ist, was die Studierenden in den Hörsaal mitbringen

Ein Gutteil der Bewertung der Lehre liegt nicht in der Hand der Lehrenden selbst, sondern hängt von der Komposition der Teilnehmerschaft ab. Systematische Unterschiede im studentischen Urteilsverhalten lassen sich dabei u.a. an folgenden Befunden festmachen: (1) Die Note der Hochschulzugangsberechtigung der Studierenden ist ein guter Prädiktor für das spätere studentische Urteil. (2) Personen mit einem mathematisch-naturwissenschaftlichen Schwerpunkt in der Schule (z.B. Informatik, Mathematik, Physik) beurteilen formal anspruchsvollere Lehre wohlwollender. (3) Studentische Urteile zur Qualität von Einführungsvorlesungen korrelieren statistisch überzufällig mit den Bewertungen von Lehrveranstaltungen durch dieselbe Person in höheren Semestern. (4) Wahlveranstaltungen werden signifikant besser evaluiert als Pflichtveranstaltungen. (5) Das Vorinteresse am Veranstaltungsthema ist – insbesondere bei Veranstaltungen, in denen ein Leistungsnachweis erworben werden muss – für eine positive Bewertung entscheidend. Man könnte nun einwenden, dass diese Befunde unproblematisch sind, da es gerade ein Merkmal guter Lehre ist, Lernräu-

me zu schaffen, welche dem Kenntnisstand und den Interessenlagen der Studierendenschaft Rechnung tragen. Dies trifft sicherlich zu. Gerade in Zeiten einer Flexibilisierung des Hochschulzuges ist jedoch die Hörschaft häufig sehr heterogen, sodass es für Lehrende kaum möglich ist, allen an sie gerichteten Ansprüchen gleichzeitig gerecht zu werden. Schlechte Lehrbewertungen können folglich auf strukturelle Probleme hindeuten und lassen nicht zwingend auf eine mangelhafte Lehrleistung der evaluierten Personen schließen.

„Wie Du mir, so ich Dir!“

Neben diesen Unterschieden studentischer Erwartungshaltungen wird in der Literatur die Gefahr von Verzerrungen aufgrund lehrunabhängiger Einflüsse diskutiert. In den eigenen empirischen Analysen konnte etwa ein Effekt der physischen Attraktivität der Dozierenden auf die Lehrbewertung dokumentiert werden. Das Auge hört bei der Lehre mit. Die Effekte sind im Durchschnitt relativ schwach, können jedoch in Extremfällen Unterschiede in der Bewertung von 0,3 bis 0,9 Punkten auf einer fünfstufigen Notenskala bewirken. Während dieses Messproblem nur in Einzelfällen schwerwiegende Konsequenzen hat, ist ein zweites Resultat von deutlich allgemeiner Relevanz: Falls vor der Evaluation Noten vergeben werden, hat das individuelle Abschneiden in Prüfungen *unabhängig von der Lehrleistung* einen sehr starken Einfluss auf die studentische Lehrbewertung (vgl. Abbildung 1). Gemäß dem Motto „Wie du mir, so ich Dir!“ werden Dozierende von den Studierenden für die Testgestaltung und Notengebung in der Lehrevaluation belohnt oder bestraft. Selbst wenn die Lehre vor entsprechenden Prüfungen evaluiert wird, sind solche Reziprozitätseffekte aufgrund



A U T O R

Dr. **Tobias Wolbring** ist Postdoctoral Fellow an der ETH Zürich. Seine Forschungsschwerpunkte sind Methoden der empirischen Sozialforschung, Evaluation von Forschung und Lehre sowie Wirtschaftssoziologie.

von Antizipation, Gerüchten und Erwartungsbildung nicht auszuschließen.

Wer bewertet hier überhaupt?

Hinzu kommt, dass keineswegs die gesamte ursprüngliche Hörerschaft in die Evaluation einbezogen wird. Problematisch ist dies deshalb, da sich die befragten Studierenden systematisch von denjenigen Personen unterscheiden, die nicht an der Befragung teilnehmen wollen oder können. Evaluiert man online, so besteht zwar prinzipiell die Möglichkeit, auch diejenigen Studierenden zu kontaktieren, welche den Kurs nicht regelmäßig besuchen oder abgebrochen haben. Online-Evaluationen haben aber fast immer geringe Rücklaufquoten und liefern verzerrte Ergebnisse, da die Zufriedenheit mit der Lehre und die Teilnahmebereitschaft miteinander zusammenhängen. Führt man die Befragung dagegen in Papierform während der Veranstaltung durch, so wird zwar das Stimmungsbild der zum Evaluationszeitpunkt anwesenden Hörerschaft erfasst, die Lehrbewertung abwesender Studierender, die im Schnitt unzufriedener sind (vgl. Abbildung 2), wird jedoch außen vor gelassen. Wie eigene empirische Analysen zeigen, unterscheiden sich Lehrveranstaltungsrankings, welche diese abwesenden Studierenden einbeziehen, deutlich von unadjustierten Ranglisten. Veränderungen ergeben sich dabei besonders in der Mitte der (relativ dichten) Verteilung, aber auch bezüglich der Spitzen- und Schlussgruppe, welche für die Setzung von Anreizen und Maßnahmen von besonderem Interesse sind.

Gut gemeinte Anreize und ihre unintendierten Folgen

Diese Ergebnisse implizieren freilich nicht, dass studentische Lehrveranstaltungsbewertungen uninformativ sind. Anhand von Validierungsstudien, die systematische Zusammenhänge mit anderen Maßen der Lehrqualität aufzeigen und Wirkungen in Kombination mit Weiterbildungsangeboten dokumentieren, lässt sich erkennen, dass das Einholen des studentischen Stimmungsbilds durchaus einen Mehrwert erbringt. Der Beitrag ist daher keineswegs als Plädoyer für die Abschaffung studentischer Lehrveranstaltungsbewertungen zu verstehen. Die vorgestellten Befunde legen jedoch einen auf Leitungsebene zurückhaltenden Umgang nahe. Grenzen des Wettbewerbs ergeben sich insbesondere aus der Möglichkeit unintendierter Folgewirkungen von

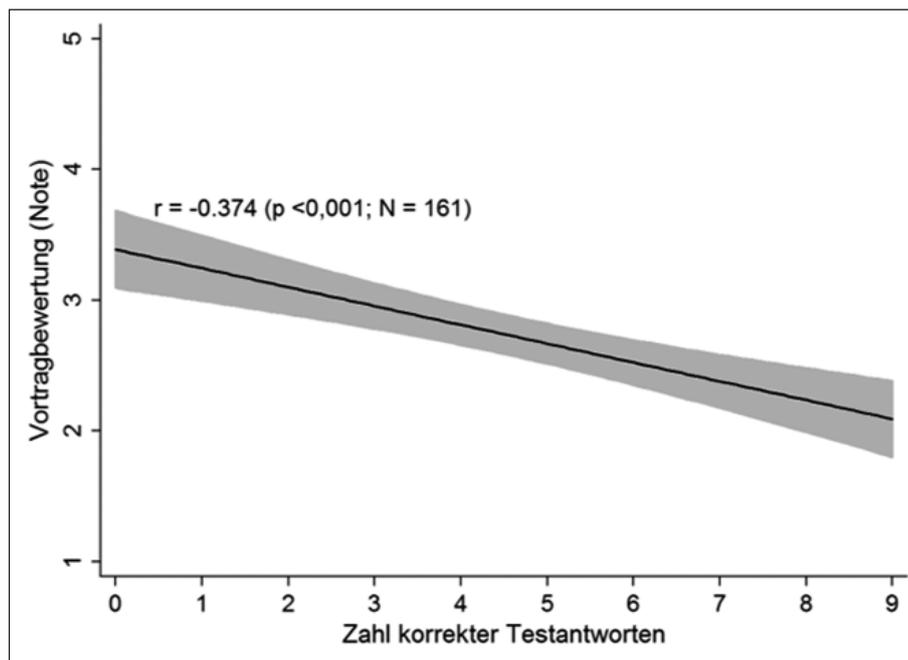


Abb. 1: Lehrbewertung nach individuellem Testerfolg in einem Experiment

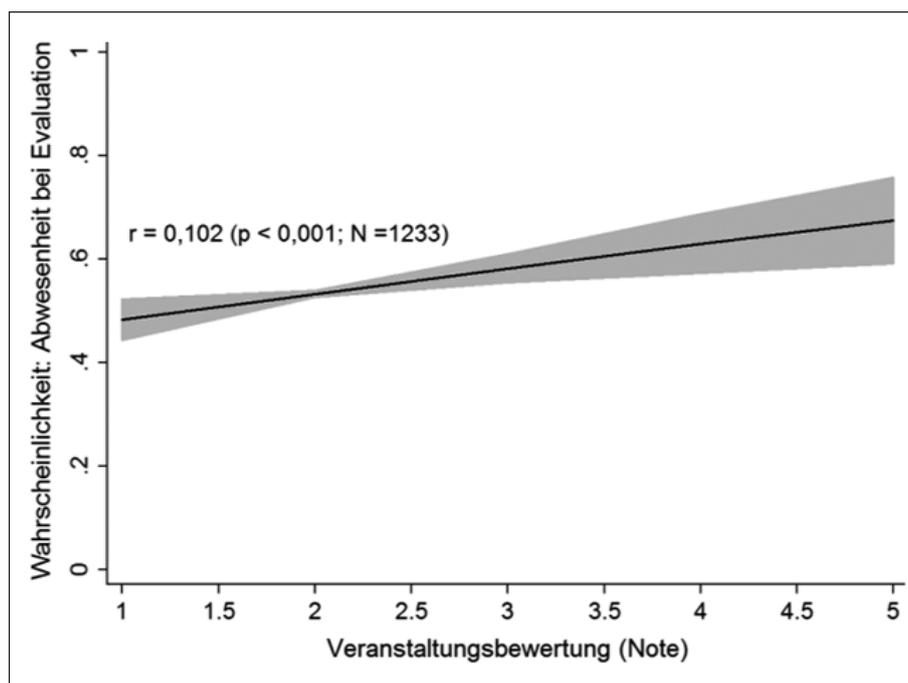


Abb. 2: Effekt der Zufriedenheit ($t = 3./4.$ Woche) auf den späteren Veranstaltungsbesuch ($t = 9./10.$ Woche)

Verteilungsmechanismen, die auf verzerrten und manipulierbaren Messungen basieren. Unerwünschte Konsequenzen der Setzung von Anreizen können dabei von der Verdrängung und Vernichtung intrinsischer Motivation über verschiedene Ausweichstrategien bis hin zur systematischen Manipulation z.B. durch Senkung des Anspruchsniveaus und Noteninflation reichen. Im Gegensatz zu einer Nutzung als Steue-

rungsinstrument liegt es daher nahe, Lehrrevaluationen als ein formatives Instrument zu begreifen, das bereits früher als bisher üblich im Semester eingesetzt wird und das – ergänzt durch zeitnahes Feedback und Fortbildungsmöglichkeiten – den Lehrprozess begleitet.

Von dem Autor ist gerade das Buch „Fallstricke der Lehrrevaluation“ im Campus Verlag erschienen.