

Tobias Wolbring

FALLSTRICKE DER LEHREVALUATION

*Möglichkeiten und Grenzen der Messbarkeit
von Lehrqualität*

campus

Fallstricke der Lehrevaluation

Hochschule und Gesellschaft

Herausgegeben von Georg Krücken

Band 1

Tobias Wolbring, Dr. rer. pol., ist Postdoctoral Fellow am Departement Geistes-, Sozial- und Staatswissenschaften der ETH Zürich.

© Campus Verlag GmbH

Tobias Wolbring

Fallstricke der Lehrevaluation

Möglichkeiten und Grenzen der Messbarkeit
von Lehrqualität

Campus Verlag
Frankfurt/New York

© Campus Verlag GmbH

Zugl.: Dissertation an der Ludwig-Maximilians-Universität München, 2013

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-593-39961-4

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlags unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Copyright © 2013 Campus Verlag GmbH, Frankfurt am Main.

Umschlaggestaltung: Campus Verlag GmbH, Frankfurt am Main

Druck und Bindung: Beltz Bad Langensalza

Printed in Germany

Dieses Buch ist auch als E-Book erschienen.

www.campus.de

Inhalt

Danksagung	7
1 Einleitung	9
I Theoretische und methodologische Rahmung	
2 Kausalität: Subjektive oder objektive Wirklichkeit?	19
3 Menschenbild: Das Rationalitätsprinzip und seine Spezifizierungen	29
4 Methodologie: Experimente als ideale Evaluationsdesigns	75
II Lehrunabhängige Einflüsse	
5 Attraktivität und Notengebung: Theoretische Überlegungen	101
6 Attraktivität, Testzeitpunkt und LVE: Empirische Befunde	131
III Personenspezifische Beurteilungsstandards	
7 Beurteilerstandards bei LVE: Problematik und Nachweis	165
8 Standards bei der Gesamtbeurteilung: Erklärungen und Evidenz	189
9 Standards bei der Urteilsdifferenzierung: Erklärungen und Evidenz	215
IV Selbstselektion und studentischer Dropout	
10 Kurswahl, Besuch und Dropout: Theoretische Überlegungen	235
11 Determinanten des Absentismus: Empirische Befunde	257
12 Verzerrungen von Ratings und Rankings: Nachweis und Korrektur	279
13 Zusammenfassung und Fazit	297
Literatur	307
Anhang	347

Danksagung

Die vorliegende Arbeit ist die leicht modifizierte Fassung meiner Doktorarbeit, die 2012 an der Sozialwissenschaftlichen Fakultät der LMU München eingereicht wurde. Sie wäre ohne die Hilfe zahlreicher Personen nicht in der jetzigen Form entstanden.

Mein größter und persönlicher Dank gilt meinem Doktorvater Norman Braun, ohne den diese Arbeit nicht möglich gewesen wäre und dem dieses Buch gewidmet ist. In seiner Rolle als Studiendekan an der Sozialwissenschaftlichen Fakultät der LMU München hat er mein Vorhaben, die regulären Lehrveranstaltungsevaluationen genauer in den Blick zu nehmen und dafür zusätzliche Daten zu erheben, maßgeblich angeregt, tatkräftig unterstützt und mich in allen Phasen meiner Arbeit stets außerordentlich gefördert. Seine vielzähligen konstruktiven Anregungen haben entscheidend zur Verbesserung dieser Arbeit beigetragen, ohne jedoch der Verwirklichung meiner eigenen Ideen und Überlegungen Schranken zu setzen. In diesem Zusammenhang sind überdies auch Prof. Dr. Josef Brüderl und Prof. Dr. Bernhard Gill zu nennen, die als Zweit- und Drittgutachter ebenfalls wertvolle Anregungen zur Verbesserung dieser Arbeit gegeben haben. Ihnen gilt ebenso mein herzlicher Dank wie Prof. Dr. Martin Kocher, der sich ohne Zögern bereit erklärt hat, die Rolle des dritten Prüfers in meiner Disputation zu übernehmen.

Christian Ganser, Sebastian Haag und Marc Keuschnigg haben die Arbeit komplett gelesen und maßgeblich zu ihrer Verbesserung beigetragen. Eine Reihe substanzieller Ideen sowie wesentlicher inhaltlicher und sprachlicher Verbesserungen schulde ich Christiane Bozoyan, Werner Fröhlich und Patrick Riordan. Spontane Unterstützung bei so manchem Ärgernis habe ich zudem von Johannes Bauer, Roger Berger, Rudolf Farys, Jochen Groß, Roman Hoffmann, Fabian Kratz, Eva Negele, Judith Offerhaus, Johannes Schubert, Maren Träger und Dominikus Vogl erfahren. Qualitätssicherung und -verbesserung wurde überdies von Mitarbeiterinnen des Campus-Verlags, namentlich Julia Flechtner und Stefanie Evita Schaefer, betrieben.

Für ihre tatkräftige Hilfe bei der Durchführung der in dieser Arbeit präsentierten Studien bedanke ich mich auch bei allen Hilfskräften des Lehrstuhls von Norman Braun ganz herzlich. Benjamin Gedon und Edgar Treischl haben mich besonders bei der Erfassung und Aufbereitung der Lehrevaluationsdaten sehr tatkräftig unterstützt. Anja Hellmann war an der Durchführung der experimentellen Studie maßgeblich beteiligt und hat – ebenso wie Lukas Gernand – bei der Zusammenführung der Paneldaten geholfen.

Daneben haben mir die Teilnehmerinnen und Teilnehmer des alljährlich stattfindenden Seminars *Rational Choice Sociology* an der Venice International University mit hilfreichen Anregungen zur Seite gestanden und mir das Gefühl vermittelt, ein wichtiges Thema zu bearbeiten. Gleiches gilt für alle Teilnehmerinnen und Teilnehmer der Oberseminare am Kieler, Konstanzer, Leipziger und Münchner Institut für Soziologie. Ich danke insbesondere Prof. Dr. Roger Berger, Prof. Dr. Thomas Hinz, Prof. Dr. Monika Jungbauer-Gans und Prof. Dr. Thomas Voss für die Möglichkeit, in ihren Kolloquien vorzutragen.

Den Studierenden der Sozialwissenschaftlichen Fakultät der LMU München danke ich für ihre Bereitschaft, für die vorliegende Arbeit die eine oder andere zusätzliche Information für das Lehrevaluationspanel preiszugeben. Ebenso gilt mein Dank auch all denjenigen Dozierenden an der Sozialwissenschaftlichen Fakultät, die sich bereit erklärt haben, den zusätzlichen Aufwand in Kauf zu nehmen, bereits zu Beginn des Sommersemesters 2011 eine zusätzliche Evaluation in ihren Kursen durchzuführen.

Überdies ist darauf hinzuweisen, dass Teile dieser Arbeit weiter ausgearbeitete und modifizierte Fassungen von Aufsätzen darstellen, die in den Journalen *Evaluation Review* (Teil IV), *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (Teil II), *Zeitschrift für Evaluation* (Teil II) und *Zeitschrift für Theoretische Soziologie* (Teil I) erschienen sind (Wolbring 2010a, 2012, 2013; Wolbring/Hellmann 2011). Diese Abschnitte haben daher auch von Hinweisen im Zuge der Peer-Review-Verfahren profitiert.

Letztlich gilt mein Dank natürlich auch meiner Familie, die mich in allen Lebenslagen stets unterstützt und auch in schwierigen Zeiten an mich geglaubt hat. Für wichtige Unterstützung, positiven Rückhalt und Rücksichtnahme danke ich ihnen ebenso wie Martina. Nicht zuletzt aufgrund ihrer soziologischen Ausbildung verdanke ich ihr viele wertvolle inhaltliche Anregungen. Ohne sie wäre daher diese Dissertation in der vorliegenden Form nicht zustande gekommen.

München, im Juni 2013

Tobias Wolbring

1 Einleitung

In den vergangenen Jahren ist eine vermehrte Beschäftigung sozialwissenschaftlicher Forschungsarbeiten mit dem Thema Kausalität zu verzeichnen. Dies lässt sich etwa an einer Reihe neuerer Beiträge ablesen, die sich aus methodologischer, methodischer und statistischer Sicht mit Möglichkeiten und Grenzen des empirischen Nachweises von Ursache-Wirkungszusammenhängen (zum Beispiel Blossfeld 2009; Engelhardt 2005; Gangl 2010; Goldthorpe 2001; Heckman 2005; Imai et al. 2011; Morgan/Winship 2007; Sobel 2005; Winship/Sobel 2004), insbesondere mit Bezug zu evaluativen Fragestellungen (zum Beispiel Angrist/Pischke 2009; Blundell/Costa Dias 2009; Freedman 2006; Imbens/Wooldridge 2008; Schlotter et al. 2009; Schneider et al. 2007; Shadish/Cook 2009) befassen. Parallel dazu kann eine erneut aufwallende wissenschaftstheoretische Diskussionswelle zum Begriff der Kausalität und daran anknüpfende Anforderungen an die Gestaltung theoretischer Erklärungen beobachtet werden (zum Beispiel Bunge 2004; Hedström 2005; Mayntz 2004; Pearl 2000; Rubin 2005; Schmid 2006; Woodward 2003).

In der vorliegenden Arbeit werden diese Entwicklungen aufgegriffen und für die Messung der Qualität von Lehre nutzbar gemacht. Genauer sollen studentische Lehrveranstaltungsevaluationen (im Folgenden LVE), also die Bewertungen universitärer Lehre durch Studierende anhand weitgehend standardisierter Fragebögen, behandelt und Determinanten studentischen Urteilens herausgearbeitet werden. Es sollen dabei einerseits die Schwächen gängiger Evaluationsverfahren an Hochschulen, generelle Fallstricke bei der Qualitätsmessung mittels studentischer Bewertungen und Ansatzpunkte zu deren Behebung aufgezeigt werden. Andererseits sollen die Vorzüge sowohl experimenteller Forschungsdesigns und Vorher-Nachher-Messungen als auch fortgeschrittener statistischer Analyse- und Korrekturverfahren wie etwa Mehrebenenanalyse, Panelanalyse, Imputation und Gewichtung illustriert werden. Studentische LVE eignen sich dabei aus verschiedenen Gründen in besonderer Weise für eine solche Illustration.

Erstens ist man bei evaluativen Fragestellungen bekanntlich vorwiegend an dem Nachweis der Wirkung eines Programms, einer Intervention oder einer Maßnahme¹ interessiert (Wirkungsevaluation). Damit ist (meist implizit) die *kausale* Wirkung eines Eingriffs gemeint. Es soll also Wissen darüber gewonnen werden, ob durch die Implementierung einer Maßnahme intendierte Effekte und unintendierte Nebenwirkungen eingetreten sind. Bei studentischen LVE ist man dabei einerseits an dem Lernerfolg und Kompetenzerwerb durch das Treatment »Lehre« interessiert. Andererseits wird nicht nur die Wirkung der Maßnahme, sondern auch deren Implementierung anhand von Prozess- (zum Beispiel Erklärungen der Dozierenden, Veranstaltungstempo, Störungen) und Strukturindikatoren (Raumgröße, Textsammlung) evaluiert, wobei die Wahl des Bewertungsmaßstabs ohne weitere Vorgaben den Studierenden überlassen wird. Da an Hochschulen stets zahlreiche ähnliche und verschiedene Veranstaltungen evaluiert werden, bietet sich hier eine besonders günstige Gelegenheit, konfundierende Einflüsse ausfindig zu machen und naheliegende Alternativerklärungen für empirische Befunde durch die Bildung geeigneter Versuchs- und Kontrollgruppen auszuschließen.

Kausalanalysen können *zweitens* im Bereich der LVE eine direkte Praxisrelevanz entfalten, wenn sich mit ihrer Hilfe etwa zeigen lässt, dass schlichte statistische Zusammenhänge (zum Beispiel zwischen studentischer Bewertung und Lernerfolg) einer strengeren Prüfung nicht standhalten, externe Faktoren (wie Notengebung oder Attraktivität des Lehrpersonals) auf LVE-Bewertungen einwirken oder Kausalrichtungen anders gelagert sind als vermutet. Denn bei Evaluation steht in der Regel eine Erkenntnisfunktion (Stockmann/Meyer 2010) im Vordergrund. Ziel ist es also Programmwirkungen zu identifizieren. Gestützt auf die Resultate dieser wissenschaftlich fundierten Interventionsbewertung soll sodann entschieden werden, ob eine Maßnahme beibehalten, intensiviert, adjustiert oder unterlassen wird.

Dies hängt mit dem Wunsch nach gezielten planerischen Eingriffen in die Welt zusammen, die in modernen Gesellschaften zentrale Bedeutung erlangt haben (Beck 1986; Weber 1922). Evaluationen sind dabei Produkt dieses modernen Rationalisierungs- und Fortschrittsmythos, fördern diesen aber zugleich. Auch Schwarzs (2006) Charakterisierung von Evaluation als ein modernes Ritual weist in diese Richtung. Im Sinne Durkheims (1912) Ritualbegriff erfüllen Evaluationen aufgrund ihrer Verstetigung elementare gesellschaftliche Funktionen und tra-

1 Die Begriffe Programm, Maßnahme und Intervention werden, wenn nicht anders gekennzeichnet, synonym verwendet. Eine genauere definitorische Bestimmung dieser Begrifflichkeiten findet sich in dem Beitrag von Stockmann und Meyer (2010: 68ff.).

gen zu aktuellen Ökonomisierungstendenzen bei (Braun und Wolbring 2012). So wird mittels Evaluationen Exzellenz konstruiert (Münch 2007) und sozialer Status zugewiesen (Podolny 2005). Evaluationen stabilisieren damit moderne Gesellschaften an sich, da sie staatliches Handeln legitimieren sowie Zweifel an dessen Legitimität verarbeiten und zerstreuen (Meyer und Rowan 1977). Anregungen zur Verbesserungen bestehender Evaluationsverfahren wirken sich daher direkt auf die Erfüllung derartiger Funktionen aus und sind wichtig, um einer unhinterfragten Anwendung derartiger Methoden und damit der Publikation möglicherweise irreführender Lehrveranstaltungsrankings entgegenzuwirken.

Dieser Sachverhalt praktischer Bedeutsamkeit gilt *drittens* für LVE in verstärktem Maße, da an Hochschulen durch die Kopplung von Mittelzuweisungen an Evaluationsergebnisse individuelle und kollektive Leistungsanreize (für Dozierende, Fachbereiche, Fakultäten, Hochschulen) bestehen und noch verstärkt gesetzt werden sollen (vgl. zum Beispiel Anderson et al. 2011; Wissenschaftsrat 2008). Wenn finanzielle Entlohnungen und weitere Handlungsanreize von derartigen Messungen abhängen, sollte Klarheit darüber bestehen, welche Aussagen studentische Qualitätseinschätzungen erlauben und welche nicht. Schließlich sollten Einzelpersonen anhand meritokratischer Prinzipien entlohnt werden und nicht anhand von Kriterien, welche von zufälligen Schwankungen und systematischen Verzerrungen dominiert werden oder leicht manipulierbar sind. Die Frage nach der Datenqualität ist somit mit Pohlenz (2009) »als Schlüsselfrage der Qualitätssicherung von Lehre und Studium« zu betrachten. Neben der Validität der Messungen werden hierbei auch Aspekte der Fairness derartiger Bewertungsverfahren berührt. Würde sich etwa herausstellen, dass LVE von lehrunabhängigen Faktoren beeinflusst werden, die von den Dozierenden nur schwer zu beeinflussen sind, so wäre das ein Fairness-Problem (siehe auch Rindermann 2001). In solchen Fällen wäre an entsprechende Korrekturmaßnahmen der Qualitätsmessungen zu denken, um *ex post facto* diese Defizite der LVE zu beheben oder zumindest zu weitreichende Schlüsse aus LVE-Resultaten zu vermeiden.

Aspekte der Validität und Fairness hängen dabei *viertens* auch mit dem im Bereich studentischer Lehrvaluationen zu konstatierenden Mangel an ausgearbeiteten theoretischen Ansätzen und dem eklatanten Defizit in der Anwendung angemessener statistischer Verfahren zum Nachweis von Ursache-Wirkungszusammenhängen zusammen. So verwundert es nicht, dass die empirische Evidenz – trotz ausgesprochen umfangreicher und bis in die zwanziger Jahre des vergangenen Jahrhunderts zurückreichender Forschungsarbeiten zur methodischen Qualität studentischer Lehrveranstaltungsbewertungen (siehe zum Bei-

spiel Remmers/Brandenburg 1927)² – ein ausgesprochen uneinheitliches Bild ergibt. Die kritischen Stimmen bezüglich der Validität der Messungen sind daher keineswegs verstummt (Burzan/Jahnke 2010; Hinz 2012; Kulik 2001; Merritt 2008; Metje 2009; Onwuegbuzie et al. 2009; Pounder 2007). Auch wurde bisher keineswegs Einigkeit darüber erzielt, was überhaupt unter dem Konstrukt »gute Lehre« zu verstehen ist. Obwohl die vorliegende Arbeit diese lange zurückreichenden Debatten sicherlich nicht klären kann, trägt sie hoffentlich zumindest dazu bei, statistische Artefakte in bisherigen Studien zu identifizieren, methodische Fallstricke bei studentischen LVE aufzuzeigen und Alternativen zur bisherigen Praxis – etwa in Form möglicher Korrekturverfahren – anzubieten.

Schließlich erscheint das hier behandelte Thema auch aus einem *fünften* Grund fruchtbar: Anhand des eng begrenzten Themas studentischer LVE können auch allgemeine Einsichten darüber gewonnen werden, wie Menschen Entscheidungen treffen und wie sie sich bei Befragungen verhalten. In der Arbeit wird daher ganz im Sinne von Karl Poppers (1934) Kritischem Rationalismus der Anspruch vertreten, theoretische Erklärungsansätze harten empirischen Überprüfungen zu unterziehen. Die im folgenden präsentierten Analysen sind damit nicht nur für die Hochschulforschung von Interesse, sondern auch von sozialtheoretischer und methodischer Relevanz.

Die Arbeit gliedert sich im Weiteren in vier Teile. Zunächst werden in vier Kapiteln theoretische und methodologische Vorüberlegungen angestellt, sodann folgen drei Teile mit empirischen Anwendungen zu lehrunabhängigen Einflüssen, Urteilerstandards und Selektionseffekten.

In einem ersten Kapitel werden zwei gegenwärtig dominierende wissenschaftstheoretische Positionen zum Begriff der Kausalität kontrastierend gegenübergestellt. Wählt man eine radikal-konstruktivistische Perspektive, so werden kausale Ursache-Wirkungsbeziehungen als soziale Konstruktionen begriffen und es wird aus dieser Ausgangsprämisse gefolgert, die Vorstellung der Existenz von beobachterunabhängigen Gesetzmäßigkeiten sei überholt. Diese Position wird aus einer im Realismus und Kritischen Rationalismus verwurzelten Sichtweise kritisiert und es wird die gegenläufige Grundannahme formuliert, dass es kausale Regularitäten und damit auch kausale Mechanismen gibt, denen sich Forscher mittels systematischer Methoden des sequenziellen Versuchs und Irrtums annähern können. Theoretische Vorüberlegungen sind dabei neben geeigneten For-

2 Marsh (1987) berichtet, dass bis 1984 mehr als 1.400 Studien zu studentischen LVE in englischer Sprache veröffentlicht worden sind. Eine Recherche von Rindermann (2000) im deutsch- und englischsprachigen Raum für die Jahre 1967 bis Mai 2000 ergab über 1.500 Publikationen zu Fragestellungen der Lehrevaluation.

schungsdesigns und Analysemethoden zentral, da ein rein statistischer Nachweis kausaler Zusammenhänge nicht möglich ist.

Es wird daher im Anschluss im zweiten Kapitel ein mit dem Realismus konsistentes Menschenbild, das bereits Popper (1967) unter der Bezeichnung »situationsgerechtes Handeln« und »Rationalitätsprinzip« ausführlich beschrieben hat und das auch mit Max Webers (1922) Überlegungen zur Situationslogik große Ähnlichkeiten aufweist, eingehender erläutert und für die Erklärung studentischen Urteilsverhaltens näher spezifiziert. Nachdem herausgearbeitet wurde, weshalb die Annahme rationaler Optimierung für das studentische Urteilsverhalten möglicherweise unangemessen ist, werden Theorien zum Satisficing, zur Informationsverarbeitung und Urteilsheuristiken sowie zur Wirkung sozialer Normen eingehender behandelt. Da das sich daraus ergebende Menschenbild den theoretischen Rahmen für die weiteren Kapitel liefert, werden nachfolgend einige allgemeine forschungsleitende Hypothesen im Hinblick auf studentische LVE formuliert, die in späteren Teilen der Arbeit dann genauer spezifiziert und überprüft werden.

In einem dritten Kapitel wird sodann eine vielfältige statistische Verfahren der Kausalanalyse umfassende Perspektive vorgestellt. Anhand von Rubins Kausalitätsmodell (Imbens/Rubin 2008; Rubin 1974) wird das grundsätzliche Vorgehen bei der Bestimmung eines Treatment-Effekts dargestellt.³ Dabei wird dafür plädiert, sich dem Vorbild experimenteller Designs auch bei Befragungsstudien durch die Verwendung entsprechender statistischer Methoden anzunähern. Denn auch bei dieser Designform kommt der von Holland (1986: 947) als »fundamental problem of causal inference« bezeichneten Schwierigkeit kontrafaktischer Schlüsse vorrangige Bedeutung zu, welche im experimentellen Kontext durch Randomisierung, Verwendung von Kontrollgruppen und bewusste Manipulation des Stimulus in überzeugender Weise gelöst wird. Es sollte daher auch bei Beobachtungsstudien angestrebt werden, adäquate Versuchs- und Kontrollgruppen zu bilden, über die interessierenden Effekte hinausgehende, einflussreiche Faktoren zu kontrollieren und eine klare zeitliche Abfolge von Ursachen und Wirkungen herzustellen.⁴ Aufgrund dieser starken Orientierung am experimentellen Vorgehen werden anschließend anhand einer Unterscheidung ver-

3 Ein ähnliches Modell wurde bereits 1923 von Neyman vorgeschlagen, weshalb streng genommen vom Neyman-Rubin-Kausalitätsmodell gesprochen werden sollte.

4 Wie Cummings (2006) betont, wurde die kontrafaktische Natur evaluativer Fragestellung in der Literatur bisher weitgehend vernachlässigt, obwohl sie interessante Ansatzpunkte für den Nachweis von Programmwirkungen liefert. Insbesondere neuere ökonomische Arbeiten zur Programmevaluation heben jedoch diesen Aspekt und die sich aus dieser Perspektive ergebenden Implikationen hervor (zum Beispiel Imbens/Wooldridge 2008; Schlotter et al. 2009).

schiedener Formen von Validität häufige Fehlerquellen bei dieser Designform im Allgemeinen und bei studentischen LVE im Besonderen diskutiert.

Die darauf folgenden drei Teile sind als Anwendungen dieses theoretischen und methodologischen Rahmens zu verstehen. Abbildung 1.1 gibt einen Überblick über die verschiedenen empirischen Anwendungen, bei denen mittels unterschiedlicher methodischer Herangehensweisen versucht wird, dem formulierten heuristischen Ideal möglichst nahe zu kommen und Probleme bei der Messung der Lehrleistung mittels LVE zu identifizieren.

Abbildung 1.1: Weitere Schwerpunkte der Arbeit

Teil II Sachfremde Einflüsse	Teil III Urteilerstandards	Teil IV Selektionseffekte
<ul style="list-style-type: none"> • Attraktivität des Lehrpersonals • Geschlechtseffekte • Benotung, insbesondere Reziprozität • Wechselwirkung dieser drei Einflussfaktoren 	<ul style="list-style-type: none"> • Werden unterschiedliche Maßstäbe angelegt? • Unterscheiden sich Studierende in der Urteilsstrenge und -differenziertheit? • Wodurch lassen sich personenspezifische Standards erklären? 	<ul style="list-style-type: none"> • Ausmaß der Selektion und des Selektionsbias • Erklärungen für den studentischen Absentismus • Wirkung des Selektionsprozesses auf Veranstaltungsratings und -rankings

Zunächst werden mit der Attraktivität des Lehrpersonals, dem Studierenden- und Dozierendengeschlecht sowie der Notengebung lehrunabhängige Einflüsse bei studentischen LVE betrachtet (Teil II). In einer ersten Teilstudie wird dabei mittels Mehrebenenanalysen der Einfluss sachfremder Faktoren, in diesem Fall die Attraktivität des Lehrpersonals, auf studentische LVE herausgearbeitet. Da die Separierung von Produktivitätsunterschieden und Diskriminierung – beides Effekte die theoretisch vorstellbar sind – mit Hilfe der statistischen Analysen auf Grundlage der verwendeten prozessproduzierten Daten nicht eindeutig gelingt (zu den Schwierigkeiten bei dem Versuch der Separierung siehe Wolbring 2010b), werden anschließend in einer zweiten Teilstudie die Resultate einer experimentellen Versuchsanordnung präsentiert. Dabei wurde versucht, konfundierende Faktoren wie die Qualität der Lehre so weit wie möglich konstant zu halten und nur die interessierende Attraktivität des Lehrpersonals zu variieren. Dabei wird auch die Bedeutung der Benotung bei der Bewertung von Lehrleistung und Lehrqualität herausgearbeitet. Von besonderem Interesse sind in diesem Zusammenhang Theorien der Reziprozität sowie die Wechselwirkung dieses Mechanismus mit Geschlecht und Attraktivität.

In einer zweiten Anwendung werden Paneldaten zur studentischen Lehrevaluation herangezogen, sodass studentisches Urteilsverhalten systematisch über die Zeit beobachtet werden kann (Teil III). Zur Auswertung der Daten werden Verfahren der Längsschnittdatenanalyse eingesetzt, um zu prüfen, inwieweit personenspezifische Beurteilerstandards und Urteilsheuristiken bei der Bewertung eine Rolle spielen und damit den Vergleich der studentischen Qualitätseinschätzungen verschiedener Lehrveranstaltungen erschweren. Die Paneldaten ermöglichen zudem Veränderungen des Urteilsverhaltens über die Zeit zu beobachten. Darüber hinaus werden auch Hypothesen zu Anspruchsniveaus, Motivation und Anpassungsanpassung getestet.

Schließlich wird die Problematik der Selbstselektion bei Kausalanalysen behandelt (Teil IV). In Diekmanns (2007) Lehrbuch zu den Methoden und Techniken der empirischen Sozialforschung wird die Problematik systematischer Stichprobenausfälle u.a. anhand des Beispiels zum Zeitpunkt der Evaluation abwesender, unzufriedener Schüler vermittelt. Gemeinhin wird dabei angenommen, es komme aufgrund der Nichtbeachtung dieser speziellen Gruppe zu einer generellen Überschätzung der Unterrichtsqualität. Ähnliche Kritik wurde gegenüber studentischen LVE geäußert: Die Lehrqualität werde aufgrund der Abwesenheit der von Esser (1995, 1996a, 1997) als »No-Shows« bezeichneten Studierenden, welche Veranstaltungen nicht oder nur unregelmäßig besuchen, verzerrt wiedergegeben (vgl. auch Kromrey 1994a, 1994b, 2001). Zudem seien die Bewertungen verschiedener Veranstaltungen nicht ohne Weiteres vergleichbar, wenn Dropout-Quoten veranstaltungsspezifisch variieren.

Einen systematischen, empirischen Nachweis zur Wirkung dieser (aller Wahrscheinlichkeit nach selektiven) Ausfälle auf die durchschnittlichen Veranstaltungsbewertungen und darauf aufbauende Rankings ist die Forschung jedoch bisher schuldig geblieben. Ein Design mit Evaluationen zu zwei Messzeitpunkten bietet hier eine günstige Gelegenheit, genauer zu untersuchen, inwieweit diese bisher ungeprüften Annahmen und Vermutungen zutreffen. Dabei stellt sich in besonderer Weise das Problem kontrafaktischer Schlüsse, denn die zentrale Frage lautet: Wie hätten die zum Zeitpunkt der Evaluationen abwesenden Studierenden die Lehrveranstaltung beurteilt, wenn Sie anwesend gewesen wären? Es wird versucht, sich einer wahren Antwort auf diese Frage mittels Gewichtung der Studierendenurteile auf Grundlage von Informationen zum Veranstaltungsbesuch sowie mittels (einfacher und multipler) Imputation anzunähern. Dabei werden auch verschiedene praktische Lösungsmöglichkeiten zur Behebung von Selektionsbias diskutiert und empirisch geprüft.

Damit wird in den drei skizzierten Teilen ein weites Feld an Methoden und Techniken zur Bestimmung kausaler Zusammenhänge, die sowohl für Querschnitt- (experimentelle Designs, Mehrebenenanalyse) als auch Längsschnittdaten (Panelanalyse) anwendbar sind, und auch an Verfahren zur Behebung der Probleme systematisch fehlender Daten (Gewichtung, Imputation) aufgespannt. Bevor jedoch in den einzelnen Teilstudien deren theoretische Darstellung und empirische Anwendung erfolgen kann, sind die Bestimmung der in dieser Arbeit vertretenen methodologischen Position zum Begriff der Kausalität (2) sowie die Entwicklung einer einheitlichen theoretischen (3) und methodologischen (4) Rahmung der folgenden Teilstudien zunächst sinnvoll.

Teil I
Theoretische und
methodologische Rahmung

2 Kausalität: Subjektive oder objektive Wirklichkeit?

Philosophische Überlegungen zum Begriff der Kausalität lassen sich bis in die Antike zurückverfolgen. Bekannte Denker wie Platon, Aristoteles, Bacon, Descartes, Hume und Mill, um nur einige renommierte Vertreter zu nennen, haben sich intensiv mit dem Zusammenhang von Ursache und Wirkung beschäftigt und damit weitere Arbeiten in diesem Bereich angestoßen und maßgeblich beeinflusst.¹ Im Folgenden sollen zwei philosophische Grundpositionen kontrastierend gegenübergestellt werden, welche die gegenwärtige Debatte zum Nachweis von Ursache und Wirkung in den Sozialwissenschaften maßgeblich prägen² und auch im Bereich der Evaluation die dominierenden Perspektiven repräsentieren.³ Es handelt sich bei diesen beiden grundverschiedenen Sichtweisen auf die Welt und die Rolle, welche wissenschaftliche Forschung darin zu spielen hat, um den *Konstruktivismus* und den *wissenschaftlichen Realismus*.

Wie Popper (1970) hierzu früh bemerkt hat, lässt sich die Position des Idealismus und daher auch die damit verwandte Position des Konstruktivismus ebenso wie die des wissenschaftlichen Realismus nicht widerlegen. Die jeweiligen philosophischen Grundannahmen führen zu in sich logisch konsistenten Aussagensystemen, für die es keine empirisch eindeutigen Falsifikationen geben kann. Jedoch gibt es praktische Argumente, die für die Wahl der einen oder der anderen Position sprechen. Es erscheint daher sinnvoll, sich zunächst die Grundannahmen und Hauptaussagen beider Ansätze zu vergegenwärtigen und mit der Thematik

1 Einen guten, kompakten Überblick zu diesen historischen Entwicklungslinien und philosophischen Denktraditionen gibt Engelhardt (2005: Kap. 2).

2 Dass diese beiden Denktraditionen heutzutage zentrales Gewicht besitzen, zeigt sich anhand intensiv geführter Debatten über die Grundausrichtung und Zielsetzung von Forschung. Eine Diskussion über die Rolle von Theorie in der Soziologie, in der diese wissenschaftstheoretischen Gegensätze deutlich werden, findet sich in dem Beitrag von Norman Braun (2008) sowie der daran anschließenden Debatte (Braun 2009a; Rehberg 2009; Schmid 2009) in der soziologischen Fachzeitschrift *Soziale Welt*.

3 Stockmann und Meyer (2010) sprechen in einem einführenden Lehrbuch zur Evaluation gar vom »kalten Krieg der Paradigmen« (S.84). Sie meinen zwar eine Tendenz zur Annäherung erkennen zu können, betonen jedoch gleichzeitig die Unvereinbarkeit der unterschiedlichen Ausgangsprämissen.

studentischer LVE in Beziehung zu setzen. Da es schwierig ist, Positionen ganzer Gruppen so zusammenzufassen, dass die Darstellung allen Vertretern und den vielfältigen Ausformungen der Ansätze gerecht wird, wird versucht, nur die in diesen Bereichen weitgehend geteilten Prämissen und den gemeinsamen Kern zentraler Implikationen wiederzugeben.

2.1 Kausalität als soziales Konstrukt

*Konstruktivistische Ansätze*⁴ zeichnen sich durch folgende Grundannahmen aus:

- Die Welt ist durch ein hohes Maß an Komplexität gekennzeichnet. Die Möglichkeit, mittels wissenschaftlicher und auch anderer menschlicher Verfahren *objektive* Erkenntnisse über weltliche Zusammenhänge zu gewinnen, wird bestritten. Weitergehend werden in vielen Spielarten des Konstruktivismus derartige faktische, vom Beobachter unabhängig gegebene Regularitäten selbst als nonexistent oder irrelevant betrachtet.
- Dies hängt damit zusammen, dass der historische und kulturelle Kontext jede Art sozialer Praxis maßgeblich formt. Hierbei spielen unter anderem Prozesse der Institutionalisierung, Objektivierung, Legitimierung, Typisierung und Habitualisierung eine wesentliche Rolle (vgl. Berger/Luckmann 1966; Knorr-Cetina 1989). »Realität« ist somit stets sozial konstruiert und von dem jeweiligen Beobachter abhängig.
- Die Wahrheit, Rationalität und Gültigkeit von Aussagen sind dementsprechend nicht *a priori* gegeben, sondern werden in der sozialen Praxis erzeugt und ausgehandelt. Dabei sind Aspekte wie Macht, Autorität, Ideologie und Rhetorik von vorgelagerter Bedeutung, da sie diese diskursiven Aushandlungsprozesse bestimmen.

Als zentrale Implikation für wissenschaftliche Forschung ergibt sich aus diesen Prämissen: Das Streben nach möglichst allgemeingültigen Aussagen, Erklärungen und Theorien und die Suche nach objektivem Wissen und absoluter Wahrheit ist für (radikale) Konstruktivisten eine Farce. Wissenschaftliche Theorien sind nicht als Annäherungen an die Wahrheit, sondern als Narrationen oder rationalisierte

⁴ Nach Bunge (1993) ist der Konstruktivismus eine Variante des Antirealismus. Andere Ausformungen sind etwa Subjektivismus, Konventionalismus, Relativismus und Hermeneutik. Verschiedene Spielarten des Konstruktivismus werden beispielsweise von Knorr-Cetina (1989) und der Bochumer Arbeitsgruppe für Sozialen Konstruktivismus und Wirklichkeitsprüfung (2000) dargestellt.

Mythen, wissenschaftliche Debatten als rhetorische Spiele aufzufassen. Sie sollten deshalb gegenüber anderen alltagsweltlichen Überlegungen und Diskussionen bezüglich des Gewinns von Erkenntnis keine priorisierte Stellung einnehmen, sodass Wissenschaft mit anderen Akteuren um die gesellschaftliche Deutungshoheit ringt. Man könnte daher argumentieren, dass der Konstruktivismus nicht mehr und nicht weniger als das Fundament der modernen Wissenschaft mit ihrer Methode systematischer Eingriffe und empirischer Beobachtungen sowie dem Instrumentarium der Logik und Mathematik in Frage stellt und »eine der großen Metaerzählungen der Moderne« (Reich 2002: 100), nämlich die Objektivität der wissenschaftlichen Methode entzaubert.

Weitere Einsichten zu einer konstruktivistischen Sicht auf die Welt ergeben sich exemplarisch anhand Niklas Luhmanns Beiträgen zur Kausalität. So schreibt er aus der Warte eines Beobachters zweiter Ordnung: »Nach jahrzehntelangen Forschungen über Kausalattribution und Wahrnehmung kausaler Beziehungen kann man nicht mehr davon ausgehen, daß Beziehungen zwischen Ursachen und Wirkungen objektive Sachverhalte der Welt seien, über die dann wahre bzw. un-wahre Urteile möglich sind« (Luhmann 1995: 7). Aus seiner systemtheoretischen Perspektive ist Kausalität eine soziale Konstruktion, die von Beobachtern erster Ordnung (ebenso wie die Begriffe Akteur und Handlung) zugeschrieben und zur Strukturierung und Beschreibung der Welt genutzt wird.

Kausalität ist dabei einerseits ein Kommunikationsmedium, mit dessen Hilfe Ursachen und Wirkungen in den Blick geraten, ohne dass jedoch bereits eine Festlegung auf konkrete Ursachen und Wirkungen erfolgt. Da diese Spezifizierung auf bestimmte Zusammenhänge mittels des Mediums ausbleibt, sind Ursachen und Wirkungen weiterhin kontingent. Kausalzusammenhänge könnten also so oder auch anders sein, was auch bedeutet: »Sowohl in Richtung Ursachen als auch in Richtung Wirkungen führt Kausalität in Endlos Horizonte« (ebd.). Andererseits muss nach Luhmann mit dem Medium die konkrete Beobachtung und Beschreibung von Regularitäten, die sogenannte Form, stets mitgedacht werden. »Das Medium erscheint [...] nur in seinen jeweils realisierten Formen. Als solches bleibt es unsichtbar. Es wird nur dadurch reproduziert, daß laufend Formen gebildet werden. Würde das (aus welchen Gründen immer) nicht geschehen, gäbe es auch keine Kausalität.« (ebd.) Luhmann erscheint es daher besonders fruchtbar, Beobachtungen erster Ordnung, in denen Kausalitäten kommunikativ hergestellt werden, unter dem Blickwinkel einer Theorie höherer Ordnung zu analysieren.

Seine Variante der Systemtheorie öffnet damit, ebenso wie andere konstruktivistische Ansätze, den Blick auf die Rolle von Diskursen und die dabei vertretenen Positionen bei der Objektivierung von Wissen und Tatsachen. Das Erkenntnisinteresse verschiebt sich von der Suche nach objektiven Ursache-Wirkungszusammenhängen hin zu kommunikativen und interaktionalen Aushandlungsprozessen und deren Produkten. Im Hinblick auf die Thematik sozialwissenschaftlicher Evaluationen lässt sich mit Hilfe des konstruktivistischen Instrumentariums eine zunehmende Popularisierung des Qualitätsbegriffs in zeitgenössischen Diskursen in verschiedensten Bereichen diagnostizieren (zum Beispiel Münch 2009). Die Verbreitung einer »Qualitätssicherungsrhetorik« und eines »New Public Management« korrespondiert dabei mit der allgemeineren Entwicklung einer Ökonomisierung und hängt mit anderen gesellschaftlichen Prozessen (zum Beispiel Differenzierung, Globalisierung, Rationalisierung) eng zusammen (siehe für Details Braun/Wolbring 2012; Nassehi 2012).

Ausdruck findet die Ausbreitung der ökonomischen Denklogik somit nicht nur in der Privatwirtschaft. Ökonomische Semantiken, die sich an Begriffen wie Qualitätssicherung, Leistungsmessung und Rechenschaftslegung erkennen lassen, werden auch in öffentlichen Domänen genutzt, um die Verteilung knapper Mittel zu rationalisieren. Das Medium Geld gewinnt damit in nicht-ökonomischen Bereichen zunehmend an Einfluss. Kennziffern können hierbei eine Realität *sui generis* erzeugen, wenn Rangordnungen erstellt und leistungsabhängige, monetäre Anreize (zum Beispiel Leistungszulagen, Mittelsteuerung) daran geknüpft werden. Die entsprechenden Konstruktionen von Qualität werden auf diese Weise nämlich als faktische Gegebenheiten behandelt und entwickeln dadurch unvorhersehbare Eigendynamiken. Insbesondere werden auch im Hochschulwesen die Qualitätsmessungen häufig bei der Erstellung von Rankings und der Mittelvergabe unhinterfragt mit der tatsächlichen Leistung gleichgesetzt, wodurch Lehr- und Forschungsqualität konstruiert und objektiviert werden (zum Beispiel Gläser 2006; Münch 2007).

Diese Einsichten zur Verbreitung und Wirkmächtigkeit sozialer Qualitätskonstruktionen, die auch mit einem realistischen Wissenschaftsverständnis gut vereinbar sind, werden im weiteren Verlauf der Arbeit berücksichtigt, wenn die Folgen von Problemen bei der Erfassung von Qualitätsindikatoren im Vordergrund stehen. Denkt man die Überlegungen zur Non-Existenz einer beobachterunabhängigen Realität im Sinne eines radikalen Konstruktivismus (z.B. Gergen 1988) jedoch konsequent zu Ende, so ergeben sich keine (praktischen und politischen) Ansatzpunkte zur Verbesserung der Lehrleistung und deren Messung mittels LVE.

Fokussiert man sich alleine darauf, dass Ideen die wahrgenommene Realität formen, so ist es aufgrund des Fehlens objektiver Kriterien streng genommen nicht mehr möglich, präzise Aussagen über die Wirkungen praktischer Maßnahmen zu treffen. Dies steht im Gegensatz zur Zielsetzung dieser Arbeit, methodische Fallstricke bei der Evaluation der Maßnahme »Lehre« aufzuzeigen. Bunge (1991: 551) spricht ob dieser Praxisferne radikal-konstruktivistischer Positionen von einem »deliberate neglect of the »technical« aspects of the research process, that is the problems, hypotheses, arguments, experimental designs, and measurements that accompany the exchanges of views, plans, and findings among the members of research team(s).« Gerade die bei der LVE auftretenden praktischen Probleme sollen jedoch im Weiteren im Vordergrund stehen, wobei eine Annäherung an die Wahrheit durch die Methode des Versuchs und Irrtums für möglich gehalten und das Aufdecken von Ursache-Wirkungszusammenhängen angestrebt wird. Die Suche und Behebung methodischer Fallstricke setzt ja gerade voraus, dass es einen »wahren« Messwert gibt und dass sich die Validität von dessen Messung verbessern lässt. Üblicherweise wird ein solches normalwissenschaftliches Vorgehen mit einem realistischen Wissenschaftsverständnis begründet, weshalb eine nähere Auseinandersetzung mit den Grundannahmen und Hauptaussagen des wissenschaftlichen Realismus sinnvoll erscheint.

2.2 Kausalität als reales Faktum

Ein mit dem Vorhaben der kausalen Maßnahmenbewertung deutlich kohärenteres Weltbild bieten der wissenschaftliche Realismus und dessen wohl bekannteste Variante, der Kritische Realismus des Kritischen Rationalismus nach Popper (1970), an. Grundprämissen des *Realismus* und deren Implikationen sind nach Braun (2008; siehe auch Bunge 1993; Popper 1970):

- »Materielle Objekte [...], natürliche Ereignisse [...] und Abläufe [...] sowie soziale Phänomene [...] existieren prinzipiell autonom von menschlichen Sinneserfahrungen« (Braun 2008: 377). Es gibt eine einzige Realität, die unabhängig davon ist, wie über sie kommuniziert und diskutiert wird.
- Menschen können vor dem Hintergrund soziokulturell vorgeformter Bezüge Regularitäten erkennen und verstehen. Es handelt sich dabei um Annäherungen an die objektive bzw. absolute Wahrheit im Sinne verschiedener Grade der Übereinstimmung mit den Tatsachen (Korrespondenztheorie der Wahrheit) (vgl. Popper 1960a).

- Es wird zwischen Erscheinung und Wirklichkeit unterschieden (vgl. Popper 1970). Konzepte und Theorien werden von den Menschen nach den Regeln der Vernunft konstruiert, um Zusammenhänge aufzudecken und zu begreifen.
- Dabei ist es zweckmäßig, verfügbares Wissen unter Berücksichtigung der situativen Gegebenheiten zu nutzen. Zur Entstehung und Veränderung dieses Wissens kommt es durch Wahrnehmung der Realität und durch Prozesse des Lernens. Die Notwendigkeit der empirischen Überprüfung des Wissens anhand der Wirklichkeit trägt zum Erkenntnisfortschritt bei und verdeutlicht die Vorläufigkeit allen Wissens.

Der von Karl Popper geprägte Kritische Rationalismus stimmt in wesentlichen Punkten mit diesen Grundprämissen überein. Poppers (1953) Ausgangspunkt ist dabei David Humes (1739) Einsicht, dass aufgrund der Endlichkeit der Zahl wissenschaftlicher Beobachtungen ein Wahrheitsbeweis von All-Aussagen, welche sich auf einen unbeschränkten Geltungsbereich beziehen, nicht möglich ist (Induktionsproblem). Während Hume Induktion trotz des beschriebenen Problems jedoch als legitimes Verfahren zur Generierung wissenschaftlicher Erkenntnis betrachtet, wendet Popper sich von der induktiven Vorgehensweise ab und betont die Rolle der Methode des Versuchs und Irrtums im wissenschaftlichen und auch alltäglichen Erkenntnisprozess. Die Wahrheit wissenschaftlicher Theorien kann auf diesem Wege zwar nicht nachgewiesen werden. Sollten aber die (möglichst kühnen, d.h. informationshaltigen) Hypothesen einer Theorie (wiederholt) strenger Kritik und harten empirischen Prüfungsversuchen nicht standhalten, so gelten sie als falsifiziert. Andernfalls bezeichnet man Theorien als bewährt oder vorläufig bestätigt, nicht aber als verifiziert.

Das Popper'sche Falsifikationsprinzip ist folglich genuin realistischer Prägung, denn mit ihm wird eine objektive Welt als gegeben vorausgesetzt, an der wissenschaftliche Aussagen scheitern können. Als Demarkationslinie zwischen empirischer Wissenschaft und Nicht-Wissenschaft, die jedoch nicht immer eine trennscharfe Unterscheidung ermöglicht, gilt dementsprechend der empirische Gehalt einer Aussage sowie die Bereitschaft, diese der Gefahr des Scheiterns auszusetzen (vgl. Popper 1974). Dabei ist zu beachten: Je informationshaltiger eine Aussage ist, d.h. je größer die Menge der von ihr ausgeschlossenen Sätze, desto größer ist die Erkenntnis, welche bei einer Prüfung gewonnen wird.

Allerdings werden anhand empirischer Prüfversuche in der Regel nicht einzelne Annahmen, sondern ganze theoretische Aussagensysteme auf ihre Gültigkeit getestet (vgl. Popper 1974: 109), sodass im Falle einer Falsifikation unklar bleibt, welche Prämisse(n) für das Scheitern verantwortlich zu machen ist (sind).

Ebenso lässt eine vorläufige empirische Bestätigung keinen Schluss auf die Gültigkeit der Annahmen zu, da sich einerseits auch aus falschen Ausgangsprämissen korrekte Folgerungen ergeben können und andererseits in Zukunft immer auch gegenläufige Evidenz beobachtet werden kann. Die beiden Schwierigkeiten bei der Falsifikation und Verifikation von Annahmen werden mit Blick auf die mit dem Kritischen Rationalismus überaus kompatiblen Überlegungen von Hempel und Oppenheim (1948) zur Logik wissenschaftlicher Erklärungen und daran anschließende Arbeiten deutlich.

2.3 Hempel-Oppenheim- und mechanismische Erklärungen

Nach dem deduktiv-nomologischen Erklärungsmodell wird der Eintritt eines Ereignisses oder dessen Ausbleiben (Explanandum) mittels Randbedingungen und mindestens einer allgemeinen Gesetzmäßigkeit (Explanans) erklärt. Bedingungen für die Gültigkeit einer solchen Erklärung sind dabei, dass das Explanandum aus dem Explanans *erstens* logisch korrekt abgeleitet wird, *zweitens* mindestens eine der nomologischen Hypothesen empirischen Gehalt besitzt und *drittens* die Elemente des Explanans wahr bzw. empirisch gut bewährt sind. Abbildung 2.1 gibt ein Beispiel.

Abbildung 2.1: DN-Erklärung für reziproke Bewertungen bei LVE

<i>Explanans:</i>	(1) Studierender X erhält von Dozentin Y eine schlechte Note.
	(2) Wenn Studierende eine schlechte Note von einer Lehrperson erhalten, dann evaluieren sie diese ebenfalls schlecht.
<hr/>	
<i>Explanandum:</i>	(3) Studierender X evaluiert Dozentin Y schlecht.

Der singuläre Satz »Studierender X evaluiert Dozentin Y schlecht.« wird in diesem Fall durch ein allgemeines Gesetz zur negativen Reziprozität bei LVE sowie die Antezedensbedingung einer erfolgten schlechten Benotung erklärt (siehe für eine empirische Prüfung dieser Hypothese Kapitel 6). Es ist klar, dass deterministische Zusammenhänge, wie sie in diesem Beispiel unterstellt wurden, die meisten sozialen Regularitäten nur unzutreffend charakterisieren. Probabilistische Hypothesen beschreiben demgegenüber soziale Phänomene häufig treffender, was zu einer Modifikation des DN-Schemas geführt hat. Bei diesem induktiv-statistischen Erklärungsmodell wird der Gewinn an Realitätsnähe jedoch mit einem nicht

unerheblichen Preis bezahlt. Denn man verzichtet auf die zwingende logische Ableitbarkeit des Explanandums und damit auf die Möglichkeit punktgenaue Prognosen für zukünftige einzelne Beobachtungen zu generieren. So würde das Gesetz »Wenn Studierende eine schlechte Note von einer Lehrperson erhalten, dann steigt die Wahrscheinlichkeit, dass sie diese Person schlecht evaluieren.« wohl einen Gewinn an Realitätsnähe darstellen. Jedoch nimmt damit auch der Informationsgehalt beträchtlich ab. Denn während zuvor bereits eine einzige Beobachtung, bei welcher schlechte Noten mit guten Evaluationen korrespondieren, zur Falsifikation des Gesetzes hinreichend gewesen wäre, kann dies bei probabilistischen Aussagen erst anhand einer größeren Menge von Beobachtungen und den daraus resultierenden statistischen Durchschnitts geschehen.

In letzter Zeit kommen, angesichts der genannten und weiterer Kritikpunkte (siehe zum Beispiel Salmon 1990; Woodward 2003), vermehrt Forderungen nach der Verwendung eines dem Hempel-Oppenheim-Schema sehr ähnlichen Modells auf, das Erklärungen über soziale Mechanismen liefert (vgl. zum Beispiel Bunge 2004; Goldthorpe 2001; Hedström 2005; Maurer/Schmid 2010; Mayntz 2004; Schmid 2006).⁵ Generell lassen sich mechanismische Erklärungen dabei nach Hedström (2005: 25) folgendermaßen charakterisieren und definieren:⁶

»Mechanisms can be said to consist of *entities* (with their properties) and the *activities* that these entities engage in, either by themselves or in concert with other entities. These activities bring about change, and the type of change brought about depends upon the properties of the entities and the way in which they are linked to one another. A social mechanism, as here defined, describes a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome.«

Vom Theoretiker wird gefordert, generative Prozesse des interessierenden Tatbestandes offenzulegen, indem er diejenigen Aktivitäten von zwei oder mehr Einheiten angibt, welche Veränderungen ursächlich hervorbringen. In den Sozialwissenschaften sind etwa rationale Imitation, Vakanzketten und selbst-erfüllende Prophezeiungen verbreitete Mechanismen (für eine Zusammenschau siehe Hedström 2009). Wie Opp (2005b: 166) herausarbeitet, ist somit die Minimalvoraussetzung für das Vorliegen einer Mechanismus-basierten Erklärung, dass eine Regelmäßigkeit, meist der Zusammenhang zwischen zwei Variablen, durch

5 Der Vorschlag, Erklärungen spezifische Mechanismen zugrunde zu legen, ist keineswegs neu. Mit Gerring (2010) und Opp (2005b) kann man daher kritisch fragen, ob nicht in jeder theoretischen Erklärung auf bestimmte Wirkungsmechanismen verwiesen wird.

6 Wie so oft, findet sich auch zu diesem Begriff eine Vielzahl verschiedener Definitionen. Eine gute, kompakte Übersicht zu verschiedenen Definitionsversuchen geben Gerring (2008), Hedström (2005: 25) sowie Hedström und Bearman (2009: 6). Ausführlich werden unterschiedliche philosophische und soziologische Ansätze von Schmid (2006) behandelt.

mindestens einen dritten, intervenierenden Faktor aufgeklärt wird. In der Regel, aber nicht notwendigerweise wird die Analyse hierzu auf eine niedrigere Ebene verlagert und auf tieferliegende Mechanismen verwiesen (vgl. Stinchcombe 1991).

Das Öffnen dieser Black Box führt dabei zu einer tieferen Erklärung eines Phänomens und ist daher wünschenswert. Mechanismus-basierte Erklärungen liefern nämlich in Ergänzung zu Sätzen der Art » X beeinflusst Y « ein Narrativ, über welche Wege X auf Y wirkt. Jedoch kann es zu einem infiniten Regress kommen, wenn versucht wird, für jede Regularität auf einer bestimmten Ebene weitere Mechanismen aufzudecken, welche auf einer tieferliegenden Ebene zu verorten sind. Jede Erklärung enthält daher notwendigerweise eine Black Box, die unerklärt bleiben muss. Die Entscheidung, wann eine Erklärung als hinreichend betrachtet wird, hängt dabei von disziplinspezifischen bzw. paradigmatischen Übereinkünften ab. Diese bezeichnet Hedström (2005) in Anlehnung an Miller (1987) als »stopping rules«, da sie angeben, wann die Suche nach Mechanismen zu einem Ende kommen soll.

Je nach theoretischer Perspektive, von der aus ein Problem betrachtet wird, bedarf es demnach einer unterschiedlichen Tiefe der jeweiligen Erklärung. Für die Soziologie empfiehlt Hedström (2005: 28f.), wie viele Andere vor ihm (zum Beispiel Boudon 1980; Coleman 1990; Esser 1999), handlungstheoretische und intentionale Aspekte als Kern mechanistischer Erklärung zu betrachten. Soziale Phänomene sollen also über individuelle Entscheidungen erklärt werden. Während also das analytische Primat in der Soziologie auf der Ebene sozialer Systeme liegt, erscheint es fruchtbar, das theoretische Primat auf die Individualebene zu legen (vgl. Lindenberg 1991) und Erklärungen als Mehrebenenerklärungen anzulegen (Coleman 1990). Dies liegt schon alleine deshalb nahe, da es keine direkten kausalen Verbindungen, sondern nur Korrelationen zwischen zwei Makro-Variablen geben kann (siehe zum Beispiel Opp 2012).⁷

Die methodologische Regel der Mikrofundierung impliziert jedoch keineswegs, strukturelle Gegebenheiten gänzlich auszublenden. Ganz im Gegenteil plädiert Hedström in seinen Arbeiten für einen Strukturindividualismus – eine methodologische Doktrin, nach der alle sozialen Tatsachen, sowohl ihre Struktur als auch ihr Wandel, mit dem Begriffsinstrumentarium von Individuen, ihren Eigenschaften, Handlungen und ihren Beziehungen zueinander explizierbar sind (vgl. Hedström/Bearman 2009: 8). Es ist somit auch bei der Auseinandersetzung

⁷ So hängen beispielsweise Krieg und Mortalität eng miteinander zusammen, die hierbei relevanten Ereignisse, wie das Auslösen einer Waffe, eine mangelnde Nahrungsaufnahme und die Infektion mit Krankheiten, sind jedoch auf der Individualebene angesiedelt.

mit studentischen LVE erforderlich, auf Grundlage eines handlungstheoretischen Kerns (Logik der Selektion), einerseits strukturelle Einflüsse auf Individualvariablen herauszuarbeiten (Logik der Situation) und andererseits Aggregationsprozesse individueller Handlungen aufzuzeigen (Logik der Aggregation) (Esser 1999).

2.4 Schlussfolgerungen für die Untersuchung von LVE

Fasst man die bisherigen Überlegungen zusammen, so lassen sich folgende Punkte festhalten: Ausgangspunkt dieser Arbeit ist die philosophische Position des Realismus, nach der es eine objektive Welt gibt. Spezifiziert wird dieses Weltbild durch Poppers Kritischen Rationalismus. Danach ist eine Methode des Versuchs und Irrtums anzuwenden, um einer objektiven Wahrheit möglichst nahe zu kommen. Kühne Hypothesen, strenge empirische Prüfungen und wissenschaftliche Kritik fördern dabei den Erkenntnisfortschritt. Ziel ist die Aufdeckung von (deterministischen und probabilistischen) Regularitäten, die in wissenschaftlichen Erklärungen eine zentrale Rolle spielen. Ziel ist jedoch nicht nur, robuste Zusammenhänge aufzuzeigen, sondern vor allem, kausale Mechanismen hinter diesen Regularitäten freizulegen, also sie theoretisch zu erklären. Für die Soziologie ist dabei ein sozialer Mechanismus nur dann vollständig beschrieben, wenn die Schritte auf der Mikroebene, über welche dieser fungiert, expliziert werden.

Ein sinnvoller Ausgangspunkt ist demnach die Auseinandersetzung mit Theorien individuellen Handelns. Erst durch ein Verständnis der Handlungsmotive und Mikroprozesse kann man zu befriedigenden Tiefenerklärungen gelangen (zum Beispiel Coleman 1990; Esser 1999; Schmid 2006). Dies gilt auch für die Bewertung universitärer Lehre durch Studierende. Will man erklären, wie LVE zustande kommen, so ist ein Rekurs auf das individuelle Bewertungsverhalten unabdingbar. Schließlich ergibt sich die Lehrveranstaltungsbewertung aus dem Mittel der studentischen Einzelurteile. Dabei ist die Suche nach Regularitäten auf der Individualebene im Sinne mechanistischer Erklärungen ein sinnvolles Vorgehen. Insbesondere die Auffassung, dass menschliches Handeln anreizgeleitet ist, erscheint tragfähig zu sein. Im folgenden Kapitel wird daher der entsprechende handlungstheoretische Rahmen für diese Arbeit gelegt, denn ein rein statistischer Nachweis kausaler Zusammenhänge ist ohne vorhergehende theoretische Überlegungen bekanntlich nicht möglich. Poppers Rationalitätsprinzip bildet dazu einen geeigneten Ausgangspunkt, der allgemein genug gehalten ist, um verschiedene Ausdeutungen situationsangemessenen Handelns zu umfassen.

3 Menschenbild: Das Rationalitätsprinzip und seine Spezifizierungen

Ausgehend von Poppers methodologischem Postulat des Rationalitätsprinzips werden in diesem Kapitel verschiedene mögliche Spezifizierungen dieser Ausgangshypothese aufgezeigt. Es wird argumentiert, dass Ausdeutungen des Rationalitätsprinzips in Form klassischer Theorien rationaler Wahl für die behandelte Fragestellung ungeeignet sind. Stattdessen wird auf Arbeiten zur begrenzten menschlichen Rationalität (*bounded rationality*) zurückgegriffen. Anhand des Beispiels studentischer LVE wird illustriert, inwiefern Menschen sich bei ihrer Bewertung an Aspirationsniveaus orientieren (*satisficing*), Entscheidungen anhand von Urteilsheuristiken treffen und je nach situativen Gegebenheiten soziale Normen befolgen. Aus diesen Überlegungen ergeben sich einige allgemeine Hypothesen, welche in den nachfolgenden Kapiteln noch genauer ausformuliert werden.

3.1 Poppers Rationalitätsprinzip

Poppers Gedanken zur Situationsanalyse werden von manchen als sein wichtigster soziologischer und sozialwissenschaftlicher Beitrag bezeichnet – so zumindest die Hauptthese eines Beitrags von Hedström et al. (1998), in dem sich die Autoren mit der Bedeutung des Rationalitätsprinzips für die zeitgenössische Soziologie beschäftigen (siehe aber Schmid 1996 für eine Kritik an Poppers Vorschlag). Dieses kann als methodologisches Postulat aufgefasst werden, das für Popper (1967: 355) »integraler Bestandteil jeder, oder fast jeder, überprüfbaren Gesellschaftstheorie« ist. Es besagt, dass handelnde Wesen immer entsprechend der Situation, in der sie sich befinden, angemessen handeln (vgl. Popper 1967: 354). Deutlich erkennbar ist eine große Ähnlichkeit zu Max Webers (1922) Verständnis menschlichen Handelns als mit Sinn behaftete Aktivität, ebenso wie bezüglich Poppers weiteren Ausführungen zur Situationsanalyse (vgl. zum Beispiel Jacobs 1990). Die Situation wird dabei nicht als objektive Gegebenheit betrachtet, son-