

Daniel Großmann
Tobias Wolbring *Hrsg.*

Evaluation von Studium und Lehre

Grundlagen, methodische
Herausforderungen und Lösungsansätze



Springer VS

Evaluation von Studium und Lehre

Daniel Großmann • Tobias Wolbring
Herausgeber

Evaluation von Studium und Lehre

Grundlagen, methodische
Herausforderungen und
Lösungsansätze

 Springer VS

Herausgeber

Daniel Großmann
Universität Leipzig
Leipzig, Deutschland

Tobias Wolbring
Universität Mannheim
Mannheim, Deutschland

ISBN 978-3-658-10885-4

ISBN 978-3-658-10886-1 (eBook)

DOI 10.1007/978-3-658-10886-1

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

Springer VS

© Springer Fachmedien Wiesbaden 2016

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Lektorat: Cori Antonia Mackrodt

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Fachmedien Wiesbaden GmbH ist Teil der Fachverlagsgruppe Springer Science+Business Media (www.springer.com)

Inhalt

Grundlagen

Stand und Herausforderungen der Evaluation an deutschen Hochschulen	3
Daniel Großmann und Tobias Wolbring	
Entstehung und Grundlagen der Evaluation	27
Reinhard Stockmann	
Evaluation, Kausalität und Validität	57
Tobias Wolbring	

Input

Soziale Determinanten der Studienaufnahme und Fachwahl	93
Werner Georg und Tino Bargel	
Studienmotivation und Evaluation	123
Daniel Großmann	
Messung von studentischem Workload	185
Roger Berger und Bastian Baumeister	

Prozess

Lehrveranstaltungsevaluation an Hochschulen	227
Heiner Rindermann	

Konstruktvaliditätsprobleme von Lehrevaluationen und die Potentiale einer Methodenintegration zur Entwicklung von Befragungsinstrumenten – eine Mixed-Methods-Studie	263
Brigitte Metje und Udo Kelle	
Studierendenbefragungen als Panelstudie	289
Philipp Pohlenz, Frank Niedermeier, Melinda Erdmann und Juliana Schneider	
Output	
Studienabbruchquoten als Evaluationskriterium und Steuerungsinstrument der Qualitätssicherung im Hochschulbereich	323
Daniel Klein und Volker Stocké	
Kompetenzdiagnostik an der Hochschule	367
Felix Wolter und Jürgen Schiener	
Berufseinstiege von AkademikerInnen mit Migrationshintergrund in Österreich – Empirische Untersuchungsstrategie und zentrale Befunde	403
Roland Verwiebe und Melek Hacıoglu	

Geleitwort

Die Evaluation von Lehre und Studium, in den 1990er-Jahren allenfalls in Form von Pilotprojekten an einigen wenigen Hochschulen durchgeführt, ist heute in Deutschland allgegenwärtig. Dazu beigetragen hat die grundlegende Infragestellung der Qualität von Lehre und Studium vor dem Hintergrund hoher Studienabbrucherquoten und einer durchschnittlichen Studiendauer, die weit über der Regelstudienzeit lag. Als Reaktion auf die Kritik verabschiedete die Hochschulrektorenkonferenz 1995 eine Entschließung „zur Evaluation im Hochschulbereich unter besonderer Berücksichtigung der Lehre“ und der Wissenschaftsrat legte 1996 „Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation“ vor (für eine ausführliche Darstellung der Genese von Evaluationen und anderer Qualitätssicherungsverfahren im deutschen Hochschulwesen siehe vbw 2013, S. 15–19).

Nach der Novelle des Hochschulrahmengesetzes im Jahr 1998 und der damit einhergehenden Einführung von Bachelor- und Masterstudiengängen wurde die Qualitätssicherung in Lehre und Studium in den Landeshochschulgesetzen verankert. Es verwundert daher nicht, dass einer Umfrage der Hochschulrektorenkonferenz zufolge, inzwischen ein ganzer Werkzeugkasten von Instrumente und Verfahren für die Qualitätssicherung im Bereich von Lehre und Studium existiert. Das Instrument mit der stärksten Verbreitung ist die Lehrveranstaltungsbewertung durch Studierende. Sie wird in 96,5 Prozent aller Fachbereiche durchgeführt (vgl. HRK 2010, S. 26). Ebenfalls stark verbreitet sind u. a. Programmakkreditierungen (85,9 %), Absolventenstudien (74,7 %), die Evaluation von Studiengängen (57,5 %) und Erhebungen studentischer Arbeitsbelastungen (52,9 %), die – mit Ausnahme der Programmakkreditierung – den Fokus des vorliegenden Sammelbandes bilden.

Die Herausgeber des Sammelbandes weisen zu Recht darauf hin, dass in den frühen 1990er-Jahre durch studentische Lehrveranstaltungsbewertungen,

Absolventenbefragungen und Studiengangsevaluationen die Qualität von Lehre und Studium verbessert werden sollte. Durch die flächendeckende Einführung der Programmakkreditierung Ende der 1990er-Jahre kam es zu einem von vielen heute bedauerten Bedeutungsverlust der entwicklungsorientierten Verfahren der Qualitätssicherung. Mit der Einführung der Systemakkreditierung im Jahr 2008 und der Erprobung von Quality Audits werden Evaluationsverfahren, welche die Verbesserung der Qualität von Lehre und Studium zum Ziel haben, zweifelsohne wieder an Bedeutung gewinnen. Aus diesem Grunde werden in dem vorliegenden Sammelband weit verbreitete Evaluationsverfahren für den Bereich Lehre und Studium einer kritischen Würdigung unterzogen. Darüber hinaus zeigen die Autorinnen und Autoren in ihren Beiträgen auf, wie die wissenschaftliche Qualität der eingesetzten Evaluationsverfahren verbessert und damit die Belastbarkeit und Nützlichkeit von Evaluationen erhöht werden können.

Die Beiträge formulieren hohe Ansprüche an die Wissenschaftlichkeit von Evaluationen. Evaluationen sollen viel stärker als bisher Hypothesen zur Wirkungsweise von Lehrveranstaltungen und Studiengängen formulieren und im Rahmen von experimentellen und quasi-experimentellen Evaluationsdesigns (z. B. Kohortenstudien bzw. Studienverlaufsanalysen) mit validen Daten und adäquaten statistischen Analyseverfahren (z. B. Mehrebenenanalyse) testen. Dass die hohen Ansprüche in der Evaluationspraxis auch tatsächlich eingelöst werden können, zeigt in diesem Band nicht zuletzt die Beschreibung der Studierendenbefragungen an der Universität Potsdam, die als Studienverlaufsanalysen konzipiert sind.

Zürich

August 2015

Hans-Dieter Daniel

Literatur

HRK – Hochschulrektorenkonferenz. 2010. *Wegweiser 2010: Qualitätssicherung an Hochschulen*. Bonn: Projekt Qualitätsmanagement, Beiträge zur Hochschulpolitik 8/2010.

vbw – Vereinigung der Bayerischen Wirtschaft e.V, Hrsg. 2013. *Qualitätssicherung an Hochschulen: von der Akkreditierung zur Auditierung (Gutachten des Aktionsrat Bildung)*. Münster: Waxmann.

Über die Autoren

Tino Bargel Jahrgang 1943. Studium der Soziologie in Bonn, Tübingen und Berlin. Ab 1968 Mitarbeit in einem internationalen Projekt zur Hochschulsozialisation an der Universität Konstanz (SFB 23). Mitbegründer des bundesweiten Studierendensurveys, des Studienqualitätsmonitors (SQM) sowie des Arbeitskreises „Qualität von Schule“. Moderation des internationalen Réseau Uni 21 (FREREF) zur Hochschulforschung. Anfang 2008 aus dem Hochschuldienst ausgeschieden. Seitdem verschiedene Berichte und Gutachten zur Studienqualität, zur sozialen Ungleichheit im Studium, zuletzt zum Teilzeitstudium und zur Studieneingangsphase. Kontakt: tino.bargel@uni-konstanz.de

Bastian Baumeister Jahrgang 1985. Studierte seit 2006 an der Universität Leipzig Soziologie und schloss im Jahr 2011 das Masterstudium ab. Ist seit 2012 wissenschaftlicher Mitarbeiter am Leipziger Institut für Soziologie, Lehrstuhl Soziologie und Methodenlehre. Forschungsschwerpunkt ist die Methodologie quantitativer Erhebungsverfahren. Kontakt: baumeister@sozio.uni-leipzig.de

Roger Berger Jahrgang 1969. Studium der Soziologie, Neueren Allgemeinen Geschichte und Wirtschaftsgeschichte. Promotion zum Dr. rer. pol. an der Universität Leipzig. Habilitation mit der Venia „Soziologie“ an der LMU München. Seit 2010 Professor für Soziologie und Methodenlehre an der Universität Leipzig. Forschungs- und Interessengebiete sind Methoden der empirischen Sozialforschung (insbesondere spieltheoretisch experimentelle Designs) und Kooperationsforschung (insbesondere fundierte soziale Dilemmata mit Anwendungen in der Wirtschaftssoziologie). Kontakt: berger@sozio.uni-leipzig.de

Hans-Dieter Daniel Jahrgang 1955. Diplomstudium der Psychologie. Promotion (1983) und Habilitation (1992) an der Universität Konstanz, Verleihung der

Venia legendi für das Fach Psychologie. 1992 bis 1995 Leiter des baden-württembergischen Modellprojekts „Evaluation der Lehre“ an der Universität Mannheim. 1995 bis 2001 Professor für Empirische Sozialforschung am Fachbereich Gesellschaftswissenschaften der Universität Kassel sowie Direktoriumsmitglied des Wissenschaftlichen Zentrums für Berufs- und Hochschulforschung (heute INCHER) Kassel. Seit 2001 Leiter der Evaluationsstelle der Universität Zürich und seit 2002 Professor für Sozialpsychologie und Hochschulforschung an der ETH Zürich. Forschungsschwerpunkte sind Hochschulforschung, Wissenschaftsforschung, Evaluationsforschung, Peer-Review-Forschung, Bibliometrie sowie Empirische Sozialforschung.

Melinda Erdmann Jahrgang 1983. Studierte bis 2012 an der Universität Potsdam Soziologie mit den Schwerpunkten empirische Sozialforschung und Sozialstrukturanalyse. Während und nach dem Studium arbeitete sie als Dozentin am Lehrstuhl für Methoden der empirischen Sozialforschung. Ihre Arbeitsschwerpunkte sind Methoden der empirischen Sozialforschung, Hochschulforschung und Bildungsforschung. Seit 2012 arbeitet sie als wissenschaftliche Mitarbeiterin am Zentrum für Qualitätsentwicklung in Lehre und Studium (ZfQ) an der Universität Potsdam in den Geschäftsbereichen Evaluation und Akkreditierung. Im Bereich der Evaluation leitet Sie das Pilotprojekt Lernzielorientierung und Lehrveranstaltungsevaluation und unterstützt das Potsdamer Studierendenpanel. Kontakt: melinda.erdmann@uni-potsdam.de

Werner Georg Studium der Politikwissenschaft und Europäischen Ethnologie, Promotion zum Dr. phil an der Universität Marburg, Habilitation mit der Venia „Soziologie mit Schwerpunkt Empirische Sozialforschung“ an der Universität Siegen, Seit 1996 Professor an der Universität Konstanz. Forschungs- und Interessengebiete sind Bildungssoziologie, Soziale Ungleichheit, Sozialstrukturanalyse sowie Lebensstil- und Jugendforschung. Kontakt: werner.georg@uni-konstanz.de

Daniel Großmann Jahrgang 1974. Medizinische Berufsausbildung, Fachoberschule mit psychologisch-pädagogischem Schwerpunkt, Studium der Soziologie und Politikwissenschaft (Magister Artium). Wissenschaftlicher Mitarbeiter für Evaluation an der Fakultät für Sozialwissenschaften und Philosophie der Universität Leipzig und Lehrbeauftragter am Institut für Soziologie. Mitglied des Zentrums für Quantitative Empirische Sozialforschung (QUANTEL) an der Universität Leipzig. Forschungs- und Interessengebiete sind Methoden der empirischen Sozialforschung (insbesondere Evaluationsmethoden) sowie Bildungsforschung. Kontakt: daniel.grossmann@uni-leipzig.de

Melek Hacıoglu Jahrgang 1983. Studium der Soziologie. Promoviert derzeit zum Dr. phil. am Institut für Soziologie an der Universität Wien und an der European University Institute Fiesole (Italien). Forschungs- und Interessengebiete sind soziale Ungleichheit und Diskriminierung, Migration, Bildung, Arbeitsmarktforschung, Mixed-Methods. Kontakt: melek.hacioglu@univie.ac.at

Udo Kelle Jahrgang 1960. Studium der Medizin, Psychologie und Soziologie in Bochum, Hannover, Bielefeld und Bremen, dort promoviert 1992. Habilitation mit der Venia „Soziologie“ an der Universität Bremen im Jahr 2005. Von 2005 bis 2010 Professor für Methoden Empirischer Sozialforschung an der Universität Marburg und seit August 2010 Professur für die Methoden empirischer Sozialforschung und Statistik an der Helmut-Schmidt-Universität der Bundeswehr Hamburg. Forschungs- und Interessengebiete sind Methoden qualitativer und quantitativer Forschung und deren wissenschaftstheoretische Grundlagen, methodische und methodologische Probleme der Evaluationsforschung und Methodenforschung in der Bildungs- und Lebenslaufsoziologie. Kontakt: kelle@hsu-hh.de

Daniel Klein Jahrgang 1981. Studium der Soziologie mit dem Schwerpunkt „Empirische Sozialforschung“ an der Otto-Friedrich-Universität Bamberg. Wissenschaftlicher Mitarbeiter an der Professur für Methoden der empirischen Sozialforschung der Universität Kassel. Seit 2014 Direktoriumsmitglied des Kompetenzzentrums für empirische Forschungsmethoden der Universität Kassel. Forschungs- und Interessengebiete sind quantitative Methoden der empirischen Sozialforschung sowie Bildungssoziologie und soziale Ungleichheit. Kontakt: daniel.klein@uni-kassel.de

Brigitte Metje Jahrgang 1953. Studium der Diplom-Gerontologie an der Hochschule Vechta. Promotion zur Dr. phil. am Institut für Soziologie an der Universität Marburg. Seit September 2010 wissenschaftliche Mitarbeiterin an der Helmut-Schmidt-Universität der Bundeswehr Hamburg. Forschungs- und Interessengebiete sind: Validitätsprobleme von Lehrveranstaltungsevaluationen, Mixed-Methods in der Evaluationsforschung, Evaluation militärischer Ausbildung. Kontakt: metjeb@hsu-hh.de

Frank Niedermeier Jahrgang 1981. Studium der Soziologie, Politikwissenschaft und Geschichte an den Universitäten Oldenburg und Potsdam. Seit 2010 wissenschaftlicher Mitarbeiter am Zentrum für Qualitätsentwicklung in Lehre und Studium der Universität Potsdam wo er das Studierenden-Panel sowie das Projekt ASEAN-QA, Qualitätsmanagementkurse für den südostasiatischen

Hochschulraum, leitet und koordiniert. Seine Forschungs- und Arbeitsschwerpunkte sind die Hochschul- und Studierendenforschung, Evaluation von Lehre und Studium, Methoden der empirischen Sozialforschung sowie Qualitätsmanagement und -entwicklung an Hochschulen. Kontakt: frank.niedermeier@uni-potsdam.de

Philipp Pohlenz Jahrgang 1971. Studium der Soziologie und Geographie an den Universitäten Hamburg und Potsdam. Promotion an der dortigen Wirtschafts- und Sozialwissenschaftlichen Fakultät 2008. Von 2000 bis 2013 wissenschaftlicher Mitarbeiter an der Universität Potsdam im Bereich der Evaluation von Lehre und Studium sowie Leiter des dortigen Zentrums für Qualitätsentwicklung in Lehre und Studium (ZfQ). Freiberufliche Tätigkeiten als Berater in Reformprojekten im Hochschulbildungsbereich in verschiedenen Ländern Südostasiens und Südosteuropas. Seit 2014 Inhaber der Professur für Hochschulforschung und Professionalisierung der akademischen Lehre an der Otto-von-Guericke-Universität Magdeburg. Kontakt: philipp.pohlenz@ovgu.de

Heiner Rindermann Jahrgang 1966. Studium der Psychologie (und in Nebenfächern ein- bis zweisemestrig Philosophie, Ethnologie, Soziologie, Medizin, Politologie und Altphilologie) in Heidelberg. 1995 Promotion ebendort. 2005 Habilitation im Fach Psychologie an der Universität Landau. 2008–2010 Professur für Evaluation und Methodik der Entwicklungspsychologie, KFU Graz. Seit 2010 Professur für Pädagogische und Entwicklungspsychologie an der TU Chemnitz. Forschungsthemen: Unterrichtsqualität und Lehrevaluation, kognitive Fähigkeiten (Intelligenz, Schulleistung) auf individueller und gesellschaftlicher Ebene, ihre Bedingungen, Förderung und Folgen sowie Kulturvergleich und Humankapital. Kontakt: heiner.rindermann@psychologie.tu-chemnitz.de

Jürgen Schiener Jahrgang 1968. 1990–1998 Studium der Soziologie, Politikwissenschaften und Ethnologie (Magister Artium) an der Johannes-Gutenberg Universität Mainz. Von 1998 bis 2005 wissenschaftlicher Mitarbeiter, seit 2005 Lehrkraft für besondere Aufgaben und seit 2014 akademischer Direktor am Institut für Soziologie der JGU Mainz u. a. als Mitglied des Zentrums für Bildungs- und Hochschulforschung. Forschungs- und Interessensgebiete sind vor allem soziale Strukturen und Ungleichheiten in den Bereichen Bildung, Weiterbildung, Hochschule und Arbeitsmarkt. Kontakt: juergen.schiener@uni-mainz.de

Juliana Schneider Jahrgang 1989. Bachelorstudium der Soziologie an der Technischen Universität Chemnitz und Masterstudium der Soziologie an der

Universität Potsdam. Seit 2015 wissenschaftliche Mitarbeiterin am Zentrum für Qualitätsentwicklung in Lehre und Studium der Universität Potsdam. Sie unterstützt im Bereich der Evaluation die Durchführung und Koordination der Lehr-evaluation sowie des Studierendenpanels. Ihr Forschungsinteresse liegt im Bereich der Methoden empirischer Sozialforschung insbesondere der Längsschnittdaten. Kontakt: juliana.schneider@uni-potsdam.de

Volker Stocké Jahrgang 1966. Studium der Soziologie, Wissenschaftslehre, Politische Wissenschaften, Öffentliches Recht und Germanistik an der Universität Mannheim. Promotion 2001 zum Dr. rer. phil. und Habilitation mit der *Venia Legendi* im Fach „Soziologie“ ebenfalls an der Universität Mannheim. Erstberufung zum Professor für Längsschnittliche Bildungsforschung an der Otto-Friedrich-Universität Bamberg. Seit 2012 Professor für Methoden der empirischen Sozialforschung an der Universität Kassel. Seit 2012 Direktoriumsmitglied des Internationalen Zentrums für Hochschulforschung (INCHER) und seit 2014 geschäftsführender Direktor des Kompetenzzentrums empirischer Forschungsmethoden der Universität Kassel. Forschungs- und Interessengebiete sind Methoden der empirischen Sozialforschung, Bildungssoziologie und Stratifikationsforschung. Kontakt: volker.stocke@uni-kassel.de.

Reinhard Stockmann Jahrgang 1955. 1978–1982 Studium der Fächer Soziologie, Sozialpsychologie, Zeitgeschichte und Politische Wissenschaften an der Universität Mannheim. 1982 Diplom, 1986 Promotion, 1995 Habilitation mit der *Venia* „Soziologie“. Seit 1997 C4-Professur für Soziologie an der Universität des Saarlandes. Seit 1997 Geschäftsführender Direktor des Instituts für Soziologie, seit 2001 Herausgeber der Reihe „Sozialwissenschaftliche Evaluationsforschung“ im Waxmann Verlag, seit 2002 Gründer und Leiter des Centrums für Evaluation (CEval), seit 2002 Mitbegründer und geschäftsführender Herausgeber der „Zeitschrift für Evaluation“, seit 2004 Leiter des MA-Studiengangs „Evaluation“ der Universität des Saarlandes und der Hochschule für Technik und Wirtschaft des Saarlandes. Interessengebiete sind: Theorien und Methoden der Evaluation, Soziologie und Politik der Entwicklungsländer, Entwicklungszusammenarbeit, Bildung, Organisationssoziologie. Kontakt: r.stockmann@ceval.de

Roland Verwiebe Jahrgang 1971. Studium an der Humboldt-Universität zu Berlin und an der Columbia-University in New York City. Seit Anfang 2009 Professor für Sozialstrukturanalyse und Quantitative Methoden am Institut für Soziologie der Universität Wien. Arbeitsschwerpunkte: Migration, Arbeitsmarkt, Ungleichheit, Armut, Lebensqualität. Methodenschwerpunkte: fortgeschrittene

statistische Verfahren, qualitative Erhebungs- und Auswertungsverfahren, Mixed-Methods. Kontakt: roland.verwiebe@univie.ac.at

Tobias Wolbring Jahrgang 1982. Studium der Soziologie, Volkswirtschaftslehre und Psychologie. 2013 Promotion zum Dr. rer. pol. mit der Arbeit „Methodische Fallstricke bei der Evaluation universitärer Lehre“ an der Ludwig-Maximilians-Universität München. 2013 – 2014 Postdoktorand an der Professur für Sozialpsychologie und Hochschulforschung, ETH Zürich. Seit 2015 Juniorprofessor für Soziologie, insbesondere Längsschnittdatenanalyse an der Universität Mannheim. Seit 2013 Mitherausgeber der „Sozialen Welt“. Forschungs- und Interessengebiete sind Analytische Soziologie, Methoden der empirischen Sozialforschung (insbesondere Experimente, Kausalanalyse, Paneldaten), soziale Ungleichheit und Diskriminierung, Wirtschaftssoziologie sowie Wissenschafts- und Hochschulforschung. Kontakt: wolbring@uni-mannheim.de

Felix Wolter Jahrgang 1979. Studium der Soziologie, Politikwissenschaft und Betriebswirtschaftslehre in Mainz und Lyon. 2011 Promotion zum Dr. phil. an der Universität Mainz mit der Arbeit „Heikle Fragen in Interviews. Eine Validierung der Randomized-Response-Technik“. Seit 2007 wissenschaftlicher Mitarbeiter am Institut für Soziologie der Universität Mainz, 2015/2016 Vertretung einer W3-Professur für Soziologie am Karlsruher Institut für Technologie. Forschungs- und Interessengebiete sind Methoden der empirischen Sozialforschung (insbes. „heikle Fragen in Survey-Interviews“), soziale Ungleichheit (Bildung, Arbeitsmarkt), Rational-Choice- und Spieltheorie sowie Kompetenzdiagnostik. Kontakt: felix.wolter@uni-mainz.de

Grundlagen

Stand und Herausforderungen der Evaluation an deutschen Hochschulen

Daniel Großmann und Tobias Wolbring

Zusammenfassung

Evaluationen von Studium und Lehre sind in Deutschland inzwischen flächendeckend verbreitet und Bestandteil eines umfassenderen Qualitätssicherungsparadigmas geworden. Die eingesetzten Instrumente und Verfahren weisen jedoch methodische Probleme auf. In der Kritik stehen deren Genauigkeit und Verlässlichkeit sowie der Umgang mit den gewonnenen Ergebnissen. Der Beitrag führt in die Thematik ein. Er nimmt zunächst eine Bestandsaufnahme der Evaluation an deutschen Hochschulen vor und benennt anschließend drei zentrale Handlungsfelder, in denen besondere Herausforderungen bestehen: *Verhältnis von Evaluation und Anreizen, Professionalisierungsbedarf, methodische Herausforderungen*. Der Beitrag schließt mit einer Zusammenschau und thematischen Einordnung der Sammelbandbeiträge.

1 Einleitung

Evaluationen in Form von Befragungen Studierender zur Lehr- und Studiensituation sind in der deutschen Hochschullandschaft inzwischen flächendeckend verbreitet und mit der Bologna-Reform Teil eines umfassenderen Qualitätssiche-

Siehe „Über die Autoren“ für nähere Informationen.

D. Großmann (✉)
Universität Leipzig, Leipzig, Deutschland
E-Mail: daniel.grossmann@uni-leipzig.de

T. Wolbring
Universität Mannheim, Mannheim, Deutschland
E-Mail: wolbring@uni-mannheim.de

rungsparadigmas geworden. Ob als studentische Lehrveranstaltungsevaluationen oder Studiengangsbefragungen, dienen sie in den Hochschulen zur qualitativen (Weiter-)Entwicklung der Studienangebote und vielerorts zum inzwischen verpflichtenden Nachweis über die Erfüllung der Lehraufgaben und Zielvereinbarungen. Überdies werden vermehrt evaluationsbasierte Dozierenden-Ranglisten innerhalb der Fakultäten oder Universitäten veröffentlicht, zur Allokation von Ressourcen oder Boni genutzt und vereinzelt sogar der allgemeinen Öffentlichkeit zugänglich gemacht. Zudem werden Evaluationen von hochschulexternen Instanzen im Rahmen von Rankings als Bewertungsgrundlage für den Studiengangsbzw. Hochschulvergleich herangezogen.

Während Evaluationen in Studium und Lehre innerhalb der Hochschulen als Impuls- und Informationsgeber für die qualitative Entwicklung von Studienangeboten weitgehend akzeptiert sind, stößt Ihre Funktion als Steuerungs- und Kontrollinstrument – insbesondere im Kontext der Ressourcenverteilung – auf Vorbehalte bis hin zur vollständigen Ablehnung. Die vorgebrachten Einwände beziehen sich dabei häufig auf methodische Probleme, verbunden mit Kritik an der Genauigkeit und Verlässlichkeit der Verfahren sowie auf die unsachgemäße Nutzung damit gewonnener Ergebnisse.¹ Wenngleich ein Teil der in diesem Diskurs geäußerten Kritik eher affektheuristischer Art zu sein scheint, so sind die Einwände gegen die derzeitige Evaluationspraxis in Studium und Lehre doch keineswegs unbegründet. Denn abgesehen von „methodisch-handwerklichen“ Fehlern, die häufig und in Folge des in Deutschland noch immer vergleichsweise geringen Professionalisierungs- und Institutionalisierungsgrades in der (Hochschul-)Evaluation auftreten (dazu Schmidt 2009, S. 167; Stockmann und Meyer 2010, S. 40), unterliegen die Evaluationsverfahren wie nahezu alle Methoden der empirischen Sozialforschung der Gefahr von Verzerrungen durch selektive Effekte und unbeachtete Drittvariableneinflüsse. Dies gilt insbesondere deshalb, da Evaluationen häufig auf Querschnittsdesigns beruhen, während der sozialwissenschaftliche State of the Art die Nutzung von Experimental- oder Längsschnittdaten (siehe Pohlenz et al., Stockmann sowie Wolbring in diesem Band) sowie von Mixed-Methods-Ansätzen (siehe Metje und Kelle sowie Verwiebe und Hacioglu in diesem Band) nahelegt. Infolgedessen könnten Evaluationen im schlimmsten Falle zu vollkommen falschen Schlüssen verleiten und die falschen Akteure belohnt bzw. bestraft werden; in jedem Falle aber steht der scheinbaren

¹ Um einen falschen Eindruck zu vermeiden: Neben den Studierendenbefragungen dienen in den meisten Rankings zusätzlich Struktur- und Kenndaten als Bewertungs-/Vergleichsgrundlage. Methodische Probleme bestehen jedoch auch bei dieser Form der Hochschul- bzw. Studiengangsevaluation.

Simplizität, Präzision und Faktizität quantitativer Indikatoren ein erhebliches Maß von Messungenauigkeit gegenüber. Diese methodisch bedingte Unsicherheit hat zur Konsequenz, dass Evaluationsresultate häufig nicht ohne weitere Korrekturen für verschiedene Lehrveranstaltungen, Dozierende, Fachbereiche und Universitäten miteinander verglichen werden können und auch Zeitreihenvergleiche für dieselbe Bewertungseinheit nicht immer ohne Probleme möglich sind.

Dies impliziert jedoch keineswegs, dass Evaluationen in Studium und Lehre deshalb per se obsolet sind. Vielmehr bedeutet es, dass die genutzten Methoden systematisch hinsichtlich solcher Effekte und ggf. deren Kontrolle erforscht werden sollten, um die Belastbarkeit gewonnener Evaluationsergebnisse besser abschätzen zu können. Die *Forschung über Evaluation* ist somit auch für die Bewertung von Studium und Lehre von zentraler Bedeutung (vgl. Hense und Widmer 2013). Während international ein entsprechender Trend zur Evaluation der Evaluation seit längerer Zeit zu beobachten ist, beginnt das Thema in der deutschen Evaluationslandschaft erst langsam an Bedeutung zu gewinnen (ebd., S. 171f.). Gerade im Bereich der Evaluation von Studium und Lehre ist eine solche Entwicklung zu begrüßen (Mittag et al., 2012). Denn erst anhand von Erkenntnissen aus der Forschung über Evaluation lässt sich abwägen, für welches Evaluationsparadigma zwischen Erkenntnisgewinn, Entwicklungsbegleitung, Steuerung und Kontrolle (dazu Kromrey 2001, S. 113f.) die jeweiligen Verfahren geeignet sind und für welche eher nicht.

In diesem Sammelband sollen solche Problemzonen der Evaluation von Lehre und Studium aufgezeigt und zugleich ein möglicher Umgang mit ihnen erörtert werden. In den Beiträgen werden historische und methodische Grundlagen von Evaluationen besprochen, Potenziale und Grenzen spezifischer Evaluationsverfahren aufgezeigt (z. B. Workload-Analysen, Kompetenz-Messung, Absolventenstudien) und mögliche externe Einflüsse auf Evaluationsergebnisse betrachtet (z. B. soziale Herkunft, Studienmotivation). Zunächst werden aber in diesem einleitenden Beitrag nach einer kurzen Bestandsaufnahme von Maßnahmen zur Qualitätssicherung an deutschen Hochschulen entsprechende Herausforderungen skizziert. Der Beitrag schließt mit einer Synopse, in der die Sammelbandbeiträge zusammengefasst und aufeinander bezogen werden.

2 Stand der Evaluation an deutschen Hochschulen

Aussagekräftiges Datenmaterial für eine systematische Untersuchung zur aktuellen Verbreitung verschiedener Verfahren zur Sicherung, Bewertung und Entwicklung der Qualität von Studium und Lehre ist für die deutsche Hochschullandschaft

leider Mangelware. Noch am ehesten wird diesem Interesse eine von der Hochschulrektorenkonferenz Anfang 2010 durchgeführte Online-Befragung von Hochschulleitungen und Fachbereichen deutscher Universitäten und Fachhochschulen (HRK 2010) gerecht. Die Resultate dieser Umfrage zeigen, dass die mit Abstand weiteste Verbreitung entsprechender Verfahren an den befragten Fachbereichen für die studentische Lehrveranstaltungsbeurteilung (96,5 %) zu konstatieren ist, mit abnehmender Bedeutung gefolgt von Programmakkreditierungen (85,9 %), Erstsemesterbefragungen (75,5 %), Absolventenstudien (74,7 %), Evaluationen von Studiengängen (57,5 %), Infrastruktur (51,6 %) und Modulen (49,6 %) sowie der Messung der studentischen Arbeitsbelastung (52,9 %) (ebd., S. 26f.). Der höchste Stellenwert wird dabei von Seiten der befragten Hochschulleitungen Absolventenstudien (75,6 % große/sehr große Bedeutung), studentischer Veranstaltungskritik (72,8 %) und mit etwas Abstand Programmakkreditierungen (65,0 %) beigemessen (ebd., S. 15).

Festzuhalten ist daher erstens, dass eine Vielzahl von Verfahren zur Evaluation von Studium und Lehre an deutschen Hochschulen Einsatz findet und dabei verschiedene Formen der Befragung (bspw. Studiengangsbefragungen, Lehrevaluationen, Absolventenbefragungen) eine exponierte Rolle spielen (siehe Daniel 1994, 1996 für frühe Beispiele). Zweitens zeigt sich, dass diesen Evaluationsverfahren von den entsprechenden universitären Instanzen durchaus hohe Bedeutung beigemessen wird.

In historischer Perspektive ist zudem sowohl ein Zuwachs standardisierter Verfahren der Qualitätssicherung an deutschen Hochschulen, als auch ein Institutionalisierungsschub zu verzeichnen (siehe Hüther und Krücken 2016). So illustriert die HRK-Studie eindrücklich, dass die Zahl der Stellen, die auf Hochschulebene angesiedelt sind und auf Qualitätssicherung spezialisiert sind (z. B. in Form von Stabsstellen, zentralen Evaluationsagenturen), im Vergleich zu den vorherigen HRK-Befragungen der Jahre 2005 und 2007 erheblich gestiegen ist.² Gab es 2005 (2007) noch durchschnittlich 1,16 (1,07) Stellen in diesem Bereich an deutschen Universitäten und Fachhochschulen, wuchs die mittlere Zahl um knapp 60 Prozent auf 1,73 Stellen an (ebd., S. 8). Mit knapp 69 Prozent fiel dabei der Anstieg an

² Es sei darauf hingewiesen, dass ein solcher Vergleich über verschiedene Erhebungswellen voraussetzt, dass sich die Zusammensetzung der Stichprobe in den verschiedenen Befragungswellen nicht systematisch verändert hat. Dies kann, u. a. angesichts der schwankenden Rücklaufquoten in den verschiedenen HRK-Erhebungen, nicht schlussendlich geklärt werden. In dem ausführlichen Bericht der HRK wird trotz entsprechender Vergleiche im Zeitverlauf dieser wichtige methodische Aspekt leider nicht thematisiert, was symptomatisch für weite Teile der empirischen Hochschulforschung ist.

Universitäten (2,6 Stellen) etwas stärker aus, als an Fachhochschulen (56 % bzw. 1,2 Stellen). Zu vermuten ist, dass dieser Unterschied in der absoluten Zahl an Stellen, aber auch der Entwicklung im Zeitverlauf wesentlich damit zusammenhängt, dass mittlere und größere Universitäten vermehrt zentrale Einrichtungen zur Qualitätssicherung schaffen. An Fachhochschulen ist dies dagegen häufig aufgrund mangelnder Größe nicht praktikabel.

Ungeachtet dessen, welche Ursachen dieser merkliche Stellenausbau hat, zeigen die Zahlen in jedem Falle, dass die Evaluation von Studium und Lehre in den letzten Jahren verstärkt institutionalisiert und professionalisiert wurde. Seyfried und Pohlenz (2013, S. 48) beschreiben in diesem Zusammenhang eine neue Klasse an Evaluatorinnen und Evaluatoren: Mit Whitchurch (2010) benennen sie diese als Profession des „third space“. Die neue Berufsgruppe zeichnet sich einerseits durch einen akademischen Hintergrund sowie idealerweise durch dezidierte Kenntnisse im Bereich der Methoden der empirischen Sozialforschung aus, andererseits sind diese „neue[n] Hochschulprofessionellen“ (Winter 2014, S. 3) in die universitäre Administration integriert, übernehmen flexible Verwaltungstätigkeiten und haben vermittelt über Hochschulleitungen und administrative Instanzen oft erheblichen Einfluss auf die universitäre Praxis. Obwohl damit eigentlich eine wichtige Voraussetzung zur Professionalisierung der Qualitätsentwicklung im Hochschulwesen gegeben ist, mangelt es jedoch, wie Winter (2014) betont, immer wieder an der personellen und institutionellen Kontinuität.

Dies hängt insbesondere damit zusammen, dass viele der entsprechenden Stellen nicht aus dem regulären Etat der Hochschulen finanziert werden, sondern im Rahmen von zeitlich befristeten Förderprogrammen von Bund, Ländern und Stiftungen geschaffen und somit nach dem Auslaufen der Programme gestrichen werden. Hinzu kommt, dass der Schaffung von haushaltsfinanzierten Personalkapazitäten für die Qualitätssicherung an den Hochschulen häufig strukturelle Veränderungen in der Hochschulverwaltung vorausgingen, in deren Folge mitunter nicht ausreichend qualifiziertes Personal auf die neu geschaffenen Qualitätssicherungsstellen umbesetzt wurde. Doch auch bei der gezielten Suche nach qualifizierten Evaluatorinnen und Evaluatoren müssen die Hochschulen und Fachbereiche oft pragmatisch vorgehen, da das potentielle Angebot aufgrund einer noch immer geringe Zahl an einschlägigen Studiengängen sowie Aus- und Fortbildungsprogrammen deutschlandweit begrenzt ist (dazu Stockmann und Meyer 2010, S. 43).

Trotz dieser teils mangelnden Nachhaltigkeit kann davon gesprochen werden, dass sich die Evaluation von Studium und Lehre in Deutschland schrittweise institutionalisiert hat. Winter (2014) unterteilt die Entwicklung des Evaluationsparadigmas an deutschen Hochschulen hierbei vereinfachend in drei Phasen. In den 1990er-Jahren kam es zunächst zur Etablierung studentischer Lehrveranstal-

tungsevaluationen, später von Studiengangsevaluationen. Der formative Aspekt stand dabei eindeutig im Vordergrund – es ging also weniger um *Steuerung*, als vielmehr um *Entwicklung* der Studien- und Lehrprogramme und deren Qualität. In den 2000er-Jahren dominierte im Zuge der Einführung der modularisierten Bachelor- und Master-Studiengänge die Programmakkreditierung. Damit erhielten Kontrollüberlegungen mehr und mehr Bedeutung, obgleich es zunächst einmal darum ging, ein gewisses Mindestmaß an Qualität zu sichern und Fehlentwicklungen entgegenzusteuern. Die Etablierung von Programmakkreditierungen führt als Nebeneffekt – ob gewollt oder ungewollt – zu einem Bedeutungsverlust der bisherigen, entwicklungsorientierten Verfahren der Qualitätssicherung: Diese wurden durch die Programmakkreditierung substituiert.

Zudem geriet die Evaluation von Studium und Lehre damit in die Gefahr einer zunehmenden *Routinisierung* (siehe auch Stockmann in diesem Band), da dem Vollzug der Evaluationsverfahren eine höhere Priorität zukam, als den inhaltlichen und gestalterischen Fragen. In den 2010er-Jahren erkennt Winter schließlich einen erneuten Kurswechsel, in dessen Zuge die Programmakkreditierungen zunehmend durch die Systemakkreditierung verdrängt werden. Die Bedeutung der ursprünglichen Evaluationsverfahren könnte laut Winter dadurch wieder hergestellt und deren freiwilliger, fakultativer Charakter durch verpflichtende, hochschulweite Vorgaben ersetzt werden. Diese mögliche Aufwertung hängt dabei wesentlich damit zusammen, dass ein funktionierendes hochschulinternes Qualitätsmanagementsystem zentrale Voraussetzung für die erfolgreiche Absolvierung der Systemakkreditierung darstellt: „Ohne hochschulinterne Studiengangs- bzw. Fachevaluationen gibt es keine Akkreditierung des QM-Systems der Hochschule. Und ohne Systemakkreditierung dürfen die Studiengänge von der Hochschule nicht angeboten werden.[...] Eine Folge der Ausweitung der Systemakkreditierung könnte sein, dass die Fachevaluationen weiter verbreitet sein und ernster genommen werden als damals in den 1990er-Jahren“ (Winter 2014, S. 5).

Die Studiengangs- und Lehrevaluationen könnten demnach als Entwicklungsinstrumente an Bedeutung gewinnen. Im folgenden Abschnitt werden einige Anforderungen formuliert, die mit einer solchen Bedeutungszunahme verbunden sind und damit zugleich die Leitthemen dieses Sammelbandes aufzeigen.

3 Aktuelle Herausforderungen in der Evaluation von Studium und Lehre

Zentrale Herausforderungen in der Evaluation von Studium und Lehre ergeben sich aus dem hohen Anspruch, über eine (1) umfassende Informationsbasis über die Qualität von Lehrveranstaltungen, Studiengängen und Universitäten verfügen

zu wollen (*Informationswert*), (2) die unter angemessenem finanziellen und zeitlichen Aufwand herzustellen ist (*Effizienz & Sparsamkeit*), (3) die Interessen der verschiedenen Stakeholder berücksichtigt und diesen gerecht wird (*Akzeptanz & Partizipation*), (4) eine verlässliche evidenzbasierte Bewertung des Status-Quo erlaubt (*Belastbarkeit*), (5) als Grundlage für Entscheidungen auf Personen- und Institutionenebene dienen kann (*Entscheidungsfunktion*), (6) eine Rechtfertigung für diese liefert (*Legitimation*) und (7) sich positiv auf die Qualität von Studium und Lehre auswirkt (*Impact/Effektivität*). Selbstverständlich sind diese Anforderungen idealtypisch und nicht immer miteinander in Einklang zu bringen. Zudem besteht teilweise Uneinigkeit darüber, inwieweit gegenwärtige Evaluationsverfahren diesen Ansprüchen (z. B. Informationswert und Belastbarkeit) gerecht werden. Einige Konfliktlinien zwischen diesen Ansprüchen und der gegenwärtigen Form von Evaluation von Studium und Lehre sollen nun anhand folgender drei Bereiche aufgezeigt werden: (a) dem Anspruch der Politik, Evaluation verstärkt als Anreizprogramm zu betreiben, (b) dem Stand der Professionalisierung im Feld tätiger Akteure sowie (c) methodische Fallstricke und Mängel bei der Messung der Qualität von Lehre.

3.1 Evaluation und Anreize

Die deutschen Hochschulen haben in den vergangenen Jahrzehnten von der Politik einen deutlich höheren Status an institutioneller Autonomie erhalten. Die zentralen Punkte dieser gewachsenen Autonomie sind beispielsweise ein veränderter Rechtsstatus, die Verfügung über Globalbudgets sowie die weitgehende Übertragung des Berufungs- und Ernennungsrechts auf die Hochschulebene (HRK 2011), wengleich die Bundesländer dies hochschulrechtlich in differenzierter Form umgesetzt haben. Dieses Mehr an Freiheit und vor allem Verantwortung hat freilich seinen Preis: Die Hochschulen müssen ihre Arbeit in Lehre, Studium und Forschung nachprüfbar machen und dabei das Erreichen seitens der Politik vorgegebener Ziele nachweisen. Dazu gehört u. a. die Selbstverpflichtung, die Qualität von Forschung, Lehre/Studium und Verwaltung durch die Institutionalisierung entsprechender Evaluationsregime zu sichern und in vergleichbarer Weise transparent zu machen. Eine Zunahme der institutionellen Autonomie hat daher nicht notwendigerweise auch eine Abnahme der Kontrolle zur Folge, sondern vielmehr eine Verschiebung von externer zu interner Kontrolle. Hinzu kommt, dass mit Einführung der W-Besoldung eine leistungsorientierte Ressourcenallokation zum festen Bestandteil der Bezahlung der Professorenschaft geworden ist. Da entsprechende Vergabeverfahren nicht allein im subjektiven Ermessen von Fakultäts-, Fachbe-

reichs- oder Universitätsleitungen liegen sollen, bedürfen sie ebenfalls einer transparenten Zuteilungsregel, die üblicherweise an die Quantität und Qualität der erbrachten Lehr- und Forschungsleistung geknüpft wird. Der Schritt ist damit nicht mehr weit, Mittel zwischen und innerhalb von Fachbereichen in Abhängigkeit entsprechender quantitativer Indikatoren, wie Studierendenzahlen, Dropout-Raten oder Arbeitsmarkterfolg von Absolventen, gemäß fixer Verteilungsschlüssel zu vergeben und Leistungszulagen denjenigen Lehrenden zu gewähren, die viele Abschlussarbeiten betreuen, Zusatzleistungen für die Lehre erbracht haben oder von ihren Studierenden in der Lehrveranstaltungsevaluation besonders gut bewertet wurden.

Wie all diejenigen wissen, die an einer Hochschule tätig sind, handelt es sich bei dieser Darstellung nicht mehr um eine progressive hochschulpolitische Idee, sondern um die Deskription faktischer Zustände (siehe Wilkesmann und Schmid 2011 zu Anreizsystemen an deutschen Hochschulen). Obwohl dieser unter dem Schlagwort der Ökonomisierung gemeinhin bekannte Prozess (siehe Braun und Wolbring 2012; Schimank 2008) sicherlich noch nicht dieselben Ausmaße angenommen hat wie in anderen, insbesondere angelsächsischen, Ländern, wird hierzulande bereits trefflich darüber gestritten, welche nichtintendierten Folgen eine solche kennzahlenbasierte Anreizstruktur hat. Neben den allgemeinen normativen Bedenken gegenüber einer solchen Hochschulpolitik ist eine erste wesentliche Voraussetzung für eine erfolgreiche Setzung von Anreizen, dass die verwendeten Indikatoren genau das messen, was sie messen sollen, nämlich die Qualität von Studium und Lehre (*Validität*). Zweitens ist mittel- und langfristig in den Blick zu nehmen, ob sich das Verhalten der Hochschulakteure unter der Kennziffernkontrolle in nichtintendierter Weise verändert (*Absenz von Reaktivität*) und welche Folgen dies ggf. für Lehre und Forschung hat. Drittens muss ein verlässlicher *Ausschluss von Manipulierbarkeit* der Kennziffern gegeben sein. Viertens ist schließlich zu gewährleisten, dass sich die Anreize nur auf solche Aspekte beziehen, auf welche die entsprechenden Akteure mit den ihnen zur Verfügung stehenden Mitteln Einfluss nehmen können (*Fairness*) – wie sonst sollte der Einsatz von evaluationsbasierten Anreizen die angestrebten Veränderungen bewirken.

Dies sind die zentralen Fragen, will man Evaluationen im Sinne eines *New Public Management* als Steuerungs- und Kontrollinstrument einsetzen und ihnen neben der bisher weitgehend dominierenden *formativen* Funktion auch eine *summative* Rolle zuweisen. Der Forschung über Evaluation kommt dabei eine exponierte Rolle zu, da sie die Erfüllung der Voraussetzungen in jedem der vier genannten Bereiche – theoretisch statt normativ orientiert (vgl. Hense und Widmer 2013, S. 270) – hinterfragen soll und somit den Rahmen der Machbarkeit des „Kontrollparadigmas“ (Kromrey 2001, S. 113f.) absteckt.

3.2 Professionalisierungsbedarf

Die personelle und institutionelle Entwicklung der Evaluation von Studium und Lehre wurde bereits eingehend beleuchtet. Vor diesem Hintergrund ist sicherlich – gleichwohl personelle Kontinuität und Nachhaltigkeit der Stellen nicht immer gesichert sind – die Diagnose einer zunehmenden Differenzierung, Spezialisierung und Professionalisierung der Evaluationslandschaft zutreffend. In bestimmter Hinsicht ist diese Feststellung jedoch irreführend. Zu beachten ist nämlich, dass zwar viele der mit Evaluation und Qualitätssicherung betrauten Personen zuvor erfolgreich ein Studium absolviert haben, jedoch im Zuge dieser Ausbildung nicht immer das notwendige theoretische, methodische und statistische Rüstzeug an die Hand bekommen haben, das ihre spätere Tätigkeit eigentlich erfordern würde. Dies gilt insbesondere für diejenigen Evaluatoreninnen und Evaluatoren, die durch Weiterbildungs- und Personalentwicklungsmaßnahmen oder „emergentes Lernen“ zur Evaluationstätigkeit gekommen sind (siehe Hanft 2006 für diese Typologie).

Dieser noch nicht gesättigte Professionalisierungsbedarf zeigt sich auch in neueren Publikationen und einschlägigen Konferenzen. So hat die Deutsche Gesellschaft für Evaluation DeGEval ihre gemeinsam mit Schweizer Evaluationsgesellschaft SEVAL organisierte Jahrestagung 2014 dem Thema „Professionalisierung in und für Evaluation“ gewidmet. Und ähnlich fragen Pohlenz und Oppermann (2010) noch ergebnisoffen in ihrem Sammelbandtitel „Lehre und Studium professionell evaluieren. Wie viel Wissenschaft braucht die Evaluation?“ und beziehen dann in der von ihnen verfassten Einleitung klar Stellung für eine evidenzbasierte, wissenschaftlich fundierte Qualitätsbewertung. Weiterhin wird im Call for Papers für die 10. Jahrestagung der deutschen Gesellschaft für Hochschulforschung im Jahre 2015 explizit dazu aufgefordert, „neuere Theorie- und Methodenentwicklungen in wissenschaftlichen Disziplinen und interdisziplinären Forschungsfeldern, die für die Hochschulforschung von Bedeutung sind, zu erschließen“ (Krücken et al. 2014).

Neben methodischen Fragen nach der Gestaltung von Erhebungsdesigns, der praktischen Durchführung von Erhebungen und der statistischen Korrektur von Störeinflüssen ist hierbei einerseits an die adäquate graphische und tabellarische Aufbereitung von Ergebnissen zu denken, sodass Evaluationsresultate von allen Beteiligten korrekt erfasst und interpretiert werden können. Andererseits besteht aus Sicht der Herausgeber dieses Sammelbandes aber auch bezüglich der theoretischen Anleitung von Evaluationen erhebliches Entwicklungspotenzial: Dies ist erstens der Fall, da Konstrukte wie Lehrqualität oder Studienmotivation zunächst theoretisch fundiert definiert und operationalisiert werden müssen und vermittelnde Variablen für tiefer-

gehende Analysen im Sinne einer „white box“-Evaluation miteinbeziehen werden sollten. Zweitens ist die theoretische Anleitung von Evaluationsvorhaben aber auch deshalb von zentralem Stellenwert, weil sich die inhaltliche Bedeutung empirischer Befunde oft erst im Lichte theoretischer Vorüberlegungen herauskristallisiert – ein für die Ergebnisbewertung zentraler Aspekt. Schließlich ist ein theoriegeleitetes Vorgehen drittens auch deshalb wichtig, um methodische Probleme im Vorfeld zu erkennen und durch Wahl eines geeigneten Forschungsdesigns zu umschiffen.

3.3 Methodische Herausforderungen

Ein besonderer Stellenwert kommt aus Sicht der Autoren Fragen nach der Güte von Messungen zu. Methodische Herausforderungen ergeben sich einerseits aus neuen technologischen Möglichkeiten. Dies betrifft sowohl die Gestaltung von Studium und Lehre (z. B. E-Learning-Plattformen, MOOCs) als auch die Möglichkeiten bei Evaluationsstudien (z. B. Online-Lehrveranstaltungsevaluation, Smart-Phone-Apps zur Erhebung der studentischen Zeitverwendung). Offen ist in diesem Zusammenhang etwa, ob Messungen mittels klassischer und neuer Erhebungsmodi vergleichbare Resultate ergeben, aber auch, welches Vorgehen in konkreten Anwendungsfällen zu bevorzugen ist (für das Beispiel Workload-Messung siehe Berger und Baumeister in diesem Band). Andererseits stellen sich aber auch die aus der empirischen Sozialforschung altbekannten Fragen nach der Fragebogen-gestaltung, Operationalisierung, Stichprobenziehung, Sicherung des Rücklaufs, Anlage des Forschungsdesigns und angemessenen Auswertung der Daten (siehe Bortz und Döring 2006; Diekmann 2012; Schnell et al. 2010 für einführende Lehrbücher). Diese Aspekte sind zwar nicht neu, aber für die Evaluation von Studium und Lehre aus mehreren Gründen aktueller denn je.

Erstens kommt den Evaluationsresultaten im Zuge der Verbreitung des New Public Managements an Hochschulen eine wichtigere Bedeutung zu, sodass auch die Frage nach der Validität von Messungen deutlich an Relevanz gewinnt. Zweitens rückt die Setzung von Anreizen auf Grundlage von Impact-Evaluationen methodisch anspruchsvolle Fragen nach der Kausalität in den Mittelpunkt – denn letztlich sollen diejenigen Universitäten, Fachbereiche und Lehrenden belohnt werden, die eine Verbesserung ursächlich bewirkt haben. In den Sozialwissenschaften hat sich mittlerweile eine interdisziplinäre Forschungsrichtung herausgebildet, die sich mit diesen anspruchsvollen Fragen zu Möglichkeit und Grenzen des Nachweises von Ursache-Wirkungszusammenhängen vertieft befasst (z. B. Morgan and Winship 2015; Morgan 2013; Murnane and Willet 2010). Die kausal-analytischen Einsichten aus diesem Gebiet gilt es, in zukünftigen Evaluationen von

Studium und Lehre fruchtbar zu nutzen. Drittens sind schließlich im Bereich von Studium und Lehre und auch bei allgemeinen Bevölkerungsumfragen seit einiger Zeit sinkende Rücklaufquoten zu verzeichnen (siehe z. B. Georg und Bargel in diesem Band). Es kann dabei nicht per se davon ausgegangen werden, dass dieser Ausfall keine Konsequenzen für die interessierende Qualitätsbewertung hat. Die Evaluation von Studium und Lehre muss sich daher auch mit Verfahren beschäftigen, die Aussagen über die Systematik von Stichprobenselektion und fehlende Daten erlauben sowie deren negative Effekte beheben oder zumindest abmildern können.

4 **Zusammenschau der Beiträge des Sammelbandes**

Diese und weitere Fragen tauchen in den Beiträgen des Sammelbandes immer wieder auf und können als Leitthemen dieses Kompendiums verstanden werden. Der Sammelband gliedert sich dabei in vier Teile mit jeweils drei Beiträgen, die hier überblicksartig dargestellt werden.

Teil I – Grundlagen

In Teil I werden Grundlagen der Evaluation im Allgemeinen und der Qualitätsbewertung an Hochschulen im Besonderen behandelt. Nach diesem einleitenden Beitrag, der den Stand und die aktuellen Herausforderungen im Feld der Evaluation von Studium und Lehre zusammenfasst, führt Reinhard Stockmann (Kap. [Entstehung und Grundlagen der Evaluation](#)) in die *Grundlagen der Evaluation* ein. Nach einer Einordnung der Evaluation in den ideengeschichtlichen Kontext sowie einem historischen Abriss ihrer Entwicklung in den USA und Deutschland, stellt er ein Forschungsmodell für die Evaluation vor. Dabei geht der Autor auf die verschiedenen Definitionen des Evaluationsbegriffs ebenso ein, wie auf die Vielfältigkeit möglicher Evaluationsobjekte. Besprochen werden zudem die oben bereits angerissenen Leitfunktionen von Evaluation (Erkenntnis, Entwicklung, Legitimation und Kontrolle) sowie die Merkmale und Vorteile einer *formativen* oder *summativen Ausrichtung*. Zudem gibt der Autor wertvolle methodische und praktische Hinweise für eine gelingende Evaluation und zeigt im Beitrag mit *Routinisierung* und *Professionalisierung* die derzeit größten Herausforderungen auf, die es nicht nur in der Evaluation von Studium und Lehre zu lösen gilt.

Im Kap. [Evaluation, Kausalität und Validität](#) beschäftigt sich Tobias Wolbring mit methodologischen Hintergründen der Evaluation und rückt dabei insbesondere die hier bereits angesprochenen Aspekte *Validität* und *Kausalität* in den Fokus. Er illustriert diese anhand von konkreten Beispielen, die inhaltlich an die weiteren

Beiträge des Sammelbandes anschließen. Der Autor zeigt u. a., dass neben einem theoriegeleiteten Vorgehen, der Wahl des Untersuchungsdesigns eine wesentliche Rolle zukommt, will man belastbare empirische Evidenz für Ursache-Wirkungszusammenhänge vorlegen (design-basierte Forschung). Hierzu wird im Anschluss an die Arbeiten von Kromrey das Feldexperiment als idealtypisches Referenzdesign dargestellt, anderen Untersuchungsplänen gegenübergestellt und mit der dominierenden Evaluationsrealität an deutschen Hochschulen abgeglichen. Im Ergebnis zeigen Praxisbeispiele, dass das Vorhaben, belastbare Aussagen über die kausale Wirkung von Studienprogrammen und andere Maßnahmen im Hochschulkontext zu treffen, häufig an einer theoretisch und methodisch nicht hinreichend fundierten Untersuchungsplanung krankt. Wie der Autor verdeutlicht, muss für eine fundierten Evaluation jedoch nicht immer ein experimentelles Design zur Anwendung kommen; oftmals sind bereits quasi-experimentelle und längsschnittliche Designs ausreichend, um verlässliche Befunde zu generieren.

Teil II – Input

Unter dem Label der Input-Evaluation wird üblicherweise die Evaluation von Merkmalen zusammengefasst, die als externe Ressourcen in Studium und Lehre eingehen und von denen angenommen werden kann, dass sie Erfolg und Qualität des Studiums maßgeblich mitbestimmen. Im Teil II des Sammelbandes stehen vor allem personale Ressourcen im Fokus – also Aspirationen, die Studierende in Form von Fähigkeiten, Orientierungen, Erwartungen und Verhaltensweisen mit in das Studium einbringen. Dass die Einbeziehung solcher Faktoren in die Evaluation von Studium und Lehre von zentraler Bedeutung ist – insbesondere für deren *Informationswert* (siehe Abschnitt 3) – muss hier nicht näher ausgeführt werden. Dennoch wird die Wirkung einiger wichtiger Einflussgrößen noch wenig beachtet. Daher ist deren Erhebung wenig verbreitet (z. B. soziale Herkunft), besteht bei einigen Konstrukten erheblicher methodischer Entwicklungsbedarf (z. B. Messung des studentischen Workloads) und haben einzelne Faktoren unerwünschten Einfluss auf die Bewertung im Rahmen der Evaluation selbst (z. B. Studienmotivation). Dem wird in den drei Beiträgen des Abschnitts nachgegangen.

Im Kap. [Soziale Determinanten der Studienaufnahme und Fachwahl](#) nehmen Werner Georg und Tino Bargel soziale Determinanten der Studienwahl in den Blick und zeichnen ein systematisches und differenziertes Bild der *sozialen Heterogenität unter Studierenden*. Die Autoren gehen auf zentrale Merkmale und deren Differenzierung ein, zeigen dabei beispielsweise die notwendige Unterscheidung zwischen *Studienaufnahme* und *Studienfachwahl* sowie zwischen *vertikalen* (soziale Herkunft) und *horizontalen* (fachliche Herkunft) Herkunftseffek-

ten. In der Sozialstruktur- wie auch der Hochschulforschung ist die Analyse sozialer Ungleichheiten im Kontext von Studienentscheidung und Studienfachwahl verbreitet und stellt eine exponierte Thematik dar. Wie die Autoren in dem Beitrag zeigen, ist sie deshalb von so großer Bedeutung, „[...] weil damit die soziale Herkunft der Studierenden mit ihrer sozialen Zukunft verknüpft wird [...]“ (Georg und Bargel in diesem Band, Abschn. [Einleitung](#)). Die Studienwahl ist also ein wichtiger Schlüssel zur Erklärung der gesellschaftlichen Reproduktion sozialer Ungleichheit. Nichtsdestotrotz ist die Betrachtung sozialer Heterogenität unter Studierenden in der Evaluation von Lehre und Studium kaum verbreitet und sind die im Rahmen von Evaluationsbefragungen üblicherweise erhobenen Sozialstrukturmerkmale (Geschlecht, Kinderstand etc.) auch nur bedingt dazu geeignet. In diesem Zusammenhang führen die Autoren den Begriff der „sozialen Evaluation“ ein. Natürlich kann die Lehr- und Studiengangsevaluation in Anbetracht begrenzter Ressourcen nicht die Ursachen sozialer Ungleichheit erforschen, doch allein schon die Information über die soziale Heterogenität unter den Studierenden eines Studien- oder Lehrprogramms ist von Bedeutung, um es auf die entsprechenden Anforderungen hin entwickeln zu können. So ist ein Aspekt bspw. die Herkunft Studierender aus Nichtakademikerfamilien: Wie die Erfahrung und einzelne Studien zeigen (Müller und Schneider 2013, siehe auch Stocké und Klein in diesem Band), weisen diese Studierenden eine erhöhtes Risiko für Studienabbrüche auf, wobei eine wichtige Rolle u. a. soziokulturelle Orientierungsprobleme zu spielen scheinen. Zum Tragen können solche Probleme jedoch auch auf der horizontalen Ebene kommen, bspw. wenn Studierende zwar aus Akademikerfamilien kommen, die Eltern jedoch in Fächern mit einer grundlegend anderen Fachkultur zu Hause sind. Im Rahmen der Evaluation gesammelte Informationen über die soziale Zusammensetzung von Studierendenkohorten sind also insbesondere für die Studienstartphase und die Abbruch-/Wechselproblematik vielversprechend. Die Autoren geben in dem Beitrag Hinweise, wie und welche Merkmale der sozialen Herkunft im Rahmen von Evaluationsbefragungen erhoben werden sollten und illustrieren dies an den Instrumenten des Deutschen Studierendenurveys.

Im Kap. [Messung von studentischem Workload: Methodische Probleme und Innovationen](#) setzen sich Roger Berger und Bastian Baumeister mit methodischen Fragen zur *Messung studentischen Workloads* auseinander und erläutern dies anhand einer sophistizierten Pilotstudie, die über zwei Studiensemester in den Soziologiestudiengängen der Universität Leipzig stattfand. Mit der Einführung der modularisierten Studiengänge ist der Workloadbegriff zu einer zentralen und gleichermaßen kontrovers diskutierten Größe in der Studien- und Lehrprogrammgestaltung geworden. Die Anzahl der in einem Studiengang zu absolvierenden

Module, ebenso wie die Vergabe von ECTS-Punkten, bemisst sich im modularisierten Studiengangssystem formal am Arbeitsaufwand (Workload) der Studierenden. Dieser wurde im Zuge der Studiengangsreformen zunächst theoretisch-abstrakt bestimmt, so dass ein zentraler Gegenstand der Evaluation in Lehre und Studium – insbesondere im Kontext der Akkreditierung – die Kontrolle des tatsächlichen studentischen Workloads in den betreffenden Studienprogrammen ist. Wie oben bereits angesprochen wurde, scheint dies jedoch nicht flächendeckend stattzufinden.³ Dies liegt vermutlich nicht zuletzt daran, dass die Erhebungsmethoden für den studentischen Workload noch in den Kinderschuhen stecken und daher noch erhebliche Probleme bestehen. Die Autoren zeigen in ihrem Beitrag den methodischen und empirischen Wissensstand der Messung studentischen Workloads auf. Sie unterscheiden dabei zwischen zeitnahen, prospektiven und retrospektiven Erhebungsmethoden und testen diese im Rahmen der Untersuchung auf ihre Zuverlässigkeit. Als mit Abstand zuverlässigstes – aber auch aufwändigstes – Instrument erweist sich dabei die zeitnahe Tagebuchmethode. Für diese Methode werden im Beitrag unterschiedliche Designs besprochen. Im Mittelpunkt stehen dabei u. a. Überlegungen, wie der Erhebungsaufwand sowohl für die Evaluatoren als auch für die teilnehmenden Studierenden gesenkt werden kann. Im Rahmen der Pilotstudien wird dazu eine klassische Papier-Tagebuchvariante gegen eine App-Variante für Smartphones und Tablett-Computer getestet. Wenngleich die Tagebuchmethode die zuverlässigsten Ergebnisse liefert, so ergeben sich mittels der retrospektiven Workloaderhebung erstaunlich geringe Abweichungen von der Benchmark. Auch wenn freilich weitere Forschung auf diesem Gebiet aussteht, zeigen die Befunde von Berger und Baumeister, dass die Nutzung dieser mit nur geringem Aufwand erhebbaren Workload-Daten durchaus Sinn machen kann, wenn sie um Erkenntnisse über mögliche Verzerrungen aus vorangegangenen Tagebuch-Erhebungen korrigiert werden.

Im Kap. [Studienmotivation und Evaluation: Der Einfluss motivationaler Faktoren auf die befragungsbasierte Evaluation von Studium und Lehre](#) beschäftigt sich Daniel Großmann mit der Studienmotivation Studierender. Obwohl diese in der jüngeren Hochschulforschung – insbesondere im Kontext von Studieneingangs-, Wechsel- und Abbruchsanalysen – zu den am häufigsten be-

³ Wie oben bereits angesprochen wurde, zeigt eine Befragung der Hochschulrektorenkonferenz aus dem Jahre 2010, dass lediglich an rund 53 Prozent der Fachbereiche an deutschen Hochschulen Messungen der studentischen Arbeitsbelastung stattfinden (HRK 2010, S. 26f.).