



Stellungnahme der Deutschen Gesellschaft für Soziologie zum Umgang mit studentischen Lehrveranstaltungsevaluationen¹

August 2020

Die Bewertung der Qualität universitärer Lehre durch Studierende mittels standardisierter Fragebögen – kurz studentische Lehrveranstaltungsevaluation (LVE) – hat im deutschsprachigen Raum flächendeckende Verbreitung gefunden und ist auch in Hochschulgesetzen fest verankert. Die Zielsetzungen, die sich hinter dem Einsatz des Instruments verbergen können, sind vielfältig, wobei es erhebliche Variation in den Evaluationspraktiken und -kulturen zwischen Hochschulen und Fächern gibt: Studentische LVE bieten Studierenden eine Möglichkeit, den Lehrenden die eigenen Erfahrungen und Sichtweisen auf die Lehre in anonymisierter Form zurück zu spiegeln. LVE können den Lehrenden Informationen liefern, um die Erwartungen der Hörerschaft besser zu verstehen und die Qualität ihrer Lehre zu steigern. Und LVE können zur Setzung von Anreizen für bessere Lehre (z.B. Lehrpreise), als Entscheidungsgrundlage bei Berufungs- und Entfristungsverfahren (z.B. kritische Schwellen als Bedingung für *Tenure*) sowie zur Kontrolle der Lehrleistung (z.B. hochschulinterne Publikation von LVE-basierten Lehrenden-Rankings) eingesetzt werden.

Je nach dem angestrebten Verwendungszweck unterscheiden sich die Anforderungen – z.B. im Hinblick auf die Validität, Präzision und Vergleichbarkeit der Messungen – an das Instrument. Daher sind studentische LVE für diese unterschiedlichen Zielsetzungen nicht gleichermaßen gut geeignet. Dies kann bei unangemessener Nutzung von LVE im schlimmsten Falle dazu führen, dass Fehlanreize für Lehrende gesetzt und an Hochschulen fehlgeleitete Entscheidungen auf Basis dieser vermeintlich objektiven Grundlage getroffen werden. Aus diesem Grund sieht sich die DGS zu einer Stellungnahme zu den Möglichkeiten und Grenzen einer Nutzung des Instruments veranlasst (siehe auch ASA 2019).

Hierfür erscheint zunächst eine Einordnung in die Forschung notwendig.² Grundsätzlich lassen sich zwei große Forschungszeige (mit vielen weiteren Verästelungen) unterscheiden:

¹ Der Vorstand der DGS dankt den Mitgliedern, insbesondere Tobias Wolbring, der zur Erarbeitung dieser Stellungnahme eingesetzten Kommission: Rolf Becker, Monika Jungbauer-Gans, Michaela Pfadenhauer, Larissa Schindler, Tobias Wolbring (Kommissionsleitung).

Auf der einen Seite gibt es eine Vielzahl an Studien, welche die Zusammenhänge von studentischen LVE mit anderen Maßen der Lehrqualität untersuchen (z.B. Marsh 2007; Rindermann 2001; Spooren, Brockx, Mortelmans 2013). Hierbei zeigt sich, dass das Instrument mittlere bis starke Korrelationen mit naheliegenden Außenkriterien, wie der Einschätzungen durch Peers, aufweist. Auch lassen sich durch hochschuldidaktische Weiterbildungsmaßnahmen Verbesserungen der LVE im Zeitverlauf erzielen (Dresel, Rindermann 2011; Penny, Coe 2004). Dies deutet an, dass Aspekte der Lehrqualität zu einem nicht unerheblichen Teil in die studentischen Bewertungen eingehen, dem Instrument also eine gewisse Validität nicht abzusprechen ist.

Auf der anderen Seite lässt die Stärke der empirisch dokumentierten Zusammenhänge aber durchaus Raum für unerwünschte sachfremde Einflüsse (z.B. Kromrey 1994a; Stark, Freisztat 2014; Wolbring 2013). So deuten zahlreiche Forschungsarbeiten *erstens* an, dass es unter bestimmten Bedingungen Einflussfaktoren auf die studentische Bewertung geben könnte, die unabhängig von der Qualität der Lehre wirken. So wurden die studentischen Bewertungen der Lehre in verschiedenen Studien mit dem Geschlecht (z.B. Boring 2017; Rivera, Tilcsik 2019) und der physischen Attraktivität der Dozierenden (z.B. Hamermesh, Parker 2005; Klein, Rosar 2006) in Verbindung gebracht. Auch weisen Studien für andere Länder auf mögliche Zusammenhänge der studentischen Bewertungen mit der Hautfarbe und ethnischen Herkunft der Lehrenden hin (z.B. Chávez, Mitchell 2020; Chisadza, Nicholls, Yitbarek 2019). *Zweitens* gehen Rahmenbedingungen der Lehre, wie das studentische Vorinteresse, der Pflicht- bzw. Wahlcharakter einer Veranstaltung und auch die räumliche Situation (z.B. Esser 1997; Rindermann 2016), die nicht oder nicht gänzlich in der Hand der Lehrenden liegen, in die Bewertung mit ein. Schließlich kann es aufgrund von Reziprozitäten bei der gegenseitigen Bewertung der Lehr- und Studienleistung durch Lehrende und Studierende *drittens* zu Fehlanreizen bei der Gestaltung von Lehre und Prüfungen kommen. So eröffnen sich Möglichkeiten und Anreize zur Manipulation von LVE durch eine Reduktion des Workloads oder eine Inflation der Noten (z.B. Bar, Kadiyali, Zussmann 2009; Johnson 2003; Weinberg, Hashimoto, Fleisher 2009).

Auch wenn diese Effekte je nach sozialem Kontext variieren dürften, zeigen diese Befunde, dass solche Faktoren LVE potenziell verzerren und im Einzelfall durchaus starke Effekte haben können. Daneben können Methodeneffekte auftreten, die mit der Fragebogengestaltung, dem Befragungsmodus und dem Rücklauf zusammenhängen (z.B. Dresel, Tinsner 2008; Kromrey 1994b; Treischl, Wolbring 2017). Dies erschwert ebenfalls die Vergleichbarkeit von LVE über verschiedene Veranstaltungsformate, Themen und lokale Hochschulkontexte hinweg. All dies entkräftet freilich nicht die zuvor berichteten Befunde zur Validität studentischer LVE, zeigt aber, dass LVE die Zufriedenheit der Studierenden mit der Lehre erfassen, was durchaus wünschenswert ist. LVE-Ergebnisse sind jedoch trotz ihrer häufigen Nutzung als Grundlage für

² Die Forschung zur studentischen LVE ist aufgrund der an Hochschulen standardmäßig anfallenden LVE-Daten und der direkten Betroffenheit der Forschenden durch das Instrument überaus umfangreich. Zu beachten ist dabei, dass große Unterschiede in der methodischen Güte einzelner Studien bestehen. Einzelbefunde sollten daher kritisch geprüft und mit dem dazugehörigen Forschungsstand abgeglichen werden.

Karriereentscheidungen nicht mit einer präzisen Messung der Lehrleistung von Einzelpersonen gleichzusetzen. Daher sollte das Instrument mit Bedacht und Augenmaß unter Anerkennung dieser Fallstricke verwendet werden. Die DGS leitet aus dem skizzierten Forschungsstand folgende Empfehlungen ab:

- LVE sollten zum Zweck einer formativen Evaluation, d.h. einer den Lehrprozess begleitenden Bewertung, eingesetzt werden. Hierfür ist eine Evaluationskultur zu entwickeln, bei der die Feedback- und Entwicklungsfunktion im Vordergrund steht. Dabei sollten LVE nicht als Ausdruck der Lehrleistung einzelner Lehrpersonen, sondern als das Ergebnis des Zusammenspiels aus Lehrenden, Studierenden und Rahmenbedingungen begriffen werden. Vor diesem Hintergrund liegt eine stärkere Nutzung von LVE für die systematische Weiterentwicklung von Modulen und Studiengängen im Rahmen des Studiengangs- und Qualitätsmanagements nahe.
- Für eine Begleitung des Lehrprozesses durch LVE im Sinne einer formativen Evaluation sollten diese früher als bisher üblich im Semester durchgeführt werden, wobei die Fragebögen entsprechend dieser veränderten Zielsetzung anzupassen sind. So können die Evaluationsergebnisse nicht nur im laufenden Semester mit den Studierenden besprochen, sondern auch dafür genutzt werden, um eine ggf. notwendige Nachjustierung in der Lehre vorzunehmen. Dies erhöht auch die Motivation für Studierende, sich an LVE zu beteiligen und Lehrenden ein konstruktives Feedback zu geben.
- Der in LVE eingesetzte standardisierte Fragenkanon bildet oft die Spezifika bestimmter Lehrveranstaltungen, z.B. deren didaktische Konzepte und konkrete Ausgestaltung, nur unzureichend ab. Um einen möglichst hohen Informationsgehalt von LVE zu gewährleisten, sollte Lehrenden die Möglichkeit gegeben werden, das Standardinstrument durch eigene Fragen zu ergänzen.
- Bei der Kommunikation der Resultate ist auf eine angemessene Ergebnisdarstellung zu achten. Dies betrifft insbesondere Angaben zum Rücklauf, zur Streuung der Daten und zu sinnvollen Vergleichswerten. Diese Angaben sind ebenso wie der Veranstaltungstyp (z.B. Wahl-/Pflichtveranstaltung, didaktisches Konzept, Lernziele) und -kontext (z.B. studentisches Vorinteresse, Schwierigkeitsgrad, Workload) bei der Interpretation der LVE-Ergebnisse zu berücksichtigen. Auf die Angabe von Lehrveranstaltungsbewertungen, die ganz anderen Rahmenbedingungen unterliegen und somit zu unzulässigen Vergleichen verleiten, ist bei der Ergebnisdarstellung ebenso zu verzichten, wie auf die Rückmeldung verletzender oder beleidigender Antworten auf offene Fragen.
- Der Prozess sollte durch Fortbildungsmöglichkeiten und Unterstützungsangebote (z.B. Peer-Mentoring) begleitet werden, ohne Lehrende jedoch zu entsprechenden hochschuldidaktischen Kursen zu verpflichten. Für die Rückmeldung besonders negativer LVE-Ergebnisse sind spezielle Verfahren zu implementieren, um negative Effekte, wie eine nachhaltige Frustration der Lehrenden, durch die Evaluation zu verhindern.

- Die Grenzen von LVE als Steuerungsinstrument im Sinne einer summativen Evaluation, d.h. einer abschließenden Bewertung der Lehrleistung zur Ableitung von Folgemaßnahmen, sind anzuerkennen. Sie können eine wichtige, aber sicherlich nicht die einzige Informationsgrundlage im Rahmen von zentralen Entscheidungen (z.B. Berufungen, Entfristungen oder Lehrpreisen) sein. Insbesondere sollten keine Automatismen an LVE-Ergebnisse geknüpft werden (z.B. Mindestziele oder kritische Schwellen). Je nach Kontext sind weitere geeignete Informationen zur Lehrleistung hinzuziehen. Dies kann z.B. im Rahmen von Lehrproben, Lehrkonzepten oder einer Expertenkommission unter Einbeziehung der Studierenden, erfolgen.
- Die Zielsetzungen und intendierten Nutzungszwecke von LVE sind allen Stakeholdern – insbesondere den Lehrenden und Studierenden – vorab transparent zu kommunizieren. Hierbei ist auch zu klären, wer die Ergebnisse einer LVE einsehen darf. Für bestimmte Zwecke (z.B. Nutzung im Rahmen von *Tenure-Track*-Verfahren) bedarf es darüber hinaus ggf. einer schriftlichen Fixierung in einer Evaluationsatzung. Daneben sind auch für LVE die Vorgaben aus der Datenschutz-Grundverordnung zu beachten, wobei die konkrete Umsetzung mit den Datenschutzbeauftragten an der jeweiligen Hochschule abzuklären ist.

Literatur

- American Sociological Association 2019: Statement on Student Evaluations of Teaching. www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_feb132020.pdf, letzter Aufruf 23. Juni 2020.
- Bar, T., Kadiyali, V., Zussman, A. 2009: Grade Information and Grade Inflation: The Cornell Experiment. *Journal of Economic Perspectives*, vol. 23, no. 3, 93–108.
- Boring, A. 2017: Gender Biases in Student Evaluations of Teaching. *Journal of Public Economics*, vol. 145, 27–41.
- Chávez, K., Mitchell, C.M.W. 2020: Exploring Bias in Student Evaluations: Gender, Race, and Ethnicity. *Political Science and Politics*, vol. 53, no. 2, 270–274.
- Chisadza, C., Nicholls, N., Yitbarek, E. 2019: Race and Gender biases in Student Evaluations of Teachers. *Economics Letters*, vol. 179, 66–71.
- Dresel, M., Tinsner, K. 2008: Onlineevaluation von Lehrveranstaltungen: Methodeneffekte bei der Onlineevaluation von Lehrveranstaltungen. *Zeitschrift für Evaluation*, 7. Jg., Heft 2, 183–211.
- Dresel, M., Rindermann, H. 2011: Consultation of University Instructors Based on Student Evaluations of Teaching Effectiveness: Multilevel Test of its Effectiveness under Consideration of Bias and Unfairness Variables. *Research in Higher Education*, vol. 52, no. 7, 717–737.
- Esser, H., 1997: Zweifel an der Evaluation der Lehre. *Wirtschaftswissenschaftliches Studium: Zeitschrift für Ausbildung und Hochschulkontakt*, 26. Jg., Heft 1, 45–49.

- Hamermesh, D.S., Parker, A.M. 2005: Beauty in the Classroom. Instructors' Pulchritude and Putative Pedagogical Productivity. *Economics of Education Review*, vol. 24, no. 4, 369–376.
- Johnson, V.E. 2003: *Grade Inflation. A Crisis in College Education*. New York: Springer.
- Klein, M., Rosar, U. 2006: Das Auge hört mit! Der Einfluss der physischen Attraktivität des Lehrpersonals auf die studentische Evaluation von Lehrveranstaltungen – eine empirische Analyse am Beispiel der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln. *Zeitschrift für Soziologie*, 35. Jg., Heft 4, 305–316.
- Kromrey, H. 1994a: Wie erkennt man gute ›Lehre‹? Was studentische Vorlesungsbefragungen (nicht) aussagen. *Empirische Pädagogik*, 8. Jg., Heft 2, 153–168.
- Kromrey, H. 1994b: Evaluation der Lehre durch Umfrageforschung? Methodische Fallstricke bei der Messung von Lehrqualität durch Befragung von Vorlesungsteilnehmern. In P. Mohler (Hg.), *Universität und Lehre. Ihre Evaluation als Herausforderung an die Empirische Sozialforschung*. Münster: Waxmann, 91–114.
- Marsh, H. 2007: Students' Evaluations of University Teaching: A Multidimensional Perspective. In R.P. Perry, J.C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. New York: Springer, 319–384.
- Penny, A.R., Coe, R. 2004: Effectiveness of Consultation on Student Ratings Feedback: A Meta-analysis. *Review of Educational Research*, vol. 74, no. 2, 215–253.
- Rindermann, H. 2001: *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierten Unterrichts*. Landau: Empirische Pädagogik.
- Rindermann, H. 2016: *Lehrveranstaltungsevaluation an Hochschulen. Der Einfluss der Rahmenbedingungen auf Qualität von Lehre und Ergebnisse von Lehrevaluation*. In D. Großmann, T. Wolbring (Hg.), *Evaluation von Studium und Lehre. Grundlagen, methodische Herausforderungen und Lösungsansätze*. Wiesbaden: Springer VS, 227–262
- Rivera, L.A., Tilcsik, A. 2019: Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation. *American Sociological Review*, vol. 84, no. 2, 248–274.
- Spooren, P., Brockx, B., Mortelmans, D. 2013: On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, vol. 83, no. 4, 598–642.
- Stark, P.B., Freishtat, R. 2014: An Evaluation of Course Evaluations. *ScienceOpen Research*, doi: 10.14293/S2199-1006.1.SOREDU.AOFRQA.v1.
- Treischl, E., Wolbring, T. 2017: The Causal Effect of Survey Mode on Students' Evaluations of Teaching: Empirical Evidence from Three Field Experiments. *Research in Higher Education* vol. 58, no. 8, 904–921.
- Weinberg, B.A., Hashimoto, M., Fleisher, B.M. 2009: Evaluating Teaching in Higher Education. *Journal of Economic Education*, vol. 40, no. 3, 227–261.
- Wolbring, T. 2013: *Fallstricke der Lehrevaluation. Möglichkeiten und Grenzen der Messbarkeit von Lehrqualität*. Frankfurt am Main, New York: Campus.