

Rezension zu:

Wolbring, Tobias: *Fallstricke der Lehrevaluation. Möglichkeiten und Grenzen der Messbarkeit von Lehrqualität*. Frankfurt: Campus, 2013. 365 Seiten, 39,90 EUR, ISBN 978-3-593-39961-4

*Heiner Rindermann*¹

Dr. Tobias Wolbring hat an der Ludwig-Maximilians-Universität München Soziologie studiert und dort promoviert. Er ist jetzt an der Eidgenössischen Technischen Hochschule Zürich im Bereich Evaluations- und Hochschulforschung tätig. Im Frankfurter Campus-Verlag hat er ein Buch, seine Dissertation, mit empirischen Studien zur studentischen Lehrevaluation vorgelegt. Der Titel „Fallstricke der Lehrevaluation“ legt nahe, dass ein eher kritischer Blick auf die Aussagekraft von Lehrevaluationsergebnissen und ihre Verwendung geworfen wird. Zunächst zum Inhalt:

Zu Anfang setzt sich Wolbring mit grundsätzlichen Fragen der Wissenschaftstheorie auseinander. Kann man mit Wissenschaft einen Zugang zur objektiven Wirklichkeit finden? Existieren kausale Beziehungen, können diese aufgeklärt werden und welche Rolle spielt hierbei in der Evaluationsforschung das Experiment? Lassen sich menschliches Handeln und, als ein Teilbereich von diesem, das Besuchs- und Beurteilungsverhalten von Studierenden besser durch die Modelle Rational Choice oder das der begrenzten Rationalität beschreiben?

Diese dienen als Hintergrund, um evaluationsbezogene Fragestellungen zu fokussieren: Werden studentische Beurteilungen der Hochschullehre durch nicht mit der Qualität dieser zusammenhängende Faktoren beeinflusst (Validitätsfrage)? Beispielsweise könnten das Geschlecht oder eine spezifische Geschlechtskombination Studierender mit Lehrenden die Beurteilung beeinflussen, ohne dass die Qualität der Lehre davon verändert wird. In diesem Falle wäre das Urteil nicht mehr objektiv und valide. Spielen bei Beurteilung und Qualität von Lehre auch nicht von der Lehrkraft zu verant-

wortende Faktoren eine Rolle (Fairnessfrage)? So könnten Seminare mit hohen Besucherzahlen kritischer bewertet werden, was aber valide ist, da die Lehre an sich unter einer hohen Besucherzahl leidet. Die Dozenten könnten für solche Rahmenbedingungen nicht verantwortlich gemacht werden. Die Ergebnisse von Lehrenden unter solch verschiedenen Rahmenbedingungen wären damit nicht mehr vergleichbar. In Wolbrings Studien werden in verschiedenen Designs, mit gegebenen Lehrevaluationsdaten oder durch Experimente, solche Einflussfaktoren näher untersucht. Dazu gehören unter anderen die Attraktivität der Lehrenden, Beurteilerstandards, Geschlecht, Leistungsniveau, Veranstaltungstag, Veranstaltungszeit, Teilnehmerzahl, Erwerbstätigkeit, Anfahrtsweg, Vorinteresse, Absentismus und Drop-Out.

Empirisch stellt Wolbring fest, dass studentische Lehrevaluationen von nicht mit der Qualität zusammenhängenden und nicht von den Dozenten zu verantwortenden Faktoren beeinflusst werden. Besonders interessant, auch theoretisch, ist seine Auseinandersetzung mit dem Merkmal Attraktivität von Lehrenden. Hier wählt Tobias Wolbring drei Zugänge: (a) reale Lehrevaluationsdaten mit zusätzlichen Attraktivitätsratings anhand von Fotos der Dozenten, (b) eine experimentelle Variation mit fiktiven Bildern und einem ausschließlich zu hörenden Vortrag und (c) die Untersuchung des Zusammenhangs von gegebener Attraktivität mit Besuchsverhalten in Veranstaltungen. Hierbei zeigte sich, dass die Lehre attraktiver Dozenten leicht besser eingeschätzt wird ($r=.16$, S. 137; $r=.02$ bis $.09$, S. 150). Aber auch das Besuchsverhalten wird positiv beeinflusst: Veranstaltungen attraktiver Dozenten werden häufiger besucht als die weniger

1 Technische Universität Chemnitz

attraktiver. Der Zusammenhang in realer Lehre zwischen Attraktivität und Lehrevaluationsergebnissen ist höher ($r=.16$, S. 137) als der in einer experimentellen Variation ($r=.02$ bis $.09$, S. 150). Dies ist ein Hinweis auf ein komplexes Wechselspiel zwischen Merkmalen, die in sozialen Interaktionssituationen unterstützend wirken („produktionswirksam“, S. 131). Absentismus führt zu leichten Verzerrungen, d.h. Veranstaltungsrankings werden durch selektives Fehlen angehoben.

Am Ende diskutiert Wolbring Möglichkeiten, Lehrevaluationsergebnisse so zu korrigieren, dass auf ihnen basierende Rankings von Dozenten und Veranstaltungen weniger durch Hintergrundfaktoren verzerrt werden. Er kommt hier zu einem eher skeptischen Fazit. Verzerrungen seien nie völlig bekannt und deswegen immer nur teilweise korrigierbar. Deshalb sollten an studentische Lehrevaluationen auch keine Mittelverteilungen gebunden werden. Sie sollten dagegen, um Lehre zu verbessern, formativ für Rückmeldung, Beratung und Fortbildung genutzt werden.

Tobias Wolbring zieht für seine Analysen komplexe statistische Verfahren wie Mehrebenenanalysen, Imputationen und Regressionen mit variierenden Prädiktorzusammensetzungen heran. Verschiedene Prädiktoren, theoretisch als Kausalfaktoren verstanden, werden in unterschiedlicher Kombination zu Kriterien, Lehrevaluationsergebnissen und anderen Indikatoren, wie Anwesenheit von Studierenden, in Beziehung gesetzt. Allerdings lassen sich die Ergebnisse oft schwierig interpretieren, da Korrelationen nicht durchgängig und standardisierte Regressionskoeffizienten nie dargestellt werden. Dabei wäre es unschwer, den Ergebnistabellen immer Korrelationen voranzustellen und je Prädiktor den standardisierten Regressionskoeffizienten hinzuzufügen. Ohne diese Information ist die Größe der Effekte schwer abzuschätzen. Bei dem Vergleich der Veranstaltungsrankings, etwa mit und ohne selektivem Besucherschwund,

wäre es naheliegend, die Rankings korrelativ zu vergleichen. Der Korrelationskoeffizient hätte ermöglicht, zu überprüfen, inwieweit die beiden Rangreihen ähnlich sind und wie groß die Aussagekraft von Evaluationen am Ende des Semesters ist. Platzwechsel in Rankings reichen hierfür nicht aus. Die Auswahl der Methoden und die Art der Ergebnispräsentation folgen den Fächertraditionen in der Soziologie und den Wirtschaftswissenschaften, sie erleichtern aber auch einen kritischen Blick auf Lehrevaluationen.

Wissenschaftstheoretisch könnte man noch anmerken, dass ein Hinweis auf Praxisferne als Argumentationsgrundlage gegen Konstruktivismus, Solipsismus und Agnostizismus ergänzungsbedürftig ist (S. 23). Realistische Zugänge könnten noch überzeugender begründet werden, etwa über den Nachweis einer Selbstwidersprüchlichkeit der alternativen Ansätze und insbesondere über eine epistemologische Argumentation. Bei Rational Choice lässt sich fragen, ob eine Annahme optimaler kognitiver Rationalität von Akteuren theoretisch notwendig ist: Reicht es nicht aus, wie von Wolbring mit Norman Braun angeführt (S. 31), Rational Choice auf die anthropologische Annahme zu gründen, dass Menschen nur „nach dem eigenen Urteil“ Nutzen maximierende Akteure sind?

Tobias Wolbring zeigt mit seiner Arbeit, dass das Thema Lehrevaluation durch neue Zugänge und Methoden theoretisch wie statistisch originell und sophisticated bearbeitbar ist. Auf empirischer Grundlage sind die Möglichkeiten und Grenzen, studentische Lehrevaluation als Maß von Lehrqualität zu verwenden, auslotbar. Zusammenfassend lässt sich festhalten, dass Lehrevaluationsergebnisse nicht verzerrungsfrei sind („keine vollkommen unverzerrte Messung“, S. 305). Das ist sicherlich zutreffend. Aber welches Beurteilungsverfahren ist verzerrungsfrei? Die Brauchbarkeit eines Messverfahrens hängt auch immer von den Alternativen ab, die zur Verfügung stehen. Wie könnte man Lehre ähnlich kostengünstig und valide beurteilen?