

## AUTOREN

Wer schreibt für Ökonomenstimme?

## ARTIKEL

Alle Artikel, Archiv

## THEMEN

nach JEL-Klassifizierung

## ÜBER ÖKONOMENSTIMME

Wer wir sind, FAQs

## Lehrevaluationen benachteiligen Wissenschaftlerinnen

1687 mal gelesen | [5 Kommentare](#)

Frederike Mengel, Jan Sauermann und Ulf Zölitz, 16. Sept. 2019

*Insbesondere von männlichen Studierenden und in mathematischen Fächern werden Dozentinnen schlechter bewertet als ihre männlichen Kollegen – trotz gleicher Lehrqualität.*

Obwohl der Anteil an Frauen in Graduierten-Programmen in den Wirtschaftswissenschaften seit Jahrzehnten stetig zunimmt, kann kein analoger Trend bei anschließenden universitären Karrieren verzeichnet werden (McElroy, 2016). Mögliche Erklärungsversuche für diese „leaking pipeline“ sind vielfältig und liegen sowohl auf der Angebotsseite, also z.B. an Präferenzen, aber auch auf der Nachfrageseite wie z.B. Diskriminierung. So zeigen Studien unter anderem, dass das Geschlecht für den Erfolg von Förderanträgen entscheidend ist (Broder, 1993, Van der Lee und Ellemers, 2015), dass Forschungspapiere von weiblichen Autoren eine geringere Qualität beigemessen wird (Krawczyk and Smyk, 2016), und dass der Beitrag von Forscherinnen zu wissenschaftlichen Kollaborationen als geringer bewertet wird als wenn (männliche) Forscher in Forschungsteams publizieren (Sarsons, 2017).

## Lehrevaluationen zur Leistungsbeurteilung von Dozierenden

Eine Form der Benachteiligung von Wissenschaftlerinnen kann auch in der Evaluierung von Lehrveranstaltungen liegen. Gerade für angehende Wissenschaftlerinnen und Wissenschaftler entscheiden Lehrevaluationen nicht selten über Einstellungs- oder Beförderungsentscheidungen und können somit einen Einfluss auf den beruflichen Werdegang ausüben. Bei den Betroffenen selbst sind schlechte Bewertungen durch Studierende nicht selten ein zusätzlicher Grund an der eigenen akademischen Eignung zu zweifeln. Diskriminierende Bewertungen können jedoch auch auf indirektem Weg den universitären Werdegang erschweren, wenn Wissenschaftlerinnen durch schlechtere Bewertungen dazu veranlasst werden Zeitkontingente von der Forschung in die Lehre zu transferieren, um vermeintlich schlechtere Bewertungen zu verbessern. Diese Verschiebung von ohnehin knappen Zeitressourcen führt fast unweigerlich zu qualitativ schlechteren Forschungsergebnissen, was Frauen auf dem akademischen Arbeitsmarkt weiter zurückwerfen kann.

Unter Einbeziehung eines Datensatzes von 19.952 Lehrevaluationen von Dozentinnen und Dozenten an einer niederländischen Universität, versucht unsere Studie einen möglichen *gender bias* in Lehrevaluationen zu belegen und zu quantifizieren (Mengel, Sauermann, Zölitz, 2019). In unserem Kontext beschreibt *gender bias* ein geschlechterspezifisch schlechteres Evaluationsergebnis, welches nicht durch objektive Unterschiede in der Qualität der Lehre erklärt werden kann. Zur kausalen Messung des Zusammenhangs machen wir uns die zufällige Einteilung von Studierenden in Kurse mit männlichen und weiblichen Dozenten zu eigen. Diese zufällige Einteilung schließt aus, dass geschlechterspezifische Unterschiede in der Dozentenbewertung auf Selektion von Studierenden in Kurse zurückzuführen sind. Dies wäre beispielsweise der Fall, wenn Frauen und Männer unterschiedliche Arten von Kursen unterrichteten oder kritische Studierende vermehrt Kurse von Dozentinnen besuchen. Des Weiteren enthalten die Daten neben den Evaluationsergebnissen detaillierte Informationen über die Selbststudienzeit und die akademischen Leistungen der Studierenden. Falls Lehre und Leistungsbewertung nicht von ein und derselben Person übernommen werden, kann Letzteres als objektives Leistungskriterium für die Dozierenden gesehen werden. Dies erlaubt es geschlechterverzerrte Bewertungen daraufhin zu überprüfen, ob sie auf Grundlage der Dozentenleistung gerechtfertigt sein können.

## Dozentinnen werden bei gleicher Lehrqualität schlechter bewertet

Unsere empirischen Ergebnisse belegen eine systematische Benachteiligung von Dozentinnen gegenüber ihren männlichen Pendanten, obwohl das Geschlecht der Dozierenden weder einen Einfluss auf die Noten noch den Arbeitsaufwand der Studierenden ausübt. Die niedrigeren

## Autoren

## Frederike Mengel



## Jan Sauermann



## Ulf Zölitz



## Schlagworte

Diskriminierung, Geschlecht, Lehre, Wissenschaft

## Weitersagen

[+](#) Share / Save [f](#) [t](#) [r](#)

## Ähnliche Artikel

- Ist Kompetitivität in Ausbildungs- und Arbeitsmärkten immer vorteilhaft?
- Warum ist der Anteil von MINT-Absolventinnen in Marokko höher als bei uns?
- Kann Peer-Mentoring die Wirksamkeit von Online-Lehre verbessern?
- Frauen an die Macht – aber wie?
- Was sollen wir erforschen? Wie ÖkonomenInnen über Ökonomik denken

Lehrevaluationen der Dozentinnen sind größtenteils auf die Bewertungen männlicher Studierenden zurückzuführen, auch wenn gleiches Verhalten in geringerem Ausmaß bei weiblichen Studierenden zu beobachten ist. Ungerechtfertigt schlechte Bewertungen treten dabei besonders häufig in der Evaluierung von mathematischen Kursen auf, in denen die Voreingenommenheit gegenüber Dozentinnen besonders ausgeprägt scheint. Darüber hinaus sinkt der Grad der Benachteiligung mit dem Alter der Dozentinnen, sodass vermehrt junge Dozentinnen schlechtere Bewertungen erhalten. Gegeben, dass diese meist noch am Anfang ihrer akademischen Karriere stehen, und weniger Erfahrung haben als ihre älteren Kolleginnen, sind sie besonders anfällig ihre akademischen Pläne zu überdenken oder zeitliche Ressourcen auf die vermeintliche Verbesserung ihrer Lehre zu schieben.

Interessanterweise setzt sich der beschriebene *gender bias* auch in der Evaluation von Lehrmaterialien fort. Dies scheint umso erstaunlicher, da diese Materialien unabhängig vom Geschlecht des Dozenten identisch für alle Studierenden dergleichen Kursreihe sind. Theoretisch könnte dies durch die unterschiedliche Art und Weise wie Dozentinnen und Dozenten das zur Verfügung stehende Lehrmaterial nutzen erklärt werden. Wahrscheinlicher und weniger konstruiert erscheint die Möglichkeit, dass Studierende ihre Voreingenommenheit gegenüber Dozentinnen auch auf weitere geschlechtsunabhängige Gesichtspunkte ausdehnen (Spillover-Effekte).

### Wie sinnvoll sind Lehrevaluationen?

Die Ergebnisse unserer Untersuchung legen nahe, dass die Evaluierung von Lehrveranstaltungen durch Studierende nur unter Vorbehalt genossen werden sollte. Obwohl Lehrevaluationen für Einstellungs- und Beförderungsentscheidungen genutzt werden, werden sie in der Regel nicht für mögliche geschlechterspezifische Verzerrungen, die Geschlechterzusammensetzung der Kurse oder die Teilnahmequote an der Bewertung korrigiert. Durch diese Nichtberücksichtigung partizipieren Frauen weniger an Gehaltserhöhungen und werden letztendlich seltener befördert als ihre männlichen Kollegen.

Neben diesen möglichen direkten Auswirkungen treten weitere indirekte Folgen hinzu, wie durch die Reallokation von Zeitressourcen von der Forschung in die Lehre. All dies trägt letztlich zu schlechteren Karriereaussichten für Frauen in der Wissenschaft bei. In Anbetracht dieser teils gravierenden Folgen für den Erfolg am akademischen Arbeitsmarkt scheint es erstrebenswert die Bedeutung von Lehrevaluationen für die Leistungsbeurteilung zu überdenken und ihnen im Ganzen weniger Gewicht zukommen zu lassen. So hat sich die American Sociological Association erst kürzlich explizit gegen die Nutzung von Studierendenbewertungen wegen der genannten Verzerrungen ausgesprochen (American Sociological Association, 2019).

## Literatur

American Sociological Association. 2019. Statement on Student Evaluations of Teaching, September.  
[https://www.asanet.org/sites/default/files/asa\\_statement\\_on\\_student\\_evaluations\\_of\\_teaching\\_sept52019.pdf](https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_sept52019.pdf)

Broder, Ivy E. 1993. Review of NSF economics proposals: Gender and institutional patterns. American Economic Review 83(4) 964-970.

McElroy, Marjorie B. 2016. Committee on the status of women in the economics profession (CSWEP). American Economic Review 106(5) 750-773.

Krawczyk, Michal W., Magdalena Smyk. 2016. Author's gender affects rating of academic articles - evidence from an incentivized, deception-free experiment. European Economic Review 90?326-335.

Mengel, Friederike, Jan Sauermann, Ulf Zölitz. 2019. Gender bias in teaching evaluations. Journal of the European Economic Association, 17(2) 535-566.

Sarsons, Heather. 2017. Recognition for Group Work: Gender Differences in Academia. AER Papers and Proceedings, 107.

Van der Lee, Romy, Naomi Ellemers. 2015. Gender contributes to personal research funding success in The Netherlands. Proceedings of the National Academy of Sciences of the United States of America 112(40) 12349-12353.

## Kommentare

Dieser Artikel hat 5 Kommentare.

---



### Kein Titel

geschrieben von Economist101 am 16. Sep. 2019, 15:52

[ antworten ]

"versucht unsere Studie einen möglichen gender bias in Lehrevaluationen zu belegen und zu quantifizieren"

Und wie groß ist der Effekt?

---



### Ergebnis nicht belastbar

geschrieben von Economist101 am 20. Sep. 2019, 10:26

[ antworten ]

In Ermangelung einer Antwort habe ich nun selbst in das Paper, das diesem Beitrag zugrundeliegt, geschaut, um zu erfahren, wie groß der Unterschied in den Lehrevaluationen von Männern und Frauen ist. Was ich fand war ernüchternd.

Die Untersuchung basiert laut Anmerkungen zu Tabelle 1 im Paper auf Fragen zur Lehrqualität auf einer Skala von "sehr gut" über "ausreichend" bis "sehr schlecht". Um "die Analyse zu vereinfachen" wurden diesen Angaben dann Zahlen von 1 bis 5 zugeordnet und aus diesen dann Durchschnitte gebildet und zwischen Männern und Frauen verglichen, bzw. OLS-Regressionen geschätzt.

Die Lehrqualität ist offensichtlich eine ordinale Variable und die Mittelwerte einer solchen Variable sind völlig bedeutungslos, wie man eigentlich in jedem Anfänger-Kurs zu Statistik lernt.

Insofern ist m.E. nichts, was die Autoren hier behaupten, belastbar. Es mag einen Unterschied geben, aber in welche Richtung und ob er am Ende auch relevant (und nicht nur statistisch signifikant) ist, das bleibt eine offene Frage. Wieder mal viel Lärm um nichts (oder etwas, man weiß es halt nicht).

---



### AW:Ergebnis nicht belastbar

geschrieben von JanSauermann am 22. Sep. 2019, 22:09

[ antworten ]

Die Studenten in unserer Studie werden gebeten, Antworten von 1 ("sehr schlecht") bis 5 ("sehr gut") zu beantworten. Wir haben uns entschieden, diese mit Ordinary Least Squares zu schätzen, hauptsächlich weil sie damit leicht interpretierbar sind. Wenn wir andere Methoden verwenden würden (z.B. ordered Logit) erhalten wir praktisch identische Ergebnisse. Auch wenn wir eine mehr anders definierte abhängige Variable verwenden, erhalten wir die gleichen Ergebnisse. Studenten wurden zusätzlich nach einer Gesamteinschätzung Tutors / der Tutorin gefragt. Diese ist auf einer Skala von 1-10 definiert und resultiert ebenfalls in den gleichen Ergebnissen.

Unsere Studie ist nicht die einzige, die zeigt, dass Lehrevaluationen verzerrte Performance Messungen sind. Neben einer aktuellen Studie von Anne Boring im Journal of Public Economics (Rotterdam: <https://www.sciencedirect.com/science/article/pii/S0047272716301591>) enthält Rebecca Kreitzers Website Verweise zu über 70 Studien zur Benachteiligung von Frauen bei Lehrevaluationen. Dass ähnliche Ergebnisse in unterschiedlichen Umgebungen / Universitäten gefunden werden, deutet weiter darauf hin, dass es sich hier sehr wohl um ein Problem handelt. Herzliche Grüsse,  
Jan Sauermann

---



### AW:AW:Ergebnis nicht belastbar

geschrieben von JanSauermann am 22. Sep. 2019, 22:10

[ antworten ]

Der zweite Link ging leider verloren: <http://www.rebeccakreitzer.com/summaries-of-research-articles/>

---



### Effektgröße?

geschrieben von Economist101 am 23. Sep. 2019, 10:36

[ antworten ]

Vielen Dank für die Antwort, Herr Sauermann!

"Wir haben uns entschieden, diese mit Ordinary Least Squares zu schätzen, hauptsächlich weil sie damit leicht interpretierbar sind."

Die sind überhaupt nicht mehr sinnvoll interpretierbar. Mittelwertunterschiede von ordinalen Variablen sind bedeutungslos

sind bedeutungslos.

"Wenn wir andere Methoden verwenden wuerden (z.B. ordered Logit) erhalten wir praktisch identische Ergebnisse."

Dann würde ich gerne meine Frage vom Anfang wiederholen: Wie groß ist der gefundene Effekt, wenn man eine angemessene Methode wählt? Um wie viel größer ist bspw. die Wahrscheinlichkeit mit "gut" oder besser bewertet zu werden, wenn man ein Mann statt eine Frau ist? Sprechen wir von 10 Prozentpunkten oder 0,001 (statistisch signifikanten, aber völlig irrelevanten) Prozentpunkten?

Sie müssen sich [anmelden](#) um Kommentare zu schreiben.

---

[Home](#) | [Autoren](#) | [Themen](#) | [Archiv](#) | [Über Ökonomenstimme](#)

[Partner: Vox \(UK\)](#) | [Telos \(Frankreich\)](#) | [Me Judice \(Niederlande\)](#) | [LaVoce \(Italien\)](#) | [Nada es Gratis \(Spanien\)](#)

[Impressum](#) | [Rechtliche Hinweise](#)



Ökonomenstimme ist ein Angebot der  
KOF Konjunkturforschungsstelle der ETH Zürich