

## Rezension zu:

Hense, Jan/Rädiker, Stefan/Böttcher, Wolfgang/Widmer, Thomas (Hg.): Forschung über Evaluation. Bedingungen, Prozesse und Wirkungen. Münster u.a.: Waxmann 2013. 282 Seiten, 29,90 EUR, ISBN 978-3-8309-2829-4

*Wolfgang Beywl*<sup>1</sup>

*Überblick:* Die Einleitung der vier Herausgeber benennt als Motiv für den Sammelband, den spärlichen Fundus von Forschung über Evaluation im deutschsprachigen Raum zu erschließen, aus möglichst verschiedenen Disziplinen und Anwendungsfeldern. Die meisten Beiträge behandeln dabei Evaluationen oder evaluationsähnliche Verfahren im Bildungsbereich (Schule, Hochschule, Weiterbildung). Die übergreifende Fragestellung lautet: Wie lassen sich die methodologischen Grundprinzipien und Verfahrensweisen der Evaluation durch eine systematische, empirisch wie theoretisch fundierte Evidenz (besser) begründen?

Das Buch enthält zwölf Studien in vier Hauptteilen, wobei die drei ersten einer für die Evaluation typischen Programmlogik (z.B. CIPP-Modell von Daniel Stufflebeam) folgen:

Teil 1: Ausgangsbedingungen von Evaluationen, darunter sozialer Kontext und Kompetenzen von Evaluierenden.

Teil 2: Prozess der Evaluation, darunter Interaktion mit den Stakeholdern, ethische Herausforderungen und Fragen der praktischen Umsetzung.

Teil 3: Wirkungen und Folgen von Evaluationen, insbesondere Evaluationsnutzung sowie deren intendierte und nicht intendierte Resultate.

Teil 4: Studien zu übergreifenden methodologischen Fragen und Überblick zur internationalen (englischsprachigen) Forschung.

### *Vorstellung von Beiträgen:*

Teil 1: Ursula Koch präsentiert ihre Untersuchungsergebnisse zu den „Datenauswertungskompetenzen von Lehrkräften“. Hintergrund sind die im deutschen Schulwesen seit Mitte 2003 inzwischen flächendeckend einge-

führten Vergleichsarbeiten. Da die gewonnenen empirischen Ergebnisse zur Schul- und besonders Unterrichtsentwicklung genutzt werden sollen, komme es auf die Datenkompetenz der Lehrpersonen an, so dass aufwändige (und Schüler(innen), Eltern und Lehrpersonen vielfach fordernde) Datenerhebungen überhaupt genutzt werden können. Da diese in der Regel fehlerhaft, entwickelt Frau Koch eine eintägige Fortbildung zur Visualisierung und Auswertung von Schülerleistungsdaten und evaluiert diese. Unter anderem sollten Lehrpersonen interpretationsleitende Fragestellungen formulieren können (diese Schlüsselkompetenz erfordert meines Erachtens intensives und mehrfaches Üben mit realen Fällen). Ausweislich eines selbst entwickelten „Datenkompetenztests“ (insbesondere Wissen und dessen Anwendung in Bezug auf unterschiedliche Diagrammtypen) lernen Lehrkräfte in Bezug auf den gebotenen Stoff dazu, doch fällt es ihnen sehr schwer, logische Modelle mit modulierbaren unabhängigen Variablen zu entwickeln, die für eine systematische, die Daten nutzende Unterrichtsentwicklung unabdingbar seien. Erneut wird durch Koch belegt, dass auf die Produktion von Daten oft keine oder unangemessene Nutzungen folgen, also zusätzlich zu Daten- auch Evaluationskompetenzen erforderlich sind.

Teil 2: Stefan Rädiker bearbeitet eine ähnliche Fragestellung (für den Weiterbildungsbereich). Er fokussiert Umsetzungsprobleme und Kompetenzanforderungen von Bildungsprozessen in 50 verschiedenen, meist nichtkommerziellen Weiterbildungsorganisationen, die ein bestimmtes Modell des „lernerorientierten Qualitätsmanagements“ einsetzen. Einer von elf „Qualitätsbereichen“ ist die „Evaluation der Bildungs-

1 Univation – Institut für Evaluation, Köln

prozesse“. Mittels Inhaltsanalyse vorliegender Selbstbeschreibungen der jeweiligen Evaluationspraxis wird der Status quo erfasst: Die Anwendenden verfügen über kein systematisch erworbenes Evaluationswissen; entsprechend dominiert als Methode der Einschätzfragebogen. Für die Hälfte der Weiterbildungsorganisationen werden mittels Leitfadenterviews Herausforderungen und Kompetenzanforderungen konkretisiert. Vielfach berichten die Befragten über einen geringen Nutzen der Datenerhebungen bei gleichzeitig hoher subjektiv empfundener Belastung. Abschließend erstellt Rädiker ein Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen (wobei der Bezug auf die einschlägige Kompetenztaxonomie von Russ-Eft, Bober, de la Teja, Foxon und Koszalka in *Evaluator Competencies* (2008, San Francisco: Jossey-Bass) fehlt).

Matthias Müller schreibt zur Evaluation in der Sozialen Arbeit. Die Nichtnutzung resultiere hier eher aus einem Sicherungsreflex im Interesse der eigenen Autonomie, zum Beispiel von Verbandsfunktionären oder von vornehmlich an Rechenschaftslegung interessierten Sozialpolitikern. Von einer produktiven „Kultur der Evaluation“ scheint dieser Handlungsbereich meist weit entfernt.

Teil 3: Holger Gärtner untersucht das vielfach als unverzichtbar für die Unterrichtsentwicklung bezeichnete „Schülerfeedback“. An einem in Berlin und Brandenburg für alle Lehrkräfte kostenlos verfügbaren Selbstevaluationsportal nehmen 2% von 50.000 Lehrkräften teil, davon beteiligen sich 25% an Gärtners Online-Befragung zum Nutzen des gebotenen Instrumentariums (der Autor weist auf die vermutlich positive verzerrende Selbstselektion hin). Wie in vielen anderen onlinebasierten Feedbackverfahren werden Fremdeinschätzungen von Lernenden mit Selbsteinschätzungen von Lehrenden verglichen. Dies ist Ausgangspunkt für lösungsorientierte Klassendiskussionen. Zwar fänden solche statt, doch würden die Ergebnisse wenn überhaupt nur sehr individuell verarbeitet. Bei offensichtlicher Akzeptanz und eingeschätzter Nützlichkeit in der Stichprobe funktionieren das Verfahren oft affirmativ und seltener innovativ-entwicklungsanregend.

André Nowakowski präsentiert ein variantenreiches Untersuchungsdesign, um die Effekte von lernprozess- versus lernergebnisfokussierten Einschätzskalen auf die Lehrhaltungen der Lehrenden (Lehrenden- versus Lernenden-Fokussierung) zu überprüfen.

Mary Sandermann behandelt ebenfalls das Thema der Bedeutsamkeit von Kompetenzen, hier der sozialen Kompetenzen bei Evaluieren-

den und Nutzenden. Hierzu wird einschlägige aktuelle Evaluationsfachliteratur eingeregelt aufbereitet. Daraus resultiert ein „Minimalset“ von sechs Kompetenzbereichen für Evaluierende. Sandermanns Empirie beschäftigt sich mit den externen Evaluationen von sieben universitären Mentoringprogrammen. Es werden „fünf Pfade zur direkten Nutzung“ aufgezeigt, darunter ausgeprägte Teamorientierung derjenigen Person, die zwischen Evaluationsteam und Programmverantwortlichen vermittelt.

Kathrin Frey und Thomas Widmer legen eine Untersuchung aus einem ganz anderen Bereich vor: Sie fragen: „[W]erden Evaluationen eher beigezogen, wenn sie sich auf die Politik im eigenen Land beziehen, oder werden Befunde bevorzugt, die von Evaluationen stammen, die andernorts zu vergleichbaren Gegenständen erarbeitet wurden?“ Sie untersuchen, ob im nationalen Gesetzgebungsprozess anhand von in- und ausländischer Evidenz gelernt wird (Basis sind zwölf Fälle aus den Politikbereichen Verkehrssicherheit, Drogengesundheit, Versorgung, Asyl). Erklärt werden soll, welche Art von Rationalität die Entscheidenden heranziehen – eine analytische, die auf Verbesserung durch Lernen aus Evidenz abzielt, oder eine politische, die auf Lernen verzichten kann. Im Ergebnis dominiert die politische Verwendungsrationalität, wobei ausländische Evidenz wenn dann fast ausschließlich als zusätzliche Bestätigung der eigenen Position herangezogen wird.

Teil 4: Wolfgang Böttcher, Nina Hogrebe und Sabrina Schulz versuchen eine dialogische Klärung, wie Evaluationen (aus der Erziehungswissenschaft) mittels systematischer Reviews synthetisiert werden können (mittels Meta-Analysen, Meta-Ethnographie oder thematischen Reviews). Wie können sowohl praktische Tauglichkeit als auch sozialwissenschaftliche Güte von Evaluationen erhöht werden? Ähnlich wie andere Autoren des Sammelbandes stoßen auch sie vielfach auf nicht rationale Nutzungen von Daten aufgrund im Voraus festgelegter bildungspolitischer Verwendungsabsichten. Sie kommen zum Schluss, dass Daten und Interpretationen nicht ausreichen, sondern dass die Evaluierenden auch inhaltliche Aussagen über die konkrete Ausgestaltung zu ergreifender Maßnahmen treffen müssen (womit sie zum Programmkonzept beitragen). Mehr noch als Bildungsforschung müsse Bildungsevaluation dafür sorgen, dass ihre Ergebnisse auch genutzt werden, namentlich durch zielgeführte Dissemination. Da sie sich von einer Synthetisierung mehrerer Evaluationen zum gleichen Thema Zusatzschub versprechen (aktuelles Beispiel: Visible Learning

von John Hattie), plädieren sie – obgleich utopisch – für die Institutionalisierung via einer „Agentur für Systematische Reviews und das Makeln von Evidenz in Politik und Praxis“.

Wolfgang Böttcher, Jan Hense und Miriam Keune geben einen Forschungsüberblick zur Schulinspektion, oft auch als externe Evaluation bezeichnet. In Bezug auf vorliegende Schulinspektionsmodelle unterscheiden sie drei Forschungsansätze: Studien zur Akzeptanz, zu Maßnahmen infolge von Inspektionen, zur Wirkung von Inspektionen auf Schülerleistungen. Während Akzeptanz oft (mehr bei Schulleitenden als bei Lehrpersonen) vorliege, folgten oft unerwartet wenige Maßnahmen (insbesondere zum Unterricht): Mittlerweile im Überfluss vorliegende Daten resultierten in wenig systematisch abgeleitete Therapien. Schließlich gebe es wenig Anhaltspunkte dafür, dass Schulinspektionen auch nur in geringem Maße zu verbesserten Schülerleistungen beitragen. Ihr Fazit: Während Schulinspektionen mithin Schwachpunkte aufdecken, gelingt die Übersetzung in Maßnahmen der Unterrichts- und Schulentwicklung selten. Die Autoren empfehlen – wie bereits andere – Erhöhung der Evaluationskompetenzen, hohe Kommunikationsqualität, Einbeziehung der Stakeholder.

Wolfgang Meyer und Jörg Rech bemühen sich um eine Erschließung des Zentralbegriffs der (in vielen anderen Aufsätzen ja auch als defizitär definiert angesprochenen) „Evaluationskultur“. Ihre kreativ und heuristisch vorangetriebene Darlegung mündet – recht prosaisch – in der Vorstellung des Deutschen Evaluationsinstituts für die Entwicklungszusammenarbeit, in dessen Programmatik die Autoren Hoffnungen für die Entwicklung der Evaluationskultur sehen.

Der abschließende Beitrag von Jan Hense und Thomas Widmer leistet einen breiten aktuellen Aufriss zum internationalen Stand der Forschung über Evaluation (englischsprachig, vorwiegend aus den USA stammend). Sie zeichnen die Entwicklung von den naiven Wirkungserwartungen zur reflektierten Nutzungsforschung nach und listen angelehnt an amerikanische Standardliteratur zentrale Fragestellungen für ein Forschungsprogramm zur Evaluation auf. Die in den USA zuletzt intensiviertete Forschung über Evaluation deuten sie als Zeichen für zunehmende Professionalisierung. Analog zur Gesamtgliederung des Sammelbandes referieren sie Forschungsergebnisse zu den Bedingungen der Evaluation (unterschieden nach Angebots- und Nachfrageseite), solche zum Evaluationsprozess (wobei mich das Kriterium der Implemen-

tierungstreu verwirrt, insofern sie ein Merkmal des Programms und nicht der Evaluation ist) und zu den Wirkungen und Folgen von Evaluationen. Hier leisten die beiden Autoren einen bemerkenswerten Beitrag zur Verdichtung des Forschungswissens, wobei ich die von ihnen bewusst ausgeklammerten (wenigen) deutschsprachigen Forschungsberichte vermissen.

*Diskussion:* Das Buch ist ein verdienstvoller Ansatz, eine schon viel zu lange bestehende dramatische Lücke zu schließen: Es ist ein Statusbericht zur Forschung über Evaluation im deutschsprachigen Raum als Basis für künftig zu intensivierende Bemühungen. Mehr als Vorhandenes kann nicht aufbereitet werden, doch das machen die Autorinnen und Autoren informativ und argumentativ, meist empirisch gestützt.

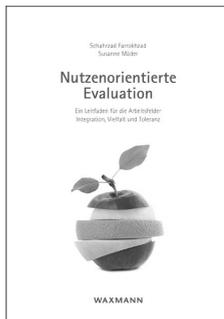
Einige der Beiträge sind verdichtende Resümeees aus umfangreichen Dissertationen (Frey, Koch, Rädiker, Sandermann), dabei ein Gewinn für die Lesenden, da die Texte mit Abstand geschrieben und dabei gereift sind. Mehrere Autoren bleiben ihrem jeweiligen feldspezifischen bzw. disziplinären Referenzsystem verhaftet (anders z.B. Frey/Widmer), so dass die für das Buch programmatisch gesetzte inter- oder gar transdisziplinäre Perspektive selten, z.B. im Schlussbeitrag, aufblitzt. Auch die zu wünschende Theoriebasierung der Forschung ist nicht immer gegeben.

Die Herausgeber verzichten auf eine die Beiträge zusammenfassende Definition von Evaluation. So bleibt die Relation zu Qualitätsmanagement und Auditierung (aus welchen Domänen einige Beiträge stammen) unklar. Und auch der babylonischen Terminologieverwirrung (z.B. bzgl. Outcomes oder „Ergebnisse“) wird damit freie Bahn gelassen (Sandermann – S. 148 – bemerkt kritisch, es werde „eher intuitiv mit Begrifflichkeiten umgegangen“ sic!). Die Herausgeber wären gut beraten, diesen Wink bei ihren künftigen Arbeiten aufzugreifen bzw. die für Forschung über einen Gegenstand unabdingbare sprachliche Klärung zumindest zu problematisieren. Zwei weitere Anregungen: Vielleicht wäre es erkenntnisfördernd, unterschiede man Evaluationen, die in Politikarenen stattfinden von solchen in organisationalen Kontexten. Ein weiteres Desiderat ist die Aufarbeitung der reichlich vorhandenen Forschung über die Nutzung von (sozialwissenschaftlicher) Forschung und von deren Bezügen zur Forschung über Evaluation.

Nimmt man das Buch als – wenn nicht repräsentative so jedenfalls bedeutsame – Stich-

probe aus Forschungen zur Evaluation, so kann als zentrale Erkenntnis festgehalten werden, dass sorgfältige kommunikativ gerahmte Nutzungsvorbereitung und die dafür erforderlichen Kompetenzen der Evaluierenden ausschlaggebend sind dafür, dass Evaluationen einen hohen sozialen Wert erlangen können. Selbstverständlich sind auch empirisch-methodische Kompe-

tenzen von Belang – und hier gibt die Dominanz der in den Beiträgen vorzufindenden Einschätzungsskalen zu denken. Dass derlei Schlüsse herausgearbeitet werden können, belegt, dass das Buch die versprochene Basis tatsächlich schafft, professionelle Evaluation evidenzbasiert weiterzuentwickeln.



Schahrazad Farrokhzad, Susanne Mäder

## Nutzenorientierte Evaluation

Ein Leitfaden für die Arbeitsfelder Integration, Vielfalt und Toleranz

2014, 144 Seiten, br., 24,90 €

ISBN 978-3-8309-3065-5

E-Book: 21,99 €; ISBN 978-3-8309-8065-0

Strategien und Methoden der nutzenorientierten Evaluation werden hier vorgestellt, insbesondere für Evaluationen in den Themenfeldern Integration, Vielfalt und Toleranz. Am Beispiel konkreter Evaluationsinstrumente veranschaulichen die Autorinnen evaluationstheoretische und -praktische Ausführungen.

Das Buch richtet sich an Projektverantwortliche, die Evaluationen selbst durchführen oder externe Evaluationen beauftragen wollen. Auch für Studierende und Evaluierende mit Interesse an einem nutzenorientierten Evaluationsansatz hält der Band praxisnahe Informationen bereit.



WAXMANN

## 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*<sup>1</sup>

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

<sup>1</sup> 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?