

Große Vorsicht geboten

Was bewerten Studierende, wenn sie Lehrveranstaltungen evaluieren?

| WOLFGANG STROEBE | **Studentische Lehrevaluationen gelten als solides Instrument zur Bewertung von Professoren, Dozenten und deren Lehrveranstaltungen. Rankings von Universitäten werden durch Evaluation durch Studierende beeinflusst. Neue Studien belegen jedoch die Gefahren, wenn diese Informationen als maßgebliche Grundlage für die Beurteilung akademischer Lehre benutzt werden.**

Die ersten Fragebögen für die Evaluation von Lehrveranstaltungen wurden in den zwanziger Jahren des letzten Jahrhunderts von den amerikanischen Psychologen Hermann Remmers und Edwin Guthrie mit dem alleinigen Ziel entwickelt, Lehrenden Rückmeldungen zu geben, wie ihre Lehrveranstaltungen von Studierenden wahrgenommen werden. Guthrie warnte, dass es ein ernsthafter Missbrauch dieser Informationen sei, sie als alleinige Grundlage für die Beurteilung der Eignung eines Dozierenden zur akademischen Lehre zu benutzen. Dennoch dauerte es nur wenige Jahrzehnte, bis Institutsdirektoren und Dekane in den USA entdeckten, dass ihnen diese Bewertungen scheinbar objektive Evidenz für die Lehrqualität von Dozierenden lieferten. Inzwischen sind studentische Lehrevaluationen in den USA an nahezu allen Universitäten eine wichtige Informationsquelle für Entscheidungen über Festanstellung und Gehaltserhöhungen,

ein Trend, der sich auch in Deutschland durchzusetzen scheint.

Neuerdings gewinnt die Evaluation der Lehre durch Studierende sogar Einfluss auf das Ranking von Universitäten. So wird die Zufriedenheit der Studierenden mit der Lehre an ihrer Universität bei dem in England geplanten „Teaching Excellence Framework“ die Bewertung von Universitäten sowie die

»Die Beurteilung kommerzieller Dienstleistungen erfolgt unilateral, während zwischen Studierenden und Lehrenden eine bilaterale Beziehung besteht.«

Höhe der erlaubten Studiengebühren mitbestimmen. Damit folgt die englische Regierung dem Beispiel des Wirtschafts magazins Forbes, das die Zufriedenheit von Studierenden mit ihrer Universität als einen der Beurteilungsfaktoren bei ihrem Ranking amerikanischer Universitäten berücksichtigt. Als ein Maß der Zufriedenheit benutzt Forbes die Beurteilungen von Lehrenden auf „RateMyProfessors.com“, einer populären Webseite, auf der Studierende anonym ihre Professoren beurteilen können. Forbes begründet dies damit, dass Kundenzufriedenheit auch von Konsumentenzeitschriften für die Beurteilungen von Produkten und Dienstleistungen benutzt würde.

Die Gleichstellung von Kundenzufriedenheit mit kommerziellen Diensten

mit der Zufriedenheit mit universitären Lehrveranstaltungen übersieht wesentliche Unterschiede zwischen diesen beiden Urteilen. Während es bei kommerziellen Dienstleistungen durchaus sinnvoll ist, dass der Preis als ein Faktor in das Zufriedenheitsurteil eingeht, ist es weniger wünschenswert, wenn der von Dozierenden geforderte Arbeitsaufwand die Beurteilung von Lehrveranstaltungen (negativ) beeinflusst. Weiterhin sind Universitäten vermutlich daran interessiert, dass Lehrveranstaltungen Kenntnis vermitteln und wissenschaftliches Denken fördern. Für viele Studierende ist hingegen der Unterhaltungswert einer Lehrveranstaltung von größerer Bedeutung, der nicht unbedingt mit dem in einer Lehrveranstaltung erzielten Kenntnisgewinn einhergehen muss. Der wohl wichtigste Unterschied ist aber vermutlich, dass die Beurteilung kommerzieller Dienstleistungen unilateral

erfolgt, während zwischen Studierenden und Lehrenden eine bilaterale Beziehung besteht. Studierende bewerten die Qualität der Lehrenden, die wiederum Leistungen der Studierenden beurteilen.

In amerikanischen Untersuchungen wird in der Regel ein positiver Zusammenhang zwischen den von Studierenden in einer Lehrveranstaltung erzielten Noten und der studentischen Bewertung der Lehrveranstaltung gefunden. Verteidiger der Evaluation von Lehrveranstaltungen werten dies als Zeichen für die Gültigkeit dieser Studierendenurteile: Studierende lernen mehr von guten Dozierenden. Gute Lehrende erzielen deshalb bessere Notendurchschnitte und werden auch besser beurteilt. Kritiker argumentieren hingegen, dass Lehrevaluation den Unterhaltungswert von

AUTOR



Wolfgang Stroebe ist Gastprofessor an der Abteilung für Sozialpsychologie der Universität Groningen, Niederlande.



Foto: mauritius-images

Lehrveranstaltungen und nicht den Wissensgewinn reflektieren. Weiterhin haben viele Studierende eine Abneigung gegenüber Lehrenden, die großen Arbeitsaufwand fordern, da dies zu Einschränkungen in ihrem Sozialleben führen kann. Da viele Lehrende sich dieses Zusammenhangs bewusst sind und die Lehrevaluationen bei vielen Universitäten Einfluss auf Personalentscheidungen haben, besteht die Versuchung, die Anforderungen in Lehrveranstaltungen zu senken, um die Lehrevaluation zu verbessern.

Empirische Untersuchungen nähren diese Zweifel an der Validität der studentischen Lehrevaluation. So fanden mehrere Studien einen positiven Zusammenhang zwischen der Einstufung eines Lehrenden als jemanden, bei dem gute Noten leicht zu erzielen sind, und der Beurteilung der Lehrveranstaltungen dieser Person. Weitere Evidenz kommt von Analysen von Beurteilungen auf der Internetwebseite „RateMyProfessors.com“. Dort beurteilen Studierende nicht nur die Qualität der Lehrveranstaltungen, sondern auch, wie einfach es ist, bei diesem Dozierenden ohne viel Arbeit gute Noten zu erzielen. Diese beiden Urteile sind hoch positiv miteinander korreliert. Je einfacher es ist, bei einem Lehrenden eine gute Note zu erzielen, umso besser wird dieser beurteilt.

Die eindeutigsten Ergebnisse stammen allerdings von Untersuchungen, die ein neues Kriterium für den Lernerfolg in Lehrveranstaltungen einführten. Statt den Wissenszuwachs von Studierenden über deren Note in der evaluierten Lehrveranstaltung zu prüfen, nutzten sie die Note in einer auf dieser Lehrver-

anstaltung aufbauenden Veranstaltung als Indikator für den Lehrerfolg. Wenn jemand zum Beispiel in einer Einführungsveranstaltung für Statistik viel gelernt hat, sollte er oder sie in Prüfungen der Veranstaltung „Statistik für Fortgeschrittene“ besser abschneiden als jemand, der in der Einführungsveranstaltung wenig gelernt hatte. In mehreren großen Untersuchungen wurde gezeigt, dass Studierende, die an Einführungsveranstaltungen teilnahmen, die gute Lehrevaluationen erzielt hatten, in den darauf aufbauenden Veranstaltungen für Fortgeschrittene schlechtere Noten bekamen als Studierende, die an weniger gut beurteilten Einführungsveranstaltungen teilgenommen hatten. Mit anderen Worten, je besser die Evaluation der Einführungsveranstaltung war, umso weniger schienen diese Studierenden gelernt zu haben.

Diese Befunde lassen sich schwer mit der Annahme vereinbaren, dass die Evaluation von Lehrveranstaltungen durch Studierende den in der Veranstaltung erzielten Wissensgewinn reflektieren und damit als Indikator für die Qualität eines Lehrenden dienen können. Einer der Gründe ist, dass Studierende den in einer Lehrveranstaltung erzielten Wissenszuwachs nicht einschätzen können, zumindest wenn man ihre Leistungen in einer auf dieser Lehrveranstaltung aufbauenden Veranstaltung als Kriterium verwendet. So wurde kein Zusammenhang zwischen studentischen Antworten auf die Frage, wie viel sie in einer Einführungsveranstaltung gelernt hatten, und ihren Leistungen in einer Folgeveranstaltung gefunden. Der Hauptgrund ist aber vermutlich, dass

gute Lehrende von ihren Studierenden viel Arbeit verlangen, was vor allem von schlechten Studierenden nicht gern gesehen wird. Da studentische Bewertungen von Lehrveranstaltungen die allgemeine Zufriedenheit mit der Veranstaltung zu erfassen scheinen und da Strategien, die Wissensgewinn fördern (wie etwa großes Literaturstudium und regelmäßige Teilnahme an der Lehrveranstaltung), die Zufriedenheit mindern, sind studentische Beurteilungen von Lehrveranstaltungen ein unzureichendes Maß der Effektivität von Lehrenden.

Einschränkend muss allerdings bemerkt werden, dass meine Schlussfolgerungen nahezu ausschließlich auf amerikanischen Untersuchungen beruhen und dass die beschriebenen Zusammenhänge hauptsächlich für weniger begabte und motivierte Studierende gelten. Obwohl die grundlegenden Prozesse – wie etwa der negative Einfluss von verlangtem Arbeitsaufwand und Lehrevaluation – vermutlich auf alle westlichen Länder zutreffen, ist es möglich, dass hohe Studiengebühren, wie sie etwa in den USA und England gängig sind, die Konsumenteneinstellung von Studierenden erhöhen und diese Effekte verstärken. Da die einzige europäische Untersuchung an einer Privatuniversität durchgeführt wurde, können diese Befunde diese Bedenken nicht ausräumen. Ich kann deshalb nur den Schlussfolgerungen zustimmen, die in einem Bericht der Universität Stuttgart im Jahre 2001 zu „Lehrveranstaltungsevaluation durch Studierendenbefragung an der Universität Stuttgart“ gezogen wurden. „In Anbetracht der beschränkten Aussagekraft von Lehrveranstaltungsevaluationen durch Studierendenbefragungen ist große Vorsicht bei der Verwendung der Evaluationsergebnisse geboten. Das primäre Ziel soll und kann nur die individuelle Verbesserung der evaluierten Lehrveranstaltung durch den jeweiligen Dozenten auf Basis des Feedbacks der Studierenden sein. Weitergehende Schlussfolgerungen aus den Ergebnissen, z.B. in Bezug auf die Eignung des Dozenten als akademischer Lehrer, sind nur schwer zu begründen.“ Dies ist genau der Zweck, für den die Fragebögen für Lehrevaluationen von Remmers und Guthrie ursprünglich entwickelt wurden.

Ausführlichere Angaben zu den Ergebnissen sind zu finden in: Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. Perspectives on Psychological Science, 11. 800– 816.