

## 1 | 2020

### **Schwerpunkt: Politikevaluation und Evaluationspolitik**

- › Politikevaluation und Evaluationspolitik
- › Institutionalisierung der Evaluation
- › Wechselwirkungen von Politik und Evaluation
- › Rigorose Wirkungsevaluierung in der Entwicklungszusammenarbeit
- › Formen der Evaluierung experimenteller Politikgestaltung
- › Regierung der Wissenschaft im Peer Review
- › A conceptual framework for ex-ante evaluation in policymaking

### **Abhandlungen**

- › Wirkung der DOPPIK-Reform
- › Landesregierungen und die Unterbringung von Asylbewerbern
- › Kommunale Schuldenhilfen in Deutschland
- › Auswirkungen von Gemeindestrukturereformen in der Steiermark

### **Rezensionen**



13. Jahrgang  
1 Halbjahr 2020  
ISSN 1865-7192  
Verlag Barbara Budrich

**Herausgeber\*innen:**

Jörg Bogumil, Ruhr-Universität Bochum; Hermann Butzer, Leibniz Universität Hannover; Marian Döhler, Leibniz Universität Hannover; Anke Hassel, Hertie School, Berlin; Thurid Hustedt (Co-Geschäftsführung), Hertie School, Berlin; Werner Jann, Universität Potsdam; Christoph Knill, Ludwig-Maximilians-Universität München; Sabine Kuhlmann, Universität Potsdam; Susanne Lütz, FernUniversität in Hagen; Renate Meyer, Wirtschaftsuniversität Wien, Edward Page, London School of Economics and Political Science; Isabella Proeller, Universität Potsdam; Caspar van den Berg, Rijksuniversiteit Groningen; Sylvia Veit (Co-Geschäftsführung), Universität Kassel; Johanna Wolff, Freie Universität Berlin.

**Beirat:**

Harald Baldersheim, Universitetet i Oslo; Michael W. Bauer, Deutsche Universität für Verwaltungswissenschaften Speyer; Geert Bouckaert, Katholieke Universiteit Leuven; Martin Burgi, Ludwig-Maximilians-Universität München; Andreas Busch, Universität Göttingen; Jørgen Grønnegaard Christensen, Aarhus Universitet; Lars P. Feld, Walter Eucken Institut, Freiburg; Jobst Fiedler, Hertie School, Berlin; Edgar Grande, Wissenschaftszentrum Berlin für Sozialforschung (WZB); Heike Grimm, Universität Erfurt; Rolf G. Heinze, Ruhr-Universität Bochum; Stephan Hobe, Universität zu Köln; Martin Höpner, Max-Planck-Institut für Gesellschaftsforschung, Köln; Vincent Hoffmann-Martinot, CNRS Sciences Po Bordeaux; Lars Holtkamp, FernUniversität in Hagen; Michael Holoubek, Wirtschaftsuniversität Wien; Peter L. Hupe, Katholieke Universiteit Leuven; Ellen Immergut, Europäisches Hochschulinstitut Florenz; Desmond King, Nuffield College Oxford; Tanja Klenk, Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg; Sabine Kropp, Freie Universität Berlin; Andreas Lienhard, Universität Bern; Elke Löffler, Governance International, Birmingham; Veith Mehde, Leibniz Universität Hannover; Frank Nullmeier, Universität Bremen; Herbert Obinger, Universität Bremen; Janbernd Oebbecke, Universität Münster; Francesco Palermo, Europäische Akademie Bozen; Salvador Parrado Diez, Universidad Nacional de Educación a Distancia Madrid; Christoph Reichard, Universität Potsdam; Adrian Ritz, Universität Bern; Friedbert W. Rüb, Humboldt-Universität zu Berlin; Kuno Schedler, Universität St. Gallen; Uwe Schimank, Universität Bremen; Utz Schliesky, Christian-Albrechts-Universität Kiel; Volker Schneider, Universität Konstanz; Eckhard Schröter, Deutsche Hochschule der Polizei, Münster; Fritz B. Simon, Universität Kaiserslautern; Tobias ten Brink, Jacobs Universität Bremen; Annette Elisabeth Töller, FernUniversität in Hagen; Reimut Zohlnhöfer, Ruprecht-Karls-Universität Heidelberg.

**Redaktion:**

Prof. Dr. Thurid Hustedt  
Professor of Public Administration and Management  
Hertie School  
Friedrichstraße 180, D-10117 Berlin  
Tel.: +49 (0)30-25-92-19358  
E-Mail: dms@hertie-school.de

**Verlag:**

Verlag Barbara Budrich GmbH, Opladen, Berlin & Toronto.  
Stauffenbergstr. 7, D-51379 Leverkusen  
Tel +49 (0)2171 79491 50. Fax +49 (0)2171 79491 69  
E-Mail: info@budrich.de.  
<https://budrich.de>

**Erscheinen und Bezugsbedingungen:**

Jährlich zwei Hefte. Institutionenpreis print 150,00 €; Institutionenpreis nur online 248,00 €; Institutionenpreis print und online 273,- €; Einzelbezieher print 109,00 €, Einzelbezieher print und online 143,00 €; Einzelbezieher nur online 99,00 €; Versandkosten Printversion 6,00 €, Ausland 14,- €. Einzelheft: 75,00 € plus Porto.  
Alle Preise unterliegen der Preisbindung. Die Bezugspreise enthalten die gültige Mehrwertsteuer. Kündigungen des Abonnements müssen spätestens drei Monate vor Ablauf des Kalenderjahres schriftlich erfolgen.

© Edmund Budrich. Beratung und Betreuung von Verlagsprojekten. Alle Rechte vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Copyright-Inhabers vervielfältigt oder verbreitet werden. Unter dieses Verbot fallen insbesondere die gewerbliche Vervielfältigung per Kopie, die Aufnahme in elektronische Datenbanken und die Vervielfältigung auf CD-Rom und allen anderen elektronischen Datenträgern.

**dms ist aufgenommen in u.a. folgende Datenbanken & Externe Webseiten:**

CNKI | CNPeReading | Crossref | EBSCO | GBI-Genios | Google Scholar | ProQuest Politics Collection | ProQuest Social Science Premium Collection | scholars-e-library | Ulrichsweb

Satz: Glaubitz. Redaktion und Satz, Leverkusen (E-Mail: [glaubitz.rs@t-online.de](mailto:glaubitz.rs@t-online.de))

Druck und Verarbeitung: paper & tinta, Warschau. Printed in Europe

ISSN 1865-7192 / ISSN Online: 2196-1395

## Inhaltsverzeichnis

**Themenschwerpunkt: Politikevaluation und Evaluationspolitik  
(Herausgeber: Holger Straßheim und Oliver Schwab)**

|                                                         |                                                                                                                                |     |
|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Holger Straßheim</i><br><i>Oliver Schwab</i>         | Politikevaluation und Evaluationspolitik .....                                                                                 | 3   |
| <i>Wolfgang Meyer</i><br><i>Reinhard Stockmann</i>      | Institutionalisierung der Evaluation in<br>den politischen Systemen Europas.<br>Eine vergleichende Analyse .....               | 24  |
| <i>Thomas Widmer</i>                                    | Wechselwirkungen von Politik und Evaluation:<br>Befunde aus der Schweiz .....                                                  | 44  |
| <i>Jörg Faust</i>                                       | Rigoreuse Wirkungsevaluierung – Genese, Debatte<br>und Nutzung in der Entwicklungszusammenarbeit ....                          | 61  |
| <i>Petra Kaps</i>                                       | Formen der Evaluierung experimenteller<br>Politikgestaltung .....                                                              | 81  |
| <i>Cornelia Schendzielorz</i><br><i>Martin Reinhart</i> | Die Regierung der Wissenschaft im Peer Review .....                                                                            | 101 |
| <i>Regine Paul</i>                                      | Analyse and rule? A conceptual framework for<br>explaining the variable appeals of ex-ante<br>evaluation in policymaking ..... | 124 |

## Abhandlungen

|                                                                           |                                                                                                                                                  |     |
|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Jens Weiß</i><br><i>Dino Schubert</i>                                  | Doppelte Reform mit geteiltem Erfolg:<br>Zur Wirkung der DOPPIK-Reform auf<br>Haushaltsführung und Verwaltungssteuerung in<br>den Kommunen ..... | 143 |
| <i>Svenja Bauer-<br/>Blaschkowski</i>                                     | Politik nach Präferenzlage oder<br>pragmatisches Problemlösen? Landesregierungen und<br>die Unterbringung von Asylbewerbern (1995-2016) ...      | 165 |
| <i>Christian Person</i><br><i>René Geißler</i>                            | Ein Fass ohne Boden? Vier Jahrzehnte<br>kommunaler Schuldenhilfen in Deutschland .....                                                           | 191 |
| <i>Sandra Stötzer</i><br><i>Michaela Stangl</i><br><i>Christian Brödl</i> | Auswirkungen von Gemeindestrukturreformen –<br>Eine Einzelfallstudie einer Gemeindefusion<br>in der österreichischen Steiermark .....            | 217 |

## Rezensionen

|                         |                                                                                                                            |     |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Manfred Röber</i>    | Vom Staat zum Markt. Privatisierung aus<br>politikwissenschaftlicher Sicht (Detlef Sack) .....                             | 241 |
| <i>Werner Jann</i>      | Handbuch Politikberatung (Svenja Falk,<br>Manuela Glaab, Andrea Römmele, Henrik Schober<br>& Martin Thunert (Hrsg.)) ..... | 245 |
| <i>Nadin Fromm</i>      | Public Administration in Times of<br>Transnationalisation and Crisis (Martina McCowan) ..                                  | 249 |
| <i>Julia Schwanholz</i> | Legitimität in unsicheren Zeiten.<br>Der demokratische Verfassungsstaat in der Krise?<br>(Alexander Thiele (Hrsg.)) .....  | 254 |



*Holger Straßheim, Oliver Schwab*

## Politikevaluation und Evaluationspolitik

### **Zusammenfassung**

Es gibt inzwischen kaum ein Politikfeld, in dem Evaluierung nicht in irgendeiner Form praktiziert wird. Die Evaluierungslandschaft zeichnet sich dabei sowohl durch Vereinheitlichungs- als auch Pluralisierungstendenzen aus. In unterschiedlichen Evaluierungsarrangements kommen spezifische Akteure, Praktiken und Standards zusammen. Zudem handelt es sich nicht um politikfreie Räume. Vielmehr sind diese „Varieties of Evaluation“ selbst Ausgangspunkt, Medium und Gegenstand von Politiken der Evaluation, also der Auseinandersetzung um legitime Formen der Politikbewertung. Der Themenschwerpunkt unternimmt eine Bestandsaufnahme, indem er die Formen und Folgen von Politikevaluation als Evaluationspolitik sowohl in theoretisch wie empirisch ausgerichteten Beiträgen nachzeichnet. In interdisziplinärer Absicht werden aktuelle Befunde der Politik- und Verwaltungswissenschaft, der Governance- und Policy-Analyse, der politischen Soziologie und der Soziologie der Bewertung vorgestellt.

**Schlagwörter:** Evaluierung, Evaluationspolitik, Wirkungen, Wertungen, Kausalität

### **Abstract**

*The Evaluation of Politics and the Politics of Evaluation*

Instruments and modes of evaluation are expanding across policy areas and jurisdictions. The emerging landscape can be characterized by dynamics of both standardization and pluralisation. Forming distinct arrangements of evaluation, particular actors, practices and standards are closely interlinked. In and between these “varieties of evaluation”, a multiplicity of actors such as policy makers, administrators, experts, consultants or stakeholders is struggling over legitimate modes of appraising and valuating public policy. This special issue brings together theoretical as well as empirical studies on the politics of policy evaluation. It presents findings from a broad interdisciplinary background including public policy and governance, comparative policy analysis, administration research, political sociology and the sociology of valuation and evaluation.

**Keywords:** Evaluation, Politics of Evaluation, Causality, Valuation, Impact

## 1 Einleitung

Verfahren der Evaluierung haben ursprünglich im Fahrwasser von New-Public-Management-Reformen, in den vergangenen zwanzig Jahren jedoch vor allem im Zuge der Bestrebungen einer „evidenzbasierten Politik“ auf nationaler, internationaler und transnationaler Ebene eine hohe Verbreitung erfahren (Pollitt & Bouckaert, 2011; Straßheim, 2017b; Vedung, 2008). Es gibt inzwischen kaum ein Politikfeld, in dem

Evaluierung nicht in irgendeiner Form praktiziert wird. Die Evaluierungslandschaft zeichnet sich dabei sowohl durch Vereinheitlichungs- als auch Pluralisierungstendenzen aus, in deren Ergebnis sich alleine in Deutschland ein großes Spektrum unterschiedlicher Arrangements<sup>1</sup> herausgebildet hat. Die Bandbreite reicht von der anlassbezogenen Evaluierung experimenteller Politik (z. B. die Evaluierung der Hartz-Reformen in der Arbeitsmarktpolitik, siehe auch Petra Kaps in diesem Band), über die Etablierung von Evaluierungsfunktionen innerhalb von Organisationen (z. B. bei der Evaluierung von Lehre in den Hochschulen oder den Peer-Review-Verfahren der Deutschen Forschungsgemeinschaft (DFG), siehe auch Cornelia Schendzielorz und Martin Reinhart in diesem Band) und die häufig praktizierte Evaluierung einzelner fachpolitischer Programme bis zur Einrichtung von Strukturen mit spezifischen Evaluierungskompetenzen (z. B. das Deutsche Institut für Entwicklungspolitik, siehe auch Jörg Faust in diesem Band). Mit den Arrangements variieren auch Organisationsformen, Ansätze, Methoden und Designs der Evaluierungen erheblich. Es ist davon auszugehen, dass wir es mit „Varieties of Evaluation“ (Schwab, 2017) zu tun haben, die sich in Wechselwirkung mit ihrem jeweiligen Kontext entwickelt haben und jeweils spezifische Akteurskonstellationen, Evaluierungspraktiken und Standards aufweisen. Diese Evaluierungsarrangements bilden keine politikfreien Räume. Vielmehr sind sie selbst Ausgangspunkt, Medium und Gegenstand von Politiken der Evaluation, also der Auseinandersetzung um legitime Formen der Politikbewertung (Bjørnholt & Larsen, 2014; Boswell, 2018; Bovens, Hart, & Kuipers, 2008; Stephenson, Schoenefeld & Leeuw, 2019; Weiss, 1993 [1973]).

Evaluierungsarrangements sind durch eine besondere Positionierung im Verhältnis zur Wissenschaft einerseits und zu Politik und Verwaltung andererseits sowie durch bestimmte Anforderungen gekennzeichnet (DeGEval, 2017): Zum einen sind Evaluierungen durch ein systematisches Vorgehen gekennzeichnet und empirisch fundiert, womit sie anschlussfähig an (fach-)wissenschaftliche Diskurse sind. Zum anderen sind sie auf eine Bewertung des Nutzens im Hinblick auf einen bestimmten Verwertungskontext oder der allgemeiner verstandenen Güte des Gegenstandes ausgerichtet, womit sie mit politisch-administrative Verfahren koppelbar werden.<sup>2</sup> Evaluierungen stellen Transparenz über die Bewertungskriterien her. Sie können sich auf unterschiedliche Gegenstände beziehen (Projekte, Programme, Politiken, Organisationen). Es ist diese besondere Positionierung zwischen Wissenschaft und Politik/Verwaltung, die Evaluierungsprozesse prägt und die Entwicklung von Evaluierungsarrangements für Einflüsse von beiden Seiten öffnet (Bovens, Hart & Kuipers, 2008). Jenseits der normativen Orientierung durch die DeGEval - Gesellschaft für Evaluation (DeGEval) als Fachgesellschaft ist in der Praxis die Verwendung des Begriffs „Evaluierung“ genauso vielfältig wie die sich herausbildenden Evaluierungsarrangements, und die begriffliche Auffächerung ist eng mit dieser Differenzierung verknüpft.

Derzeit gibt es – vor allem im deutschsprachigen Raum – wenig Forschung und damit wenig Erkenntnisse über die institutionelle und kulturelle Einbettung von Evaluationsarrangements einerseits und über die Formen und Folgen von Evaluationspolitiken andererseits. Dies jedoch wäre eine zentrale Voraussetzung, um im inter- und transnationalen Vergleich wie auch im Vergleich von Politikfeldern Aussagen über zukünftige Entwicklungen der Evaluierungslandschaft und über Bedingungen der Lernfähigkeit und Reflexivität von Politikevaluierungen und Evaluierungspolitiken treffen zu können. Unter anderem ergeben sich dabei folgende Forschungsfragen:

- Welche *Varieties of Evaluation* im Sinne distinkter Arrangements von Akteurskonstellationen, Evaluierungspraktiken und Standards lassen sich identifizieren? Durch welche Merkmale unterscheiden sich diese Evaluierungsarrangements? Handelt es sich um gebietskörperschaftliche Varianten, die etwa mit bestimmten Verwaltungskulturen einhergehen? Oder konstituieren sich die Varianten in bestimmten gesellschaftlichen Sektoren und Politikfeldern? Was sind die Mechanismen der Konstitution, Transformation oder auch des Verschwindens solcher Arrangements?
- Inwiefern erfüllt sich die Hoffnung, aus Evaluierungsverfahren Impulse für politische Lernprozesse zu gewinnen? Führen bestimmte Ausprägungen der Evaluationspolitik statt zu einer kritischen Beurteilung zu einer ‚Disziplinierung und Zählung‘ kritischer Perspektiven? Wohnt bestimmten Evaluationspolitiken eine Tendenz der Verhüllung von Politikversagen und nicht intendierten Nebeneffekten inne? Unter welchen Bedingungen erhellen bzw. verdunkeln Evaluationen das politische Verständnis von komplexen Problemlagen?

Der Themenschwerpunkt unternimmt eine Bestandsaufnahme zu diesen Fragestellungen und Forschungslinien, indem er die Formen und Folgen von Politikevaluation als Evaluationspolitik sowohl in theoretisch wie empirisch ausgerichteten Beiträgen nachzeichnet. Er präsentiert in interdisziplinärer Absicht aktuelle Forschungen der Politik- und Verwaltungswissenschaft, der Governance- und Policy-Analyse, der politischen Soziologie und der Soziologie der Bewertung.<sup>3</sup> Diese Einführung entwickelt zunächst auf der Grundlage der Beiträge und der einschlägigen Literatur mögliche Ansätze und Perspektiven einer Analyse von Politikevaluation als Evaluationspolitik. Dazu werden in einem ersten Schritt Grundlagen und Konzepte von Evaluation als politischer Praxis vorgestellt. In einem zweiten Schritt sollen dann Erkenntnisse über die Ausprägung und den Wandel der *Varieties of Evaluation* zusammengetragen werden. In einem dritten Schritt stellen wir knapp die Beiträge vor. Abschließend skizzieren wir Dimensionen und Typen der Politisierung und des Politischen in Evaluationen. Diese reichen von einer öffentlich und kontrovers auf der „Vorderbühne“ ausgetragenen Evaluationspolitik über deren symbolische oder sogar simulative Aufführung, einer „stillen“ Evaluationspolitik auf der „Hinterbühne“ bis hin zu einer Entpolitisierung von Evaluationen.

## 2 Evaluation als politische Praxis

Vor nunmehr 47 Jahren plädierte Carol Weiss (Weiss, 1993; 1993 [1973]) in ihrem für ein ganzes Genre der Forschungsliteratur vorausweisenden Artikel „Where Politics and Evaluation Research Meet“ dafür, Evaluationen als einen politischen Prozess zu begreifen. Sie wendet sich damit gegen ein rational-instrumentelles Modell der Trennung zwischen Evaluation und Entscheidung, das Evaluationen als Orte der Produktion von objektiven, politikfreien Fakten auffasst, auf deren Grundlage politische Entscheidungen optimiert werden können (dazu auch Bovens, Hart & Kuipers, 2008). Evaluationen sind jedoch laut Weiss aus drei Gründen mit Politik eng verwoben: Erstens bilden die Objekte von Evaluationen – Politikprogramme, Instrumente und andere Interventionen – gewissermaßen Epizentren politischer Spannungen, Kontroversen und Verhandlungen.

gen (1993 [1973], pp. 94-97). In ihnen manifestieren sich die Positionen konkurrierender, um Einfluss ringender Akteure mit teilweise weitreichenden sozialen und politischen Folgewirkungen. Über die Lebenszeit eines Programmes hinweg geraten Evaluationen direkt in das Kräftefeld dieser Auseinandersetzungen. So divergieren unter Umständen die Interpretationen der Programmziele zwischen Politik und Verwaltung einerseits und der Evaluierenden andererseits. Mitunter kommt es zu Zielkonflikten. Veränderungen der Programmgestaltung oder -ausrichtung ordnen das politische Kräftefeld neu und sorgen so dafür, dass sich die an der Evaluation Beteiligten mit geänderten Erwartungskonstellationen konfrontiert sehen. Alice O’Connors (2001) historische Analyse sozialpolitischer Evaluationen in den USA zeigt auf eindrückliche Weise, wie spätestens mit den *Welfare*-Reformen in den 1980er Jahren Expertinnen<sup>4</sup> und Evaluierende von einem Wandel des politischen und ideologischen Klimas gleichsam überholt und teilweise marginalisiert wurden – einem Wandel, dem sie allerdings zuvor durch einen rivalisierenden Wettbewerb um die einflussreicheren Evaluationsmethoden („analytic warfare“) den Weg bereitet hatten.

Zweitens finden die Ergebnisse von Evaluationen in Entscheidungsprozessen häufig keine oder eine nur schwer nachverfolgbare Verwendung (Beck & Bonß, 1984; Caplan, Morrison & Stambaugh, 1975; Weiss, 1980; 1993 [1973], pp. 97-99). Das liegt zum einen an den Zeithorizonten und Rhythmen politischer Entscheidungen und Verfahren, die von denen der Evaluationsforschung abweichen. Wahl- oder Budgetzyklen erzeugen einen Handlungs- und Rechtfertigungsdruck, dem sich politische Akteure kaum entziehen können und der sich nicht ohne weiteres mit den Abläufen von Evaluationsprojekten synchronisieren lässt. Zum anderen folgen politisch-administrative Prozesse eigenen Relevanzen und Prioritäten, die mit darüber bestimmen, ob und in welcher Form Evaluierungsergebnisse zur Geltung kommen. Um resonanzfähig zu sein, müssen Evaluationsberichte in die Präferenzhierarchien und Darstellungslogiken von Ministerien oder Behörden übersetzt werden. Außerdem können im Zuge von geänderten legislativen oder administrativen Interessenlagen die ursprünglichen Ziele und Bewertungskriterien in einem anderen Licht erscheinen. Der Evaluation selbst wird dann möglicherweise eine geringere (oder eine höhere) Bedeutung zugemessen als erwartet. Evaluationen sind auch deswegen notorisch mit „Übersetzungsproblemen“ konfrontiert (Schwab, 2013).

Weiss hebt jedoch noch einen dritten Aspekt hervor, der sich nicht auf den politischen Kontext – den Gegenstand oder die Wirkung – von Evaluationen bezieht und den eigentlichen Kern ihrer Argumentation ausmacht: Evaluationen sind implizit politisch (Weiss, 1993 [1973], pp. 100-106). Weiss überschreitet damit die Grenze zwischen Evaluation und Entscheidung und verortet das Politische im Evaluieren selbst (zur Überwindung des Trennungsmodells Jasanoff, 2011). Es liegt in der Natur von Geltungsaussagen in Evaluationen, dass sie immer gleichzeitig auch eine politisch-wertende Haltung („political stance“, Weiss, 1993 [1973], p. 94) einnehmen. Durch die Bewertung einiger Programme als problematisch und anderer als erfolgreich, durch die Akzeptanz bestimmter Programmziele und die Hinnahme ihrer Prämissen wirken Evaluationen legitimierend. Die Fokussierung auf spezifische Kontextfaktoren und Wirkungen einer Intervention blendet andere Problemdimensionen und Strukturbedingungen aus. Was als methodische Notwendigkeit ausgegeben wird, hat möglicherweise weitreichende politische Implikationen, weil es die Aufmerksamkeit und Wahrnehmungsweisen politischer Akteure durch Schemata strukturiert (dazu Scott, 1998 und

Regine Paul in diesem Band). Das ist unvermeidbar und sollte nicht zur Suche nach neuen Technologien und Methoden der Neutralisierung und Objektivierung verleiten. Vielmehr, so schlussfolgert Weiss, sollte es um eine Sensibilisierung für die politische Dimension des Evaluierens gehen, damit für alle Beteiligten die politischen Implikationen und Nebenfolgen des Evaluierens sichtbar, kommunizierbar und strategisch bearbeitbar werden (zu politiksensiblen Evaluationskonzepten Bovens, Hart & Kuipers, 2008; Fischer, 1999).

Auch die neuere kritische Politikforschung greift auf dieses grundlegend politische Verständnis von Evaluationen zurück, wenn sie sich – wie vor mehr als vier Dekaden auch Weiss – gegen ein nach wie vor wirksames rational-instrumentelles Evaluationsverständnis wendet (Adelle, Jordan & Turnpenny, 2012; Bjørnholt & Larsen, 2014; Bovens, Hart & Kuipers, 2008; Stephenson, Schoenefeld & Leeuw, 2019). Vor allem der Aufschwung der evidenzbasierten Politik seit Ende der 1990er Jahre hat zu einer (Wieder-)Belebung des kybernetischen Modells geführt, wonach Evidenzen als augenscheinlich überzeugende und verlässliche Wissensbestände so systematisch in den Politikkreislauf einzuspeisen sind, dass es zu einer kontinuierlichen Verbesserung und Wirkungssteigerung politischer Maßnahmen und Entscheidungen kommt (Straßheim, 2018a). Demgegenüber gelten Evaluationen im kritischen Strang der Forschung als inhärent politischer Akt – oder besser eine Verkettung politischer Akte – in denen sich Wertungen und Beurteilungen manifestieren (Bovens, Hart & Kuipers, 2008, p. 1).

Evaluationen weisen in mehrfacher Hinsicht Charakteristika eines politischen Handelns auf, das auf kollektiv verbindliche Ordnung zielt und dazu (Auswahl-)Entscheidungen trifft (Fischer & Gottweis, 2012; Nullmeier & Pritzlaff, 2009): Erstens sind Evaluationen eng mit kollektiv verbindlichen Regelungen und Programmen verzahnt, die erst in der Ausgestaltung des Verfahrens, den Prozeduren und Methoden, den Berichts- und Bewertungskriterien ihre konkrete Umsetzung erfahren. Dabei kommt es immer auch zu Abweichungen, Neubewertungen und Ausnahmetatbeständen. Evaluationen ermöglichen deswegen eine meist mikroskopische, mitunter aber pfadabweichende Rekalibrierung existierender Regel- und Programmbestände, die in sich selbst bereits den Anspruch kollektiver Verbindlichkeit trägt und deswegen auch besonders zäh umstritten sein kann – selbst wenn es scheinbar nur um Kleinigkeiten geht (Vedung, 2008). Zweitens können sich in Kontroversen über Zieldefinitionen und Indikatorenkonstruktionen sehr unterschiedliche Auffassungen über das Verständnis beispielsweise von Arbeitsmärkten, Strukturwandel oder Leistungserbringungen widerspiegeln. Wer sich durchsetzt, bekräftigt in der Konsequenz eine spezifische politische Perspektive und schafft so mögliche Anknüpfungspunkte für die Rechtfertigung (oder Kritik) legislativer oder administrativer Entscheidungen (Fischer, 2003). Drittens definieren Evaluationen mit der Festlegung von Zielgruppen, der Unterteilung von Interventions- und Kontrollgruppen, der Bestimmung einer sozialen Ebene, auf der Wirkungen gemessen werden und der Annahme über Wirkungsbeziehungen und -bedingungen die (Kollektiv-)Subjekte politischen Handelns. Damit verbunden sind oft Zuweisungen von Handlungsverantwortungen. Im Aufstieg der randomisierten Kontrollstudien in der Armuts- und Wohlfahrtsstaatsforschung sehen beispielsweise Kritiker eine Verschiebung politischer Verantwortungszuschreibungen von der strukturellen Ebene auf die Ebene der individuellen Betroffenen (so bereits Straßheim, 2018b und Faust in diesem Band; Weiss, 1993 [1973], p. 105). Evaluationen setzen sich insofern aus Sequenzen von einzelnen Handlungen oder Kommunikationen zusammen, die in wertender Weise

Regelungs-, Kollektivitäts-, Verbindlichkeits- und Konfliktbezüge herstellen, bekräftigen oder mit Akzeptanz versehen. Evaluation ist in diesem Sinne eine inhärent politische Praxis (Nullmeier & Pritzlaff, 2009, S. 13-16).

Wie Schendzielorz und Reinhart in diesem Band anhand von Peer Review-Verfahren zeigen, können Evaluationen durch die sequentielle Verklammerung von Wertungs- und Geltungsbehauptungen eine hohe Akzeptanz und Autorität erreichen. Sie bilden eine Art Scharnier zwischen evaluatorischen und epistemischen Aussagen, indem sie verschiedenartige Praktiken der Wertzuschreibungen und -abwägung in eine Infrastruktur von Evaluationsprozeduren einbetten und mit politisch-administrativer, aber zugleich auch fachlicher Gültigkeit ausstatten. Evaluationen profitieren insofern von der „Legitimation durch Verfahren“ (Luhmann, 1989 [1969]).<sup>5</sup> Natürlich stehen Evaluationsergebnisse im politischen Prozess unter Vorbehalt, werden angezweifelt oder bestritten. Es gibt auch gute Gründe, anzunehmen, dass der permanente Druck zur Performanzmessung und zum Leistungsvergleich langfristig die Autoritäts- und Vertrauensbasis von Evaluationen in der Öffentlichkeit erodieren lässt (Boswell, 2018; Straßheim & Kettunen, 2014).

Die Fabrikation von Vertrauen ist insofern fragil: Evidenzen können niemals aus sich selbst heraus überzeugen, sondern müssen in Bezug zu einem laufenden Argumentations- und Bewertungsprozess gesetzt werden (Majone, 1989). Akteure verweisen in ihren Einschätzungen aufeinander (oder grenzen sich voneinander ab), werfen ihre Reputation in die Waagschale oder berufen sich auf Professionsstandards. Evaluationsberichte und *Policy Papers* bilden nicht nur eine Entscheidungs- und Bewertungsgrundlage, sondern gewinnen auch ein symbolisches Potential, indem sie für den Abschluss eines im Nachhinein nur schwer rekonstruierbaren Prozesses der Verhandlung von Definitionen, des Streits um Schlussfolgerungen und des Feinschliffs von Formulierungen stehen. Wenn sie dieses symbolische Potential erreichen, dann entwickeln sich bestimmte Evaluationsverfahren wie etwa *Peer Reviews* zu regelrechten Erfolgsgeschichten, die zur Nachahmung anregen und über den konkreten Anwendungskontext hinaus „Meta-Stabilität“ erreichen (Schendzielorz und Reinhart in diesem Band). Auf solche Evaluationstechniken wird dann immer wieder und in sehr unterschiedlichen Kontexten zurückgegriffen und gerade durch ihren symbolischen Gehalt zu einem gewissen Grade gegen Kritik abgesichert; zugleich erfahren sie je nach institutionellen und organisatorischen Rahmenbedingungen eine sehr unterschiedliche Anwendung.

### 3 Varieties of Evaluation

In bemerkenswerter Übereinstimmung zeichnen die Beiträge dieses Bandes das Bild einer expandierenden Evaluationslandschaft, die sich einerseits durch Professionalisierungstendenzen und Vereinheitlichungen auszeichnet, andererseits jedoch durch eine breite Vielfalt in der Institutionalisierung und Umsetzung von Evaluationsverfahren und -praktiken (insbesondere die Beiträge von Wolfgang Meyer und Reinhard Stockmann wie auch Thomas Widmer).

Hellmut Wollmann und andere sehen die Entwicklung dieser Landschaft durch drei Wellen gekennzeichnet (Stern, 2009; Wollmann, 2003): Schon Ende des 19. Jahrhunderts hatten sich Allianzen zwischen wissenschaftlichen und sozialpolitischen Akteuren gebildet, die dann die Legitimations- und Wissensbasis für spezifische Vorstellun-

gen über die Leistungsfähigkeit, Interventionstiefe und Extensivität von Wohlfahrtsstaaten lieferten (Bulmer, 1991; Wittrock, Wagner, & Wollmann, 1991). Die Expansion des Wohlfahrtsstaates in den 1960er und 1970er Jahren (erste Welle) ist nicht denkbar ohne die massive Pluralisierung solcher Arrangements der Politikberatung und -evaluierung und jener als „epistemic drift“ bezeichneten Bewegung, die wissenschaftliche und politische Qualitäts- und Relevanzkontrollen immer stärker miteinander verzahnte (Elzinga, 1985; Wagner, Hirschhorn Weiss, Wittrock & Wollmann, 1991). Eine zweite Welle von Evaluationen seit dem Ölpreisschock Mitte der 1970er Jahre war vor allem durch Budgetkürzungen und Kostensenkungen motiviert. Die „modelling wars“ zwischen US-Denkfabriken in den späten 1970er und frühen 1980er Jahren waren das Resultat dieser zweiten Expansionswelle. Sie mündeten in den Bemühungen von Experten und Beratungseliten, durch die Bewaffnung mit Mikrosimulationen, Modellierungen und ersten randomisierten Kontrollstudien die Wirksamkeit Kosteneffizienz staatlicher, vor allem sozialstaatlicher Maßnahmen in großem Umfang zu bewerten und so einen Anteil an der „poverty research industry“ zu gewinnen (O'Connor, 2001). Die dritte Welle von Evaluationen in den späten 1980er Jahren im Zuge der *New-Public-Management-Reformen*, seit Ende der 1990er Jahren eng verbunden mit der „evidence based policy“-Bewegung, führte zu einer verstärkten Institutionalisierung und einer systematischen, durch Leitlinien gestützten Verbreitung entsprechender Instrumente (Krapp, Pannowitsch & Heinelt, 2015; Straßheim, 2018a).

Wollmanns historische Analyse wäre aus heutiger Sicht durch eine vierte Welle zu ergänzen, welche seit mindestens drei Dekaden die ersten drei Wellen begleitet, verstärkt und partiell überrollt: Eine zunehmende Zahl inter- und transnational agierender Organisationen, Expertennetzwerke und -gemeinschaften prägen die Expansion von Evaluationen (Adelle & Weiland, 2013; Stephenson, Schoenefeld & Leeuw, 2019; Stockmann, Meyer & Taube, 2020 und Stockmann & Meyer in diesem Band; Stone, 2019; Straßheim, 2017b). Dazu gehört zum einen die Evaluationspolitik der Europäischen Union und hier insbesondere der Kommission. Durch die Implementierung von Evaluationsinstrumenten wie etwa *Impact Assessments* seit 2002, die Entwicklung eines Evaluationssystems zur Wirkungskontrolle, der „Smart Regulation“- und „Better Regulation“-Agenda, durch Expertennetzwerke, Training, Leitlinien und Standards unterstützt, beeinflusst und forciert sie entsprechende Bemühungen in den Mitgliedstaaten (Eußner & Blomeyer, 2020; Stern, 2009). Als besonders einflussreich hinsichtlich der Stärkung von Evaluationsverfahren hat sich die Kohäsionspolitik erwiesen, wo durch die direkte Einbindung nationaler Behörden im Rahmen der geteilten Mittelverwaltung Impulse besonders gut zu vermitteln sind.<sup>6</sup> Des Weiteren sorgen internationale Organisationen wie die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) oder die Weltbank in Form von *Peer Reviews* nationaler Evaluationen, durch eigene Leitlinien und Vergleiche guter Praktiken für eine grundsätzlich freiwillige, in ihrer Bindungswirkung jedoch nicht zu unterschätzende Verbreitung von Evaluationen (Morth, 2005; Radaelli, 2004). Hinzu kommen Expertennetzwerke, epistemische Gemeinschaften, Think Tanks, philanthropische Organisationen, Consulting-Firmen, transnationale Rechtsanwaltskanzleien und eine Vielzahl weiterer Organisationen an den Schnittstellen von Politik, Wissenschaft, Wirtschaft und Zivilgesellschaft (Botzem, 2012; Quack, 2016; Stone, 2012, 2019; Straßheim & Loer, 2019). Mit ihnen formiert sich eine „globale Agora“ (Stone, 2012) der Evaluationspolitik, die in ihren ebenenübergreifenden Zusammenhängen und Auswirkungen derzeit noch zu wenig

verstanden ist. Zuletzt haben Arno Simons und Jan-Peter Voss (Béland, Howlett, & Mukherjee, 2018; Simons & Voss, 2018) darauf hingewiesen, dass zunehmend auf der inter- und transnationalen Ebene mit Akteurskonstellationen (sogenannten „instrument constituencies“) zu rechnen ist, die in der Produktion, Fabrikation und Verbreitung von Evaluationsinstrumenten ihre kollektive Aufgabe sehen – und dies weniger, um Lösungen für Probleme bereitzustellen, als vielmehr im Sinne John Kingdons (1984) mit dem Ziel, für die von ihnen entwickelten Lösungen passende Probleme zu finden.

Die sich herausbildende Evaluationslandschaft ist divers und umfasst sehr verschiedenartige, voneinander teilweise abgrenzbare, teilweise einander beeinflussende und sich überschneidende Arrangements. Grundsätzlich lassen sich drei Perspektiven der Analyse und des Vergleichs dieser Varieties of Evaluation unterscheiden (Leeuw & Furubo, 2008; Schwab, 2017; Stephenson, Schoenefeld & Leeuw, 2019):

(1) Ansätze des Vergleichs von Regierungs- und Governancesystemen betonen die prägende Kraft politisch-administrativer Institutionen, die Rolle föderaler Ordnungen für die Einheitlichkeit von Evaluationen, den Einfluss des Wahlsystems auf evaluationspolitisch relevante Akteurskonstellationen und Politikstile wie auch die Verwaltungs- und Rechtstraditionen (Berg-Schlosser & Müller-Rommel, 1991; Murswieck, 1994; Rueschemeyer & Skocpol, 1996; Welch & Thompson, 1980). In einer policyanalytischen Variante dieses Ansatzes treten die strukturellen Fähigkeiten und Kapazitäten („policy analytical capacity“) von Regierungen und Verwaltungen in den Blick (Howlett, 2009). Diese umfassen die Fähigkeiten von Regierungen, Evaluationsforschung und eigene Evaluationskapazitäten etwa in Ressortforschungseinrichtungen bereitzustellen, komplexere statistische Methoden und Evaluationsinstrumente zu entwickeln und eine entsprechende Professionalisierung zu ermöglichen, Interessengruppen und andere politische Akteure in Evaluationsprozeduren einzubeziehen, die Ergebnisse in Form eigener Leitlinien zu reflektieren, sie innerhalb des politisch-administrativen Systems und in unterschiedlichen Ressorts zu implementieren, zu kontrollieren und letztlich auch in der Öffentlichkeit zu kommunizieren. Da diese strukturellen Merkmale in den verschiedenen Politikfeldern unterschiedlich ausgeprägt sind, erklärt dies die Entstehung und Differenzierung sehr verschiedener Evaluierungsarrangements innerhalb nationaler Ordnungen.

Die Ergebnisse der Analysen von Wolfgang Meyer und Reinhard Stockmann wie auch von Thomas Widmer in diesem Band sind auch in dieser Hinsicht äußerst aufschlussreich, weil sie sowohl auf international vergleichender wie nationaler bzw. subnationaler Ebene den Einfluss von Regierungssystemen stark relativieren. Dies lässt sich anhand des Schweizer Paradoxons (siehe Widmer in diesem Band) illustrieren, einem System, dem aufgrund der fragmentierten Staatsstrukturen, des Milizsystems und der Rückkopplungsfunktion der direkten Demokratie traditionell eine geringe Evaluierungskapazität und ein ebenso geringer Evaluierungsbedarf zugeschrieben wurde – und das dennoch als Evaluationspionier einen ungewöhnlich hohen Grad der Institutionalisierung und Implementierung von Evaluationen aufweist.

(2) Ansätze der vergleichenden Analyse von Akteurskonstellationen und -gemeinschaften sind konzeptuell ähnlich angelegt wie Forschung zu epistemischen Gemeinschaften und advokatorischen Koalitionen (Haas, 2013; Leeuw & Furubo, 2008; Sabatier & Weible, 2014; Wittrock, Wagner & Wollmann, 1991). Jan-Eric Furubo und Frans Leeuw gehen in ihrem Konzept der Evaluationssysteme davon aus, dass Evaluierungsar-



rangements sich genau dann ausdifferenzieren, wenn (a) Regierungs- oder Nichtregierungsorganisationen andere Organisationen mit der Umsetzung von Evaluationen beauftragen oder sie zumindest darin fördern, (b) sich in diesem Zuge ein kollektiv geteiltes Glaubenssystem („shared epistemology“) über die Gründe, Formen und Ziele von Evaluationen stabilisiert, (c) diese Allianz auf Dauer angelegt ist und (d) sich ihre Aktivitäten auf die Nutzung der Ergebnisse und Informationen aus Evaluationen in politischen Entscheidungs- und Implementationsprozessen beziehen (Leeuw & Furubo, 2008, pp. 159-160). Im Ergebnis konfigurieren sich aus dieser Sicht die *Varieties of Evaluation* also nicht entlang von Räumen oder Systemen des Regierens, sondern entlang von epistemisch vergemeinschafteten Akteurskonstellationen. Dementsprechend lassen sich Evaluationssysteme durch ihre paradigmatische, auf gemeinsamen Annahmen beruhende Verknüpfung spezifischer Verfahren, Instrumente und Akteure voneinander abgrenzen.

Systeme der Performanzmessung etwa sind eng mit der Entwicklung des *New Public Managements* verbunden, werden auf nationaler Ebene durch Ministerien und Audit-Organisationen gefördert und auf internationaler Ebene durch die OECD unterstützt. Dabei spielt die Verdichtung auf ausgewählte Kennzahlen und die damit einhergehende Verringerung und teilweise Ausblendung von Komplexitäten eine wichtige Rolle. Das dahinter stehende Glaubenssystem verbindet Annahmen über die anreizfördernde Wirkung von Performanzanalysen und -vergleichen mit der Behauptung, dass politische und gesellschaftliche Akteure die daraus resultierenden Informationen tatsächlich zur Optimierung von Entscheidungen nutzen und dabei nichtintendierte Nebeneffekte prinzipiell kontrollierbar sind (Bjørnholt & Larsen, 2014; Leeuw & Furubo, 2008, p. 161). In einem gewissen Kontrast dazu stehen Systeme der (quasi-)experimentellen Evaluation, die randomisierte Kontrollstudien als „Goldstandard“ betrachten und in einem mittlerweile weltumspannenden Netzwerk von Expertenorganisationen wie der Campbell Collaboration oder dem Abdul Latif Jameel Poverty Action Lab die Wirkungsanalyse von Interventionen auf der Mikroebene betreiben (Donovan, 2018 und Faust in diesem Band): Sie fokussieren stark auf Verhaltensweisen und Mechanismen, die auf Ebene der einzelnen Untersuchungseinheiten greifbar werden. Mit dieser Fokussierung blenden sie noch stärker als Systeme der Performanzmessung Komplexitäten und Zusammenhänge aus. Der Kontrast zu Systemen der Performanzmessung wird dann stärker, wenn die RCT-Bewegung sich mit verhaltensökonomischen Expertenzirkeln verbindet und auf dieser Grundlage Annahmen über rationales, informiertes Entscheiden durch experimentelle Erkenntnisse über verzerrte Rationalitäten ersetzt werden (Straßheim, 2017a). Petra Kaps zeichnet in ihrem Beitrag für diesen Band nach, wie sich im Zuge der Hartz-Reformen advokatorische Koalitionen der experimentellen Politikgestaltung herausbilden. Dabei hängt die Ausprägung dieser Koalitionen und ihre Durchsetzungsfähigkeit in Deutungskämpfen besonders von der institutionellen Ausgestaltung des Evaluationsverfahrens, der Umstrittenheit des jeweiligen Deutungssystems wie auch vom Organisationsgrad und der Machtposition der beteiligten Akteure ab.

(3) Schließlich konzentrieren sich Ansätze der Evaluationskulturen auf die mehr oder weniger unhinterfragten Praktiken und Prozeduren, die damit verbundenen impliziten Normen, Prämissen und Wertordnungen und die darauf beruhenden Infrastrukturen und Institutionen des Evaluierens (vgl. für unterschiedliche Herangehensweisen Wid-

mer und Schendzielorz & Reinhart in diesem Band). Ortwin Renn (1995) hatte bereits Mitte der 1990er Jahre darauf hingewiesen, dass sich Länder wie die USA, Großbritannien oder Deutschland in ihren Stilen der Politikberatung und -evaluation stark unterscheiden. In seiner Klassifikation unterscheidet er zwischen einem „adversarial“ bzw. „pluralist style“ der öffentlichen Wissensproduktion und Politikbewertung wie in den USA, einem „embodied“ bzw. „service-based style“, der wie in Großbritannien durch ministerielle „in house“-Einrichtungen und enge Konsultationsbeziehungen geprägt ist, und einem „corporatist style“ hochgradig institutionalisierter Beziehungen zwischen verschiedenen gesellschaftlichen Interessengruppen, die in Enquetekommissionen, parlamentarischen Ausschüssen und Verhandlungsrunden zu Erkenntnissen und Bewertungen gelangen (z. B. Deutschland, Schweden, teilweise die Niederlande). Daran anschließend argumentiert Sheila Jasanoff (2005; 2011), dass Arrangements der Expertise und Evidenzproduktion in engem Zusammenhang mit (national-) spezifischen Epistemologien stehen – institutionalisierten und weitgehend selbstverständlichen Praktiken der Produktion öffentlich relevanten Wissens, die aus dem historischen Zusammenspiel sozio-technischer, wissenschaftlicher und politischer Kulturen hervorgegangen sind und mit deren Hilfe die Mitglieder einer Gesellschaft bestimmen, ob bislang ungeprüfte Geltungsansprüche im Prozess kollektiv bindender Entscheidungen Berücksichtigung finden sollen (Straßheim, 2013).

Die vergleichende Analyse von Evaluationskulturen und Epistemologien kann erheblich zum Verständnis jener Selbstverständlichkeiten beitragen, mit denen ganz bestimmte Modi der Evaluation in einigen Kontexten öffentliche Akzeptanz finden, in anderen Kontexten dagegen nicht. Das lässt sich beispielsweise gut an der Debatte über experimentelle und verhaltensökonomisch informierte Politiken nachvollziehen, die in Großbritannien und auch in den USA Zustimmung erfahren, während sie in Deutschland auf heftigen Widerstand in einer breiten Öffentlichkeit stoßen (Straßheim, Jung, & Korinek, 2015; für den Bereich der Entwicklungspolitik den Beitrag von Faust in diesem Band). Und dennoch haben Typologien wie die von Renn oder Jasanoff nur begrenzte Aussagekraft, weil sie gleichsam gefrorene Wissenslandschaften präsentieren. Das zeigen die Beiträge in diesem Band übereinstimmend ganz deutlich: Die *Varieties of Evaluation* unterliegen einem hochgradig dynamischen Wandel, der auch für interne Differenzierungen und Konflikte sorgt. So kommen Meyer und Stockmann zu dem Schluss, dass die Ausprägung von Evaluationen weniger durch die institutionellen Rahmenbedingungen der politischen Systeme oder strukturelle Besonderheiten gekennzeichnet seien als vielmehr durch Entwicklungsprozesse und Reformen in der Regierungspraxis – eine durch Wandel und Diversität gekennzeichnete Situation, die auch Widmer für die Schweiz konstatiert. Deutlich einflussreicher scheinen – so zeigen auch die Beiträge von Widmer und Kaps – politikfeldspezifische Unterschiede und Akteurskonstellationen zu sein.

Die drei genannten Ansätze der *Varieties of Evaluation* schließen einander nicht aus, sie bedürfen aber der wechselseitigen Befruchtung und der konzeptuellen Weiterentwicklung – insbesondere unter den Bedingungen der „postnationalen Konstellation“ (Habermas, 2001). Praxistheorien und Ansätze der neueren Wertungs- und Bewertungssoziologie erscheinen hier als besonders vielversprechend, um Evaluationspolitiken in ihrer Dynamik und Vielgestaltigkeit zu erfassen (Lamont, 2012; Ringel & Werron, 2019; Schendzielorz & Reinhart in diesem Band).

## 4 Die Beiträge

*Wolfgang Meyer* und *Reinhard Stockmann* haben in einer umfassenden Bestandsaufnahme die Institutionalisierung von Evaluierung in Europa untersucht. Sie analysieren die formale Verankerung in Legislative und Exekutive sowie die Nutzung von Evaluationsergebnissen. Die institutionelle Verankerung wie auch die Nutzung von Evaluierungen variiert europaweit stark, ohne dass sich ein einheitliches Entwicklungsmodell herauskristallisieren würde. Insgesamt ist die Institutionalisierung in den Niederlanden und der Schweiz am weitesten fortgeschritten. Für die Institutionalisierung der Evaluierung in den nationalen Systemen spielen deren Strukturmerkmale nur eine geringe Rolle. Wichtiger ist die Regierungspraxis und hier vor allem das Ausmaß, in dem die Regierungen Ideen aus *New Public Management* und evidenzbasierter Politik aufgreifen. Die Parlamente treten als Treiber der Entwicklung und Nutzer der Evaluierung nur im Ausnahmefall prominent in Erscheinung. Nicht für die Institutionalisierung, aber für die Professionalisierung im Bereich Evaluierung ist außerdem die Größe des Landes von Bedeutung. Größe begünstigt die Entwicklung von Fachcommunities. Übergreifend treten Trends zur Verbreitung und Professionalisierung, aber auch teils nachlassende Dynamik gerade in den Ländern mit relativ weit entwickelten Evaluierungssystemen und eine Art Ausfransung von Evaluierungsverfahren in Richtung Controlling und Audit sowie eine generelle Entwertung von Evidenz zutage.

Wo Wolfgang Meyer und Reinhard Stockmann ein breites Spektrum im Ländervergleich in den Blick nehmen, betrachtet *Thomas Widmer* den Fall der Schweiz in der Tiefe und fragt auf Grundlage der Ergebnisse eines umfassenden Forschungsprogramms nach Wechselwirkungen zwischen Politik und Evaluierung. Die Verankerung der Evaluationsfunktion ist in der Schweiz auf Bundesebene stärker als auf Kantonsebene, wo sich vor allem Unterschiede zwischen Politikfeldern, weniger zwischen Kantonen zeigen. Hinter der hohen Zahl formal verankerter Evaluationsklauseln stehen neben regionalen Differenzierungen auch inhaltlich unterschiedliche Ausgestaltungen. In den für die Schweiz typischen Verfahren direkter Demokratie scheinen Evaluationsergebnisse insgesamt wenig Eingang zu finden; sie können sich aber auf die Diskursqualität auswirken, wenn auch eher über Expertinnen als über die breite Bevölkerung. Die parlamentarische Nutzung ist deutlich erkennbar und geht über die reine Kontrollfunktion hinaus, ist aber insgesamt wenig systematisch. Evaluationsprozesse werden eher genutzt als Evaluationsergebnisse. Im Hinblick auf die Rolle institutioneller Faktoren bestätigt die Analyse der Schweiz im Wesentlichen den internationalen Vergleich: Institutionen können nur in Teilen das Bild erklären, das sich insbesondere hinsichtlich der Praxis von Evaluierungsarrangements zeigt. Auch wenn zu keinem anderen Land so systematische Analysen zur Entwicklung der Evaluierungskultur vorliegen, tun sich hier weiter vielfältige Forschungsfragen auf. Widmers Apell, die Heterogenitäten und insbesondere die Erklärungsfaktoren für die Herausbildung unterschiedlicher Evaluationskulturen besser zu verstehen, kann somit auch für die internationale Perspektive gelten.

Mit *Jörg Fausts* Beitrag zur Bedeutung rigoroser Wirkungsevaluierung in der Entwicklungszusammenarbeit wird eine sektorale Perspektive eingenommen. Damit treten auch die Besonderheiten des konkreten Politikfeldes in den Vordergrund – dort haben Meyer und Stockmann sowie Widmer wichtige Einflussfaktoren für die Gestaltung von Evaluierungsarrangements vermutet. Faust zeigt, dass einerseits grundlegende Eigenschaften der Entwicklungszusammenarbeit die Rolle von Evaluierungen in diesem Po-

litikfeld stärken und das sie dort in besonderem Maße ihre Erkenntnis-, Lern- und Kontrollfunktion zeigen kann. An den Auseinandersetzungen um die Rolle rigoroser Wirkungsevaluierungen, die Faust strukturiert nachzeichnet, wird die Rolle von Einflussfaktoren stärker konturiert, die bereits in den vorangegangenen Beiträgen anklangen: Verschiedene inhaltliche Diskussionslinien entlang derer der Stellenwert rigoroser Wirkungsevaluierungen unterschiedlich bewertet wird, treten zu Tage. Aber auch die Rolle verschiedener Akteursgruppen, insbesondere auch der internationalen Geberorganisationen sowie die Abhängigkeit von Eigenschaften des Evaluierungsgegenstandes und dem Evaluierungsansatz werden neben weiteren thematisiert.

*Petra Kaps* konzentriert sich in ihrem Beitrag auf Experimente in der deutschen Arbeitsmarktpolitik und deren Evaluierung. Anhand der Kombilohn-Experimente (1998-2005), der Experimente zur Pauschalierung von Sozialhilfeleistungen (1999-2004) und des Experiments zur Erprobung von Eingliederungsmodellen (2005-2008) geht sie den *Varieties of Evaluation* nach. Sie verfolgt dabei eine auf Akteurskonstellationen bezogene Perspektive: Auf der Grundlage des Advokaten-Koalitionen-Ansatzes lotet sie insbesondere die politischen Deutungskämpfe und den Einfluss der institutionellen Verfahrensgestaltung aus. In der Tat kann Kaps nachweisen, dass die institutionelle Einbettung – die Dezentralisierung des Evaluationsauftrags, der Grad der Institutionalisierung des Evaluationsauftrags, aber auch die Eigenschaften des zu evaluierenden Programmes und die Verfügbarkeit von Datenquellen – einen Einfluss auf die Einheitlichkeit und interne Abstimmung von Evaluierungsarrangements haben. Zudem unterliegen die Advokaten-Koalitionen im Evaluierungsprozess unterschiedlich heftigen politischen Deutungskämpfen, die jeweils von der institutionellen Ebene, der Umstrittenheit des jeweiligen Politikprogrammes sowie dem Organisationsgrad und der Machtposition der handelnden Akteure abhängen.

Aus einer bewertungssoziologischen Perspektive erschließen *Cornelia Schendzielorz* und *Martin Reinhart* jene spezifischen Praktiken, die in Peer-Review-Verfahren zur Stabilisierung von *Varieties of Evaluation* führen. Zudem leuchten sie die politische Dimension dieser Arrangements aus, indem sie die Steuerungs- und Legitimationsfunktion dieser Verfahren rekonstruieren und damit auch die wissenschaftliche (Selbst-)Regierung durch *Peer Reviews*. Die empirische Grundlage bildet eine fallvergleichende Analyse von *Peer Reviews* im Rahmen der Förderung von Sonderforschungsbereichen (SFB) durch die DFG, im Rahmen von Zeitschriften und im Rahmen von Berufungsverfahren. Als Zwischenbefund identifizieren Schendzielorz und Reinhart typische Bewertungspraktiken, die als elementare Bestandteile von *Peer Reviews* sowohl die Binnendifferenzierung als auch prozedurale Anpassungsfähigkeit von *Peer Reviews* ausmachen. *Peer Reviews* entwickeln „Meta-Stabilität“, gerade weil sie Unsicherheiten und auch die Vielfalt von Personenkonstellationen, Arbeitsteilung und Rollentrennung durch Prozessualität und Sequenzialität kanalisieren und reduzieren – all das unter einem symbolisch vereinheitlichenden Label. Im Zusammenhang damit zeigen Schendzielorz und Reinhart auch, dass die in diesen Verfahren ablaufenden, mitunter widersprüchlichen und nicht kohärenten Kriterien zuzuordnenden Bewertungspraktiken sich auf die Vereinheitlichung berufen können, die *Peer Reviews* gleichsam als die „Stimme der Wissenschaft“ leisten. So ermöglichen *Peer Reviews* Adaption wie auch Legitimation nach innen wie außen. Sie arbeiten wie ein Scharnier, das „wissenschaftliche und politische Qualitäts- und Legitimationserfordernisse ineinander überführt und füreinander kompatibel macht“ (Schendzielorz und Reinhart).

Diese politische Multifunktionalität von Evaluationen adressiert auch *Regine Paul* in ihrem konzeptuellen Beitrag zu Ex-ante-Evaluationen, der dieses Themenheft abschließt. Darin entwickelt sie in intensiver Auseinandersetzung mit der Policy-Forschungsliteratur eine interpretative Perspektive, die Evaluationen als institutionell eingebettete Instrumente begreift, die auf selektive Weise Deutungshorizonte nicht nur erschließen, sondern politische Deutungen und Erwartungskonstellationen überhaupt erst ermöglichen. Im Zentrum von Pauls Beitrag steht eine Typologie, die idealtypisch drei verschiedene, mit Ex-ante Evaluationen verbundene Logiken ausweist und damit unterschiedliche Thesen der Literatur zu Policy-Instrumenten miteinander verknüpft: Unter Umständen kann dasselbe Instrumentarium je nach Kontextbedingungen instrumentelle, legitimatorische oder machbezogene Funktionen übernehmen. Paul führt damit nicht nur Literaturen unterschiedlicher Provenienz zusammen. Vielmehr zeigt sich gerade in den Kombinationen scheinbar gegenläufiger Rationalitäten die Vielgestaltigkeit der Evaluationspolitiken.

## 5 Evaluationspolitik auf der Vorder- und Hinterbühne

Überblickt man die hier versammelten Beiträge, so erscheinen Evaluationen in der Tat als eine in sich politische Praxis, die durch Widersprüchlichkeiten und Differenzierungsbewegungen gekennzeichnet ist und doch ein erhebliches Expansionspotential aufweist. Evaluationen sind deswegen inhärent politisch, weil mit ihnen Behauptungen über kollektive Verbindlichkeit und durchaus nicht unumstrittene, potentiell konflikt-hafte und immer kontingente, auch anders mögliche Wertungen verbunden sind. Akzeptiert man diese dem vorliegenden Themenschwerpunkt zugrunde liegende These, dann kann man behelfsmäßig und sicher nur vorläufig zu einer Systematisierung von Evaluationspolitik in ihren Dimensionierungen und Erscheinungsformen gelangen. Diese Systematisierung kann je nach Theorie der Politik und des Politischen auch anders ausfallen (Bedorf & Röttgers, 2010; Marchart, 2010; Nullmeier & Pritzlaff, 2009; Zürn, 2018).

Wir gehen hier davon aus, dass Politik mindestens zwei Dimensionen aufweist: Sie erschließt immer einen Raum des auch anders Möglichen, trifft explizit oder implizit eine Auswahl und ist insofern zumindest potentiell umstritten und konflikthaft. Dort, wo Notwendigkeitskonstruktionen und Automatismen den Diskurs bestimmen, verschwindet Politik (Greven, 2010). Zum anderen realisiert sich Politik in einer Öffentlichkeit, die in sich hochgradig divers sein kann und doch zumindest vorläufig durch Politik als ein Kollektivsubjekt des gemeinsamen Handelns und Kommunizierens erfahrbar wird (Arendt, 2017 [1950]; Rancière, 2002). Je begrenzter diese Öffentlichkeit und das potentiell relevante Akteursspektrum, desto geringer der Spielraum für Politik. Im Ergebnis postulieren wir zwei Dimensionen – Konfliktpotential und Akteursspektrum –, die uns Aufschluss über mögliche Erscheinungsformen von Evaluationspolitik geben (siehe auch Zürn, 2013).

Tabelle 1: Erscheinungsformen der Evaluationspolitik

|                   |                | Akteursspektrum                                  |                                                    |
|-------------------|----------------|--------------------------------------------------|----------------------------------------------------|
|                   |                | <i>breit</i>                                     | <i>eng</i>                                         |
| Konfliktpotential | <i>hoch</i>    | (1)<br>Evaluationspolitik auf der Vorderbühne    | (2)<br>Evaluationspolitik auf der Hinterbühne      |
|                   | <i>niedrig</i> | (3)<br>Symbolische/simulative Evaluationspolitik | (4)<br>Entpolitisierung von und durch Evaluationen |

Quelle: Eigene Darstellung.

In *Tabelle 1* sind diese Erscheinungsformen abgetragen. Wir unterscheiden zwischen (1) Evaluationspolitik auf der Vorderbühne, (2) Evaluationspolitik auf der Hinterbühne, (3) symbolischer bzw. simulativer Evaluationspolitik und (4) einer Entpolitisierung der Evaluationspolitik:

(1) Evaluationspolitik findet auf der Vorderbühne statt, wenn wissenschaftliche Expertinnen, Akteure aus Parlament, Regierungen und Ministerien, zivilgesellschaftliche und wirtschaftliche Akteure öffentlich um gemeinsame Bewertungsgrundlagen und -kriterien ringen und die Konsequenzen von Evaluationsergebnissen für kollektiv verbindliche Entscheidungen ausloten (Bovens, Hart & Kuipers, 2008; zur Metapher der Vorder- und Hinterbühne Goffman, 1959; Hilgartner, 2000). Solche im eigentlichen Sinne politischen Evaluationen sind mitunter konfliktreich und lösen häufig Kaskaden von Folge- und auch Gegenevaluationen aus (Bovens, Hart & Kuipers, 2008). Als „öffentliches Drama“ (Hilgartner, 2000) liefern sie wissenschaftlichen und politischen Akteuren eine Bühne für den Gewinn (oder Verlust) von Autorität und Reputation, ermöglichen Deliberation und Austausch über fundamentale Werte und Normen (Fischer, 1999) – sind aber auch der Ort für Machtasymmetrien, Verantwortungsverschiebungen und „blame games“ (Hood, 2007). Insofern handelt es sich bei Evaluationspolitik auf der Vorderbühne auch keineswegs um ein Ideal herrschaftsfreier Kommunikation, sondern eher um einen Spielraum für vielerlei Varianten politischer Praxis.

(2) Evaluationspolitik auf der Hinterbühne betrifft demgegenüber nur einen schmalen Akteurskreis im Arkanbereich parlamentarischer und administrativer Verhandlungsarenen, in Expertennetzwerken und eher geschlossenen Akteurskonstellationen (Bovens, Hart & Kuipers, 2008). Diese Form der Evaluationspolitik unterliegt möglicherweise den Evaluationszyklen und -rhythmen von Politikprogrammen, sie kann aber auch im Zuge politischer Krisen, der Untersuchung möglicher Fehlentscheidungen, Neuwahlen oder Rücktritten auftreten und ist insofern potentiell konfliktreich. Im Grunde ist sie der Normalfall bei Verhandlungen über Indikatoren und Verfahrensstandards; insofern handelt es sich auch um eine Notwendigkeit effektiver Evaluationspraxis und eine unweigerliche Begleiterscheinung der Professionalisierung und Institutionalisierung von Evaluationen. Einige Befunde, die auf die starken Unterschiede von Evaluierungsarrangements zwischen Politikfeldern hinweisen (siehe Meyer und Stockmann sowie Widmer in diesem Band) legen die Vermutung nahe, dass die Ausgestaltung der Evaluierungsarrangements nicht nur von Seiten der Politik und Verwaltung, sondern auch von den jeweils mit den *Policies* verbundenen wissenschaftlichen Fachgemeinschaften

getrieben ist. Die Professionalisierung kann auch in technokratische Strukturen umschlagen (Fischer, 1990; Straßheim & Kettunen, 2014). Dies führt dann zu Wissensmonopolen, politischen Schematismen und mit großer Wahrscheinlichkeit zu jenen Formen des Politikversagens, die James Scott (Scott, 1998) in seiner Analyse von Regierungsprogrammen eindrucksvoll geschildert hat. Widmer betont in diesem Zusammenhang die Rolle von evaluationsfachlicher Reflexion in Ergänzung zu fachpolitischen Wissensbeständen.

(3) Evaluationen können für die politische Kommunikation durchaus eine wichtige symbolische Funktion ausfüllen, weil sie im Bestfall Orte des kollektiven Austauschs und der deliberativen Reflexion demokratischer und politischer Wertgrundlagen sind – selbst wenn die Ergebnisse nicht unmittelbar umsetzbar sind (Fischer, 2009). Wenn Evaluationspolitik jedoch zur „Evaluitis“ (siehe Meyer & Stockmann in diesem Band) und damit zu einem Ritual verkommt, das in kulissenhafter Weise lediglich die Rechtfertigung von Politikprogrammen sicherstellen soll, kann man von *symbolischer* Evaluationspolitik sprechen (im Sinne von Edelman, 1964). Evaluationen werden dann zu Maschinerien der Fabrikation von Vertrauen, wenn Rankings, Benchmarkings und Performanzmessungen in alle denkbaren Bereiche des öffentlichen Lebens ausgedehnt werden. In einer solchen „Audit Society“ (Power, 1997) findet schleichend eine Entfremdung statt, weil in paradoxer Weise immer mehr Menschen den Technologien der Bewertung ausgesetzt sind, ohne abweichende Interessen oder Vorstellungen artikulieren zu können (Boswell, 2018). Evaluationen, die tatsächlich politisch folgenreich sind, finden dann unter Umständen lediglich auf der Hinterbühne statt, die sorgfältig von der Öffentlichkeit abgeschirmt ist. Im Gegensatz dazu überschreitet *simulative* Evaluationspolitik die Grenzen zwischen Vorder- und Hinterbühne und wirkt kollektivierend (Blühdorn, 2013). Hier gibt es kein Hinterbünnenspiel mehr. Längst haben Bürger wie Eliten die Erwartungen an eine gemeinsame Wertgrundlage verloren. Evaluationen dienen nur noch der Simulation eines funktionierenden Gemeinwesens – ein postdemokratisches Spektakel, das auf allseitige Zustimmung und Beteiligung rechnen kann, weil – wie dies Ingolfur Blühdorn, Felix Butzlaff, Michael Deflorian, Mirijam Mock und Daniel Hausknost (2020) jüngst anhand der Umwelt- und Klimapolitik nachgewiesen haben – damit der wohlthuende Eindruck kollektiver Handlungsfähigkeit erzeugt werden kann, ohne die Prämissen (aber auch Paradoxien) einer spätkapitalistischen Konsumgesellschaft aufgeben zu müssen.

(4) Wenn Evaluationen weder eine Öffentlichkeit erreichen noch in irgendeiner Weise Anknüpfungspunkte für die Suche nach politischen Alternativen bieten, dann unterliegen sie Prozessen der Entpolitisierung. Ihre Ergebnisse bleiben marginal, sie laufen als Beiwerk eines behördlich-administrativen Apparats ab oder werden nur selektiv wahrgenommen (zur Nutzung von Evaluationsergebnissen Widmer und Faust in diesem Band). Evaluationen können jedoch auch selbst zu Mechanismen der Mobilisierung gesellschaftlicher Entpolitisierungsdynamiken werden. Ihre Ergebnisse werden in solchen Fällen zur Verleugnung von Alternativen, zur Konfliktreduktion und zur Behauptung von Entscheidungsnotwendigkeiten genutzt (Greven, 2010). Dies mag fallweise immer wieder vorkommen und es lässt sich vermuten, dass Entpolitisierungsstrategien in Politikfeldern mit einem besonders hohen Legitimationsdruck häufiger zu beobachten sind (Fischer, 2009). In den vergangenen Jahren wurden im Zeichen einer instrumentellen Regierungstheorie jedoch auch systematische Überlegungen angestellt, die auf eine

weitreichende Entpolitisierung und Entdemokratisierung politischer Systeme im Zeichen einer Stärkung der Wissenseiten hinauslaufen (Brennan, 2016; Min, 2015; Straßheim, 2020). In „Epistokratien“ sollen die Schwächen der Demokratie durch monopolisierte Bewertungs- und Leistungsanalysen ausgeglichen werden. Befürworter schlagen vor, Expertenkommissionen in bestimmten Bereichen ein Vetorecht gegenüber parlamentarischen Entscheidungen einzuräumen („epistocratic veto“), das Wahlrecht auf jene zu beschränken, die in Sachfragen ausreichend informiert sind („restricted suffrage“) oder Wählerstimmen je nach individuellem Abschneiden in einem Wissenstest zu gewichten („weighted voting“) (Brennan, 2016). Auch wenn solche Überlegungen wohl wenig Umsetzungschancen haben, zeigen sie doch, dass die Entpolitisierung von und durch Evaluationen gravierende Folgen haben kann. Es bleibt weiteren Forschungen vorbehalten, mögliche Zusammenhänge zwischen solchen Dynamiken der Entpolitisierung und Entdemokratisierung einerseits und populistischen Tendenzen (Meyer & Stockmann in diesem Band) andererseits auszuloten.

Welche Konsequenzen haben diese Beobachtungen für die Evaluationspraxis? Frank Fischer (1999) hat vor mehr als zwanzig Jahren ein deliberatives Evaluationssystem entwickelt, das die öffentliche und konfliktorische Reichweite von Evaluationen stufenweise erhöht und immer auch die Reflexion auf die zugrundeliegenden politischen Ideen und kollektive Legitimation erzwingt. Mark Bovens, Paul t’Hart und Sanneke Kuipers (2008) suchen nach Wegen der Vermittlung zwischen programmatischen, effektivitätsbezogenen Evaluationen und politischen, die normativen Grundlagen betreffenden Evaluationen. In beiden Fällen geht es um die Vermeidung von Entpolitisierungen und um die Sicherung einer Pluralität von Kriterien des Evaluierens in einer möglichst breiten Öffentlichkeit (Majone, 1989, p. 183). Möglicherweise muss die Frage ähnlich beantwortet werden wie es jüngst Julia Hamann, David Kaldewey und Julia Schubert in einem preisgekrönten Aufsatz zur Relevanz von Wissenschaft getan haben (Hamann, Kaldewey & Schubert, 2019). Es gehe letztlich nicht um standardisierte Bewertungsverfahren, sondern um einen sich beständig ausweitenden „Pool“ von Bewertungsperspektiven, aus dem je nach Situation und Position ausgewählt werden könne – „Autonomie wäre dann, kurz gesagt, die Freiheit zur Heteronomie“ (Hamann, Kaldewey & Schubert, 2019, S. 27). In der Unübersichtlichkeit der Evaluationslandschaft liegt insofern auch eine Chance. Die vergleichende Erforschung der *Varieties of Evaluation* könnte, einem Observatorium ähnlich, diese Landschaft für die Praxis erschließen, die Vielfalt der Evaluationspolitiken sichtbar machen und so Autonomieresourcen bereitstellen (Haus, Korinek, & Straßheim, 2018). Dieser Band will dazu einen Beitrag leisten.

## Anmerkungen

- 1 Unter Arrangements verstehen wir hier ein Bündel von Aspekten, welche die Evaluierungspraxis in einem abgrenzbaren Gegenstandsbereich charakterisieren: Diese Arrangements umfassen institutionelle Aspekte, Akteursstrukturen, aber auch Präferenzen für bestimmte Ansätze und Methoden sowie epistemologische Grundannahmen. Unterschiedliche Perspektiven zur Analyse dieser Arrangements stellen wir Teil 3 dieser Einleitung vor.
- 2 Die Positionierung zwischen Wissenschaft und Politik/Verwaltung kommt auch darin zum Ausdruck, dass in den Standards für Evaluationen neben Genauigkeits- und Durchführbarkeitsstandards auch jeweils ein Block von Standards zu Fairness und Nützlichkeit definiert wird.



- 3 Die Beiträge gehen auf die zweitägige Konferenz „Politikevaluation und Evaluationspolitik“ am Wissenschaftszentrum Berlin für Sozialforschung am 17. und 18. November 2018 zurück. Wir danken den Teilnehmerinnen und Teilnehmern für rege Diskussionen und den Kommentatorinnen und Kommentatoren sowie den Gutachterinnen und Gutachtern für zahlreiche Anregungen. Unser Dank gilt zudem der Stiftung Mercator für ihre großzügige Unterstützung.
- 4 Aus Gründen der besseren Lesbarkeit wird im Folgenden alternierend die männliche und weibliche Sprachform verwendet. Selbstverständlich sind alle Geschlechter gleichermaßen gemeint.
- 5 Was auch dazu führt, dass die Behauptung, „evaluiert“ zu haben, zur Legitimation politisch nutzbar ist, ohne dass notwendigerweise tatsächlich Prozesse durchlaufen werden, die systematisch den Wert oder Nutzen eines Gegenstandes analysieren würden. Letztlich kann der Evaluationsbegriff hier politisch als Symbol zur Legitimationsbeschaffung verwendet werden (siehe auch den Beitrag von Paul in diesem Band).
- 6 Während ansonsten Evaluierungspraktiken stark mit den originär europäischen Politiken, wie etwa der Innovationspolitik verknüpft waren, hat in den letzten Jahren eine regulativ vermittelte Evaluierungsverpflichtung auch in anderen Politikfeldern, wie etwa der europäischen Beihilfenpolitik zu einer primärrechtlich gefassten und damit unmittelbar EU-weit geltenden Evaluierungspflicht geführt.

## Literatur

- Adelle, Camilla, Jordan, Andrew & Turnpenny, John (2012). Proceeding in parallel or drifting apart? A systematic review of policy appraisal research and practices. *Environment and Planning C: Government and Policy*, 30 (3), 401-415.
- Adelle, Camilla & Weiland, Sabine (2013). Policy assessment: the state of the art. *Impact assessment and project appraisal*, 30 (1), 25-33.
- Arendt, Hannah (2017 [1950]). Was ist Politik? In Ursula Ludz (Hrsg.), *Hannah Arendt: Was ist Politik? Fragmente aus dem Nachlass* (S. 9-12). München/Berlin: Piper.
- Beck, Ulrich & Bonß, Wolfgang (1984). Soziologie und Modernisierung. Zur Ortsbestimmung der Verwendungsforschung. *Soziale Welt*, 35 (3), 389-397.
- Bedorf, Thomas & Röttgers, Kurt (2010). *Das Politische und die Politik*. Frankfurt a.M.: Suhrkamp.
- Béland, Daniel, Howlett, Michael & Mukherjee, Ishani (2018). Instrument constituencies and public policy-making: an introduction. *Policy and Society*, 37 (1), 1-13.
- Berg-Schlosser, Dirk, & Müller-Rommel, Dieter (1991). *Vergleichende Politikwissenschaft*. Opladen: Leske + Budrich.
- Bjørnholt, Bente & Larsen, Flemming (2014). The politics of performance measurement: 'Evaluation use as mediator for politics'. *Evaluation*, 20 (4), 400-411.
- Blühdorn, Ingolfur (2013). *Simulative Demokratie. Neue Politik nach der postdemokratischen Wende*. Frankfurt a.M.: Suhrkamp.
- Blühdorn, Ingolfur, Butzlaff, Felix, Deflorian, Michael, Mock, Mirijam & Hausknost, Daniel (2020). *Nachhaltige Nicht-Nachhaltigkeit: Warum die ökologische Transformation der Gesellschaft nicht stattfindet*. Bielefeld: transcript.
- Boswell, Christina (2018). *Manufacturing Political Trust. Targets and Performance Measurement in Public Policy*. Cambridge: Cambridge University Press.
- Botzem, Sebastian (2012). *The Politics of Accounting Regulation. Organizing Transnational Standard Setting in Financial Reporting*. Cheltenham, UK/Northampton, MA: Edward Elgar.
- Bovens, Mark, Hart, Paul 't & Kuipers, Sanneke (2008). The Politics of Policy Evaluation. In Robert E. Goodin, Michael Moran, & Martin Rein (Eds.), *The Oxford Handbook of Public Policy* (pp. 319-335). Oxford.
- Brennan, Jason (2016). *Against Democracy*. Princeton: Princeton University Press.
- Bulmer, Martin (1991). National contexts for the development of social-policy research: British and American research on poverty and social welfare compared. In Peter Wagner, Carol Hirschhorn Weiss, Björn Wittrock, & Hellmut Wollmann (Eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads* (pp. 148-167). Cambridge: Cambridge University Press.

- Caplan, Nathan, Morrison, Andrea & Stambaugh, Russel J. (1975). *The Use of Social Science Knowledge in Policy Decision at the National Level. A Report to Respondents*. Ann Arbor: Ann Arbor Center for Research on the Utilization of Scientific Knowledge.
- DeGEval Gesellschaft für Evaluation e.V. (2017). *Standards für Evaluation*. Verfügbar unter: <https://www.degeval.org/degeval-standards/standards-fuer-evaluation/>.
- Donovan, Kevin P. (2018). The rise of the randomistas: on the experimental turn in international aid. *Economy and Society*, 47 (1), 27-58.
- Edelman, Murray J. (1964). *The Symbolic Use of Politics*. Illinois: Illini Books.
- Elzinga, Aant (1985). Research, Bureaucracy and the Drift of Epistemic Criteria. In Björn Wittrock & Aant Elzinga (Eds.), *The University Research System. The Public Policies of the Home of Scientists* (pp. 191-220). Stockholm: Almquist & Wiksell International.
- Eußner, Ansgar, & Blomeyer, Roland (2020). Evaluation Practice in the European Union. In Reinhard Stockmann, Wolfgang Meyer, & Lena Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 463-479). London: Palgrave.
- Faust, Jörg (2020). Rigorose Wirkungsevaluierung – Genese, Debatte und Nutzung in der Entwicklungszusammenarbeit. *dms – der moderne staat*, 13 (1), 61-80.
- Fischer, Frank (1990). *Technocracy and the Politics of Expertise*. Newbury Park/London/New Delhi: Sage.
- Fischer, Frank (1999). *Evaluating Public Policy*. Chicago: Nelson-Hall.
- Fischer, Frank (2003). *Reframing Public Policy*. Oxford: Oxford University Press.
- Fischer, Frank (2009). *Democracy & Expertise. Reorienting Policy Inquiry*. Oxford: Oxford University Press.
- Fischer, Frank & Gottweis, Herbert (Eds.). (2012). *The Argumentative Turn Revisited: Public Policy as Communicative Practice*. Durham/London: Duke University Press.
- Goffman, Erving (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Greven, Michael (2010). Verschwindet das Politische in der politischen Gesellschaft? Über Strategien der Kontingenzerleugnung. In Thomas Bedorf & Kurt Röttgers (Hrsg.), *Das Politische und die Politik* (S. 68-88). Frankfurt a.M.: Suhrkamp.
- Haas, Ernst B. (2013). Epistemic Communities. In Joel Krieger (Ed.), *The Oxford Companion to Comparative Politics (Volume 1)* (pp. 351-359). Oxford/New York: Oxford University Press.
- Habermas, Jürgen (2001). *The Postnational Constellation*. Cambridge: MIT Press.
- Hamann, Julian, Kaldewey, David & Schubert, Julia (2019). Ist gesellschaftliche Relevanz von Forschung bewertbar und wenn ja, wie? *Forschung und Gesellschaft*, 14, 13-27.
- Haus, Juliane, Korinek, Rebecca-Lea & Straßheim, Holger (2018). Expertise im Nexus. Von der Verwendungs- zur Vernetzungsforschung. In Nico Lüdtke & Anna Henkel (Hrsg.), *Das Wissen der Nachhaltigkeit. Herausforderungen zwischen Forschung und Beratung* (S. 63-88). München: Oekom Verlag.
- Hilgartner, Stephen (2000). *Science on Stage: Expert Advice as Public Drama*. Stanford (CA): Stanford University Press.
- Hood, Christopher (2007). What Happens When Transparency Meets Blame Avoidance. *Public Management Review*, 9 (2), 191-210.
- Howlett, Michael (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52 (2), 153-175.
- Jasanoff, Sheila (2005). *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton (NJ): Princeton University Press.
- Jasanoff, Sheila (2011). The Practices of Objectivity in Regulatory Science. In Charles Camid, Neil Gross, & Michèle Lamont (Eds.), *Social Knowledge in the Making* (pp. 307-338). Chicago/London: University of Chicago Press.
- Kaps, Petra (2020). Formen der Evaluation experimenteller Politikgestaltung. *dms – der moderne staat*, 13 (1), 81-100.
- Kingdon, J.W. (1984). *Agendas, Alternatives, and Public Policies* (2. ed.). Boston: Little Brown and Company.

- Krapp, Max-Christopher, Pannowitsch, Sylvia & Heinelt, Hubert (2015). *Wissenspolitik und politischer Wandel. Veränderung der deutschen und britischen Arbeitsmarktpolitik seit den 1990er Jahren*. Baden-Baden: Nomos.
- Lamont, Michèle (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38 (1), 201-221.
- Leeuw, Frans L. & Furubo, Jan-Eric (2008). Evaluation Systems: What Are They and Why Study Them? *Evaluation*, 14 (2), 157-169.
- Luhmann, Niklas (1989 [1969]). *Legitimation durch Verfahren* (2. Aufl. ed.). Frankfurt a.M.: Suhrkamp.
- Majone, Giandomenico (1989). *Evidence, Argument, and Persuasion in the Policy Process*. New Haven/London: Yale University Press.
- Marchart, Oliver (2010). *Die politische Differenz*. Frankfurt a.M.: Suhrkamp.
- Meyer, Wolfgang & Stockmann, Reinhard (2020). Institutionalisierung der Evaluation in den politischen Systemen Europas. Eine vergleichende Analyse. *dms – der moderne staat*, 13 (1), 24-43.
- Min, John B. (2015). Epistocracy and democratic epistemology. *Politics in Central Europe*, 11 (1), 91-112.
- Morth, Ulrika (Ed.) (2005). *Soft Law In Governance And Regulation: An Interdisciplinary Analysis*. Edward Elgar.
- Murswiek, Axel (1994). *Regieren und Politikberatung*. Opladen: Westdeutscher Verlag.
- Nullmeier, Frank & Pritzlaff, Tanja (2009). Zu einer Theorie politischer Praktiken. *Österreichische Zeitschrift für Politikwissenschaft*, 38 (1), 7-22.
- O'Connor, Alice (2001). *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History*. Princeton: Princeton University Press.
- Paul, Regine (2020). Analyse and rule? A conceptual framework for explaining the variable appeals of ex-ante evaluation in policymaking. *dms – der moderne staat*, 13 (1), 124-142.
- Pollitt, Christopher & Bouckaert, Geert (2011). *Public Management Reform: A Comparative Analysis – New Public Management, Governance, and the Neo-Weberian State*. Oxford: Oxford University Press.
- Power, Michael (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Quack, Sigrid (2016). Expertise and Authority in Transnational Governance. In Roger Cotterrell & Maksymilian Del Mar (Eds.), *Authority in Transnational Legal Theory. Theorising Across Disciplines*. (pp. 361-386). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Radaelli, Claudio M. (2004). The diffusion of regulatory impact analysis – Best practice or lesson drawing. *European Journal of Political Research*, 43 (5), 723-747.
- Rancière, Jacques (2002). *Das Unvernehmen*. Frankfurt a.M.: Suhrkamp.
- Renn, Ortwin (1995). Styles of using scientific expertise: a comparative framework. *Science and Public Policy*, 22 (3), 147-156.
- Ringel, Leopold & Werron, Tobias (Eds.) (2019). *Rankings – Soziologische Fallstudien*. Wiesbaden: VS Verlag.
- Rueschemeyer, Dietrich, & Skocpol, Theda (Eds.) (1996). *States, Social Knowledge, and the Origins of Modern Social Policies*. New Jersey: Princeton Univ. Press.
- Sabatier, Paul, & Weible, Christopher M. (Eds.) (2014). *Theories of the Policy Process (third edition)*. Boulder: Westview.
- Schendzielorz, Cornelia & Reinhart, Martin (2020). Die Regierung der Wissenschaft im Peer Review. *dms – der moderne staat*, 13 (1), 101-123.
- Schwab, Oliver (2013). Aus Daten Sinn machen. Übersetzungsprobleme im Performanzmanagement. *dms – der moderne Staat*, Sonderheft 1/2013 „Wissen und Expertise in Politik und Verwaltung“ (Hg. Sabine Kropp & Sabine Kuhlmann), 103-125.
- Schwab, Oliver (2017). Varieties of Evaluation? Zur Rolle von Evaluierung in politisch-administrativen Systemen. In Sabine Kuhlmann & Oliver Schwab (Hrsg.), *Starke Kommunen – wirksame Verwaltung. Fortschritte und Fallstricke der internationalen Verwaltungs- und Kommunal-forschung* (S. 199-225). Wiesbaden: VS Verlag.

- Scott, James C. (1998). *Seeing Like A State. How Certain Schemes to Improve the Human Condition Have Failed*. New Haven/London: Yale University Press.
- Simons, Arno & Voss, Jan-Peter (2018). The concept of instrument constituencies: accounting for dynamics and practices of knowing governance. *Policy and Society*, 37 (1), 14-35.
- Stephenson, Paul J., Schoenefeld, Jonas J. & Leeuw, Frans L. (2019). The Politicisation of Evaluation: Constructing and Contesting EU Policy Performance. *Politische Vierteljahresschrift*, 60, 663–679.
- Stern, Elliot (2009). Evaluation policy in the European Union and its institutions. *New Directions for Evaluation*, 123, 67-85.
- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (2020). *The Institutionalisation of Evaluation in Europe* London: Palgrave.
- Stone, Diane (2012). *Knowledge Actors and Transnational Governance: The Private-Public Policy Nexus in the Global Agora*. London: Palgrave Macmillan.
- Stone, Diane (2019). *Making Global Policy*. Cambridge: Cambridge University Press.
- Straßheim, Holger (2013). Politische Expertise im Wandel: Zur diskursiven und institutionellen Einbettung epistemischer Autorität. *dms – der moderne Staat, Sonderheft 1/2013 „Wissen und Expertise in Politik und Verwaltung“* (Hg. Sabine Kropp/Sabine Kuhlmann), 65-87.
- Straßheim, Holger (2017a). Die Globalisierung der Verhaltenspolitik. *Jahrbuch Normative und institutionelle Grundfragen der Ökonomik, Band 16: Kapitalismus, Globalisierung, Demokratie*, 211-242.
- Straßheim, Holger (2017b). Trends towards evidence-based policy formulation. In Michael Howlett & Ishani Mukherjee (Eds.), *Handbook of Policy Formulation* (pp. 504-521). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Straßheim, Holger (2018a). Policy as a body of expertise. In Hal K. Colebatch & Rob Hoppe (Eds.), *Handbook on Policy, Process and Governing* (pp. 89-108). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Straßheim, Holger (2018b). Verhaltenspolitik im Wohlfahrtsstaat: Zur Mikrofokussierung in der Sozialpolitik(forschung). *Sozialer Fortschritt*, 67 (8-9), 759-782.
- Straßheim, Holger (2020). De-biasing Democracy. Behavioural public policy and the post-democratic turn. *Democratization*, 27 (3 ), 461-476.
- Straßheim, Holger, Jung, Arlena & Korinek, Rebecca-Lea (2015). Reframing Expertise: The Rise of Behavioural Insights and Interventions in Public Policy In Ariane Berthoin Antal, Michael Hutter, & David Stark (Eds.), *Moments of Valuation. Exploring Sites of Dissonance* (pp. 249-268). Oxford: Oxford University Press.
- Straßheim, Holger & Kettunen, Pekka (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy*, 10 (2), 259-277.
- Straßheim, Holger & Loer, Kathrin (2019). Global Health Promotion: Who Are the Policymakers and What Are Their Interests. In Benjamin Ewert & Kathrin Loer (Eds.), *Behavioural Policies for Health Promotion and Disease Prevention* (pp. 73-92). London: Palgrave Macmillan.
- Vedung, Evert (2008). *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Wagner, Peter, Hirschhorn Weiss, Carol, Wittrock, Björn & Wollmann, Hellmut (Eds.) (1991). *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press.
- Weiss, Carol (1993). Politics and Evaluation: A Reprise with Mellow Overtones. *Evaluation Practice*, 14 (1), 107-109.
- Weiss, Carol (1993 [1973]). Where Politics and Evaluation Research Meet. *Evaluation Practice*, 14 (1), 93-106.
- Weiss, Carol H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 8 (2), 381-404.
- Welch, Susan & Thompson, Kay (1980). The Impact of Federal Incentives on State Policy Innovation. *American Journal of Political Science*, 24, 715-729.
- Widmer, Thomas (2020). Wechselwirkungen von Politik und Evaluation: Befunde aus der Schweiz. *dms – der moderne staat*, 13 (1), 44-60.

- Wittrock, Björn, Wagner, Peter & Wollmann, Hellmut (1991). Social science and the modern state: policy knowledge and political institutions in Western Europe and the United States. In Peter Wagner, Carol Hirschhorn Weiss, Björn Wittrock, & Hellmut Wollmann (Eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads* (pp. 28-85). Cambridge: Cambridge University Press.
- Wollmann, Hellmut (2003). Evaluation in public-sector reform. Trends, potentials and limits international perspective. In Hellmut Wollmann (Ed.), *Evaluation in Public-Sector Reform. Concepts and Practice in Internationale Perspective* (pp. 231-258). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Zürn, Michael (2013). Politisierung als Konzept der Internationalen Beziehungen. In Michael Zürn & Matthias Ecker-Ehrhardt (Hrsg.), *Die Politisierung der Weltpolitik. Umkämpfte internationale Institutionen* (S. 7-35). Frankfurt a.M.: Suhrkamp.
- Zürn, Michael (2018). *A Theory of Global Governance. Authority, Legitimacy, and Contestation*. Oxford: Oxford University Press.

### *Anschriften der Autoren:*

Prof. Dr. Holger Straßheim (Korrespondenzautor), Universität Bielefeld, Fakultät für Soziologie, PF 100131, 33501 Bielefeld, E-Mail: holger.strassheim@uni-bielefeld.de.

Dr. Oliver Schwab, IfS Institut für Stadtforschung und Strukturpolitik GmbH, Lützowstr. 93, D-10785 Berlin, E-Mail: schwab@ifsberlin.de.

Wolfgang Meyer, Reinhard Stockmann

# Institutionalisierung der Evaluation in den politischen Systemen Europas. Eine vergleichende Analyse

## Zusammenfassung

Der Beitrag präsentiert die ersten vergleichenden Ergebnisse eines weltweiten Forschungsprojekts zur Institutionalisierung der Evaluation. Mehr als 30 Autorinnen und Autoren haben sich mit der Situation in 16 Ländern Europas beschäftigt und diese anhand eines einheitlichen Analyseleitfadens untersucht. Bei der Betrachtung der Integration von Evaluation innerhalb des politischen Systems standen drei Aspekte im Vordergrund: Die formalen Verankerungen in der Legislative (z. B. in Form von Gesetzen und Verordnungen) und in der Exekutive (z. B. der Evaluationspraxis in verschiedenen Politikfeldern, Ministerien und Behörden) sowie die Nutzung der Evaluationsergebnisse für evidence-based policies.

Es handelt sich um die erste umfassende systematische Bestandsaufnahme der Institutionalisierung von Evaluation in den politischen Systemen Europas. Herausgearbeitet werden die Gemeinsamkeiten und Unterschiede sowie die unterschiedlichen Einflussfaktoren auf die nationalen Institutionalisierungsprozesse. In die Betrachtung gehen z. B. Spezifika nationaler politischer Systeme, die Einflüsse transnationaler Organisationen (insbesondere der Europäischen Union), die sektoralen Besonderheiten und ihre Leitfunktionen sowie die Rolle der Zivilgesellschaft mit ein.

**Schlagworte:** Institutionalisierung der Evaluation, Politik, Europa, systematischer Vergleich

## Abstract

*Institutionalization of evaluation in European political systems. A comparative analysis*

This paper presents the first comparative results of a global research project on institutionalization of evaluation. More than 30 authors investigated the situation in 16 European countries by using a consistent analytical framework. While observing the integration of evaluation into the political system, three key elements are in focus: the formal anchoring within legislation (e. g. in form of acts, laws and decrees) and within the executive (e. g. the practice of evaluation in various policy fields, ministries and public authorities) as well as the use of evaluation results for evidence-based policies.

This is the first systematic and comprehensive stock-taking of the institutionalization of evaluation within the European political systems. It carves out the similarities and differences (especially for the drivers) of the national institutionalization processes. Among others, the specifics of national political systems, the influence of transnational organizations (notably the European Union), the leading function of particular sectors and the role of civil societies will be mentioned.

**Keywords:** Institutionalization of Evaluation, Politics, Europe, systematic comparison

## 1 Einführung

Dieser Beitrag präsentiert zentrale Befunde zur Institutionalisierung der Evaluation in Europa, welche im Mai 2020 umfassend in einer Buchpublikation vorgestellt werden (Stockmann, Meyer & Taube, 2020a). Buch und Beitrag sind Teil eines globalen Forschungsprojektes Evaluation-GLOBE des Centrums für Evaluation (CEval) an der Universität des Saarlandes mit dem Ziel, den ersten weltweiten systematischen Überblick zu diesem Thema in einer vierbändigen Buchserie zu veröffentlichen. Der zweite Band über Nord- und Südamerika befindet sich ebenfalls bereits in der Endredaktion und soll Anfang 2021 erscheinen (Stockmann & Meyer, 2021). Der vorliegende Artikel zur Lage in Europa konzentriert sich ausschließlich auf die Darstellung der vergleichenden Ergebnisse zum politischen System und verzichtet zugunsten von ersten Erklärungsversuchen auf die im Buch zusätzlich dargestellte Implementierung in der Zivilgesellschaft sowie im System der Professionen. Er stellt damit zugleich den ersten und vorsichtigen Schritt von der reinen Deskription zu einer weitergehenden Analyse dar.

Das Gesamtprojekt ist sowohl hinsichtlich seiner ambitionierten geographischen Ausrichtung als auch hinsichtlich seines Publikationsumfangs bisher einmalig. Am ehesten vergleichbar ist es mit dem 2002 erschienenen „International Atlas of Evaluation“ (Furubo, Rist & Sandahl, 2002), der ebenfalls auf Basis eines einheitlichen Analysekonzepts und mit Hilfe von Expertenratings eine Beurteilung des Stands der Evaluation in einer Vielzahl von Ländern – vorrangig in Nordamerika und Europa – vorgenommen hat. Das GLOBE-Projekt orientiert sich jedoch nur bedingt an diesem Vorbild und bietet keine vergleichbaren, aktualisierten Befunde (siehe hierzu die Studie von Jacob, Speer & Furubo, 2015). Auf die theoretischen wie methodischen Grundlagen des GLOBEs insbesondere auch in Abgrenzung zum *International Atlas* und einigen weiteren Projekten zur Darstellung der Institutionalisierung von Evaluation wird im zweiten Abschnitt dieses Beitrags näher eingegangen.

Die Fokussierung auf die Institutionalisierung im politischen System begründet sich nicht nur durch das Konzept dieses Schwerpunkthefts und die inhaltliche Ausrichtung der Zeitschrift, sondern auch durch dessen besondere Rolle für die Herausbildung, Implementierung und Nutzung von Evaluation in Europa. Während sich in der Privatwirtschaft andere, in ihrer Zielrichtung und Wirkungsweise durchaus vergleichbare Konzepte herausgebildet haben (Stockmann, 2008), ist die Evaluation sehr eindeutig als „Kind der Politik“ zu verorten und kommt primär bei politischen Programmen und Projekten zur Anwendung (hierzu z. B. Pattyn, Voorst, Mastenbrock & Dunlop, 2018; Vedung, 2017). Hieraus ergeben sich einige Spezifika hinsichtlich der Institutionalisierung, die im Rahmen der Vorstellung der Ergebnisse im dritten Abschnitt näher herausgearbeitet werden.

Besonders bemerkenswert ist allerdings die Tatsache, dass weder die unterschiedlichen Charakteristika der politischen Systeme noch die historische Entwicklung die abweichenden Muster der Institutionalisierungsprozesse in den verschiedenen Ländern erklären können. Einige der zentralen Einflussfaktoren und deren Gewichtung werden in der vergleichenden Analyse im vierten Abschnitt vorgestellt.

Schließlich bleibt zum Schluss noch ein Ausblick auf die potentielle weitere Entwicklung der Evaluation in Europa, die in den letzten Jahrzehnten angesichts der exponentiellen Verbreitung in höchst unterschiedlichen Politikfeldern zweifellos als eine

beeindruckende Erfolgsgeschichte zu beschreiben ist, nun aber an einen kritischen Punkt kommt und eventuell zum Opfer des eigenen Erfolgs werden könnte. Der fünfte und letzte Abschnitt widmet sich diesen kritischen Aspekten und den notwendigen nächsten Schritten, die für eine weiterführende Verankerung von Evaluation in den politischen Systemen Europas von entscheidender Bedeutung sein werden.

## 2 Der GLOBE Europe – Konzept und Leitfragen

Ziel des GLOBE-Projektes ist ein systematischer, weltweit ländervergleichender Überblick zum Stand der Institutionalisierung von Evaluation. Die Entscheidung für die Nationalstaaten als Einheit des Vergleichs begründet sich durch ihre zentrale Rolle für die gesellschaftliche Steuerung mittels des nationalen politischen Systems (Hay, Lister & March, 2006). Dem entsprechend sind Gesetze oder andere Formen rechtlicher Grundlagen zur Evaluation Aufgabe des Nationalstaats, selbst wenn es ergänzende internationale, regionale oder sektorale Regelungen oder Steuerungsmodelle gibt (die dann allerdings durch nationale Regulationen abgesichert sein müssen, siehe hierzu die Beispiele in Ansell & Torfing, 2016). Auch hinsichtlich der Nutzung von Evaluationen kommt dem Nationalstaat eine Führungsrolle zu, da viele öffentliche Aufgaben von nationalen Ministerien und Behörden wahrgenommen werden. Hier unterscheiden sich die Staaten bezüglich der Verteilung von Zuständigkeiten und der Organisation dieser öffentlichen Aufgabenerfüllung erheblich voneinander – und dies gilt im besonderen Maße für Europa mit seinen historisch gewachsenen Staatsformen (Hroch, 2005) und z. T. grundsätzlich voneinander abweichenden politischen Systemen (Ismayr, 2009; 2010).

In den Europaband wurden insgesamt 16 Länderkapitel und ein Beitrag über die Europäische Union (EU) aufgenommen (*Abbildung 1*). Da in der Buchpublikation aus Platzgründen nicht alle europäischen Länder zu berücksichtigen waren, musste eine Auswahl vorgenommen werden. Im ersten Schritt sind vor allem kleinere Länder (z. B. Andorra, Luxemburg, Liechtenstein, Malta und San Marino) aufgrund ihrer geringen Größe und der Wahrscheinlichkeit, hier keinen eigenständigen nationalen Evaluationsmarkt vorzufinden, ausgeschlossen worden. Im Falle der baltischen Staaten wurde versucht, einen vergleichenden Beitrag zu initiieren, was aber an den großen Unterschieden hinsichtlich der Institutionalisierung der Evaluation zwischen den drei Ländern scheiterte. Mit Lettland wurde deshalb das Land mit der fortgeschrittensten Evaluationskultur für das Buch ausgewählt. Die Balkanregion sowie Belarus und die Ukraine wurden nach Vorgesprächen mit einigen nationalen Experten aufgrund des geringen Entwicklungsstands der Evaluation nicht aufgenommen. In vier für die Entwicklung der Evaluation in Europa wichtigen Ländern – Norwegen, Österreich, Schweden und Ungarn – konnten keine Autorinnen und Autoren gefunden werden, die in dem vorgegebenen Zeitraum einen Beitrag entsprechend der Vorgaben ausarbeiten konnten. Die vorliegenden Befunde für Europa sind somit nicht vollständig, umfassen aber die bedeutendsten Länder. Von den knapp 600 Mio. Bewohnern<sup>1</sup> Europas (Russland und die Türkei wurden hier nicht mitgezählt, die beiden Staaten werden nach Rücksprache in den später erscheinenden Australasia-Band aufgenommen) leben ca. 80% in den 16 erfassten Nationen.



Abbildung 1: Ausgewählte Länder Europas



Quelle: Eigene Darstellung.

Inhaltlich ist das „casing“ nicht so einfach, wie es auf den ersten Blick erscheint. Die Entscheidung für Nationalstaaten als Untersuchungseinheiten impliziert z. B., dass die sektorübergreifenden Gemeinsamkeiten innerhalb eines Staates größer sind als die länderübergreifende Homogenität innerhalb eines Handlungsfelds. Im Falle der Evaluation lässt sich diese Annahme aber durchaus bezweifeln (siehe z. B. die Ergebnisse der Studie zu den großen Unterschieden zwischen öffentlichem und privatem Sektor in Kanada von Cousins, Goh, Elliott, Aubrey & Gilbert, 2014). Als Folge wird der Steuerungseinfluss des Nationalstaats auf die Institutionalisierung von Evaluation innerhalb eines Landes tendenziell über-, die Bedeutung sektorspezifischer internationaler Organisationen, multilateraler Vereinbarungen und transnationaler Vernetzung dagegen eher unterschätzt.

Generell stellt die Fallauswahl in der international vergleichenden Forschung immer eine besondere Hürde dar, die auf verschiedene Weise genommen werden kann und jeweils mit unterschiedlichen Schwierigkeiten und Einschränkungen verbunden ist (Goerres, Siewert & Wagemann, 2019, pp. 83-93). Durch die hier getroffenen, eher pragmatisch orientierten Auswahlentscheidungen ergibt sich ein Selektionsbias in Richtung

Überbewertung der Institutionalisierung von Evaluation in Europa (insbesondere durch den Ausschluss von Ländern, in denen keine oder nur eine sehr rudimentäre Institutionalisierung vermutet wurden) und trotz dem bevölkerungsbezogen hohen Abdeckungsgrad kann weder statistisch noch inhaltlich von Repräsentativität gesprochen werden.

Das Ziel der Studie ist allerdings nicht die Repräsentation Europas, sondern die deskriptive Darstellung der zentralen Unterschiede zwischen Nationalstaaten innerhalb Europas. Es geht darum herauszufinden, ob und inwieweit die Institutionalisierung der Evaluation gleichmäßig verläuft und welche unterschiedlichen Entwicklungspfade festzustellen sind. Die Beschreibung historischer Abweichungen und nicht der Test elaborierter Theorien der Institutionalisierung stehen (als Überblick Peters, 2019) im Vordergrund. Gleichwohl orientiert sich die Auswahl der Indikatoren an neo-institutionalistischen Überlegungen und der Fokus wird auf drei verschiedene Elemente gesetzt (*Abbildung 2*):

- Erstens geht es um die Frage des legalen Rahmens und der vorhandenen Regulationen, die den Einsatz, die Form und die Nutzung von Evaluationen steuern.
- Zweitens steht die regelmäßige Einbindung des Instruments der Evaluation in die praktische Politik sowohl generell als auch in den verschiedenen Teilbereichen im Fokus.
- Drittens schließlich beschäftigen sich die Länderstudien mit der Nutzung von Evaluationen und ihren Ergebnissen bei der Planung, (Weiter-)entwicklung, Steuerung und Umsetzung von Politik durch die verschiedenen Beteiligten im politischen System.

*Abbildung 2:* Leitfragen zur Institutionalisierung von Evaluation im politischen System

| National Regulations on Evaluation                                                                                                                                                                                                                                                          | Evaluation Practice (Evaluation Policies)                                                                                                                                                                                                                                                                                                                   | Use of Evaluation within the Political System                                                                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Are there national laws or regulations about evaluation or use of evaluation? Are there sectoral laws or regulations about evaluation or use of evaluation (e.g. a law about school evaluation or evaluation in the higher education system as example for laws in the educational sector)? | Is it possible to speak of a frequent rhythm of evaluations, for instance for every new legislation or for every national program? Or is it rather non-specific?                                                                                                                                                                                            | Which sectors are "good performer", which are "bad performer" regarding use of evaluation and evaluation findings? Please describe up to 3 sectors that can be considered as leading and 3 sectors that are lagging behind in the field of evaluation's use.                                          |
| Are there policies or strategies about evaluation or use of evaluation, either national or sectoral? Are there administrative regulations about evaluation or use of evaluation in different policy fields (instructions, guidelines, etc.)?                                                | Does evaluation take place in all sectors/ policy fields of a country? And within one sector, is evaluation applied for measures funded in different ways or maybe only the ones that received funding by the European Union?                                                                                                                               | On which aspect do most evaluations focus in these sectors (e.g. Planning and Steering, Accountability and Legitimacy, Enlightenment)? Which professional groups use evaluation and evaluation findings regularly (e.g. political decision makers, program or project manager, administrative staff)? |
| Is evaluation and use of evaluation findings embedded in parliamentary structures? Do parliamentarians in your country deal with evaluation findings for their own political work? Do parliamentarians in your country demand evaluations for their own political work?                     | How would you describe the relation between internal and external evaluations? Which form is carried out more often and for what purposes? How would you describe the relation between process and impact/ outcome evaluations? Which form is used more often and for what purposes? Do independent internal departments exist, in ministries or elsewhere? | How is the quality of evaluations guaranteed (e.g. regular conduction of meta-evaluations analyses, competence requirements for evaluators, quality requirements for evaluations)?                                                                                                                    |

*Quelle:* Stockmann, Meyer & Taube, 2020a: Appendix (von den Autoren gekürzte Fassung).

Die eigentliche Datenerhebung erfolgte durch Expertinnen und Experten vor Ort, die angehalten waren, zur Abdeckung der Breite ihrer Informationen zusätzlich Interviews mit Ressourcenpersonen durchzuführen. Die Auswahl der Expertinnen und Experten erfolgte über die Netzwerke der lokalen Evaluationsgesellschaften (VOPEs = Voluntary Organisation of Professional Evaluation) sowie entsprechende Aufrufe bei der Europäischen Evaluationsgesellschaft (EES = European Evaluation Society) und während der Kick-Off-Veranstaltung bei der EES-Tagung in Maastricht, Niederlande, im Herbst 2016. Insgesamt konnten so 30 namhafte Expertinnen und Experten als Autorinnen und Autoren für den Band gewonnen werden.

Zur Gewährleistung der Vergleichbarkeit der Beiträge wurde ein für alle Autoren verpflichtendes Analyseraster sowie ein gemeinsam entwickeltes Glossar zur Verfügung gestellt. Die eingereichten Rohfassungen sind von den Herausgebern hinsichtlich der Vollständigkeit und Plausibilität der Aussagen geprüft und mit Bitte um Vervollständigung bzw. Schärfung der Argumentation zurückgegeben worden. Dadurch konnten natürlich nicht sämtliche Lücken geschlossen, unterschiedliche Gewichtungen oder Interpretationen jedoch weitgehend vermieden werden.

Bezüglich der Datenqualität lässt sich sagen, dass einige der erfassten Aspekte (z. B. die Gesetze und Regulationen zu Evaluationen) gut dokumentiert und nachvollziehbar belegt sind, andere jedoch auf Einschätzungen beruhen (z. B. die Verbreitung der Evaluationspraxis in den verschiedenen Evaluationsfeldern). Die Herausgeber haben sich durch die beschriebene Vorgehensweise bemüht, solche Einschätzungen soweit möglich zu homogenisieren. Die methodischen Schwierigkeiten sind aber primär durch den Stand der Institutionalisierung selbst bedingt: Nur dort, wo sich dauerhaft implementierte Institutionen herausgebildet haben, ist eine entsprechende Erfassung einfach und empirisch eindeutig machbar. Noch dominiert aber ein gewisser „Wildwuchs“, d. h. ein wenig geregelter und nicht dauerhaft festgelegter Einsatz der Evaluation im politischen System.

Erschwerend kommt hinzu, dass der Begriff Evaluation in den letzten Jahren zunehmend inflationär genutzt wird und in den verschiedenen europäischen Sprachen sich zusätzlich unterschiedliche Interpretationen bzw. Konnotationen herausgebildet haben. Eine Abgrenzung der Begrifflichkeiten ist dementsprechend nicht in allen Fällen trennscharf und eindeutig möglich, auch wenn mittels des Glossars zwischen den Beteiligten an dem Europaband ein allgemein geteiltes Verständnis gewährleistet werden konnte. Trotz aller Bemühungen konnte so letztlich nicht vollständig ausgeschlossen werden, dass in manchen Ländern Dinge als Evaluationen bezeichnet werden, die in anderen mit abweichenden Begriffen oder Konzepten erfasst sind. Auch diese Schwierigkeiten sind aber weniger der Methodik denn der Dynamik eines fortschreitenden Verbreitungs- und Institutionalisierungsprozesses geschuldet (zur Komplexität solcher Prozesse siehe vor allem die Ergebnisse der Organisationsforschung z. B. Barley & Tolbert, 1997; Lawrence, Winn & Deveraux, 2001; Yu, 2013).

Im Vergleich zu Jan-Eric Furubo, Ray Rist & Rolf Sandahl (2002) und auch den eigenen Vorarbeiten zur zukünftigen Entwicklung der Evaluation (Stockmann & Meyer, 2017a) ist sowohl die Länderauswahl als auch der Datenerhebungsprozess deutlich systematischer und kontrollierter erfolgt. Zwar beziehen auch hier sich die Befunde auf Experteneinschätzungen in ausgewählten Ländern, die Expertinnen und Experten waren aber gehalten, nicht nur ihre eigenen Einschätzungen widerzugeben, sondern die Befunde empirisch abzusichern und möglichst umfassend zu belegen. Im Unterschied

zu der eher willkürlichen Länder- und Themenauswahl bei Furubo, Rist und Sandahl (2002), die wichtige Aspekte nicht behandelt und andere kaum nachvollziehbar unterschiedlich gewichtet haben, werden die Fragen aus einem theoretisch fundierten Raster abgeleitet. Insgesamt wurde ein deutlich höherer Aufwand zur Gewährleistung eines einheitlichen Bewertungsmaßstabs und des wissenschaftlich überprüfbaren Belegs der getroffenen Einschätzungen betrieben.

### 3 Institutionalisierung der Evaluation im politischen System

Die zur Untersuchung der Institutionalisierung der Evaluation in den politischen Systemen Europas verwendeten Leitfragen beziehen sich auf drei verschiedene Aspekte, welche sowohl die Legislative als auch die Exekutive betreffen. Kollektives gesellschaftliches Handeln wird idealtypischer Weise durch die auf nationaler Ebene von der Legislative festgelegten (und von der Judikative kontrollierten) Regeln und eine strikt regelkonforme Sachbearbeitung durch Regierung und staatliche Verwaltung in der Exekutive gewährleistet. Kollektive Verhaltensänderungen werden dabei primär durch die Bildung neuer bzw. die Veränderung bestehender Institutionen erreicht. Diese bestehen nicht nur aus der Verabschiedung neuer Regeln in Form von Gesetzen oder Verordnungen, sondern auch aus deren Interpretation (z. B. mittels der Ausgestaltung durch die Rechtsprechung oder die praxisorientierte Umsetzung durch zuständige Verwaltungen) sowie der Herausbildung von Kontroll- und Vermittlungsmechanismen zur Überwachung regelkonformen Verhaltens und zur Sanktion abweichenden Verhaltens.

Der erste Themenblock befasst sich mit der formalen Einbindung von Evaluation durch Regulationen in der Legislative und den darauf basierenden Kontrollen des Regierungshandelns. Solche nationalen Regeln sind für die Sicherstellung eines kohärenten Verständnisses und Einsatzes des Instruments der Evaluation von größter Bedeutung. Dies hat sich auf supranationaler Ebene durch die Initiativen und Aktivitäten zentraler multinationaler Organisationen (EU, Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), Vereinte Nationen (UN), Weltbank) gezeigt (Meyer, 2019a).

Aus globaler Perspektive ist zunächst festzuhalten, dass es eine unüberschaubare Vielfalt von Regelwerken in Form von Standards, Richtlinien, Selbstverpflichtungen, Durchführungsempfehlungen etc. zu Evaluationen mit einem mehr oder weniger hohen Verpflichtungsgrad innerhalb von nationalen und transnationalen Organisationen, Netzwerken, Politikfeldern, Verwaltungen usw. gibt (Meyer, 2019b). Für die nationale Ebene hat Barbara Rosenstein (2013) 115 Staaten untersucht und in immerhin 20 Staaten eine Nationale Evaluationspolitik gefunden, die auf bestimmten formalen Regeln beruht.

In Europa gibt es lediglich drei Länder (Frankreich, die Niederlande und die Schweiz), die Evaluation auf höchster nationaler Ebene als allgemeingültiges Instrument gesetzlich verankert haben. Vier weitere Staaten (Deutschland, Finnland, Großbritannien und Lettland) sehen Evaluationen in verschiedenen Verordnungen und Dekreten als allgemein verbindlich vor (in Deutschland z. B. in den Bundes- und Landshaushaltsordnungen sowie der gemeinsamen Geschäftsordnung der Bundesministerien, siehe ausführlich in Stockmann & Meyer, 2017b). Weitaus häufiger sind aber spezielle Regelungen in einzelnen Politikfeldern verbreitet, die zumeist durch die zuständigen

Ressorts (Ministerien) oder die ihnen nachgeordneten Behörden erstellt und umgesetzt werden. Hier bleibt es den zuständigen Stellen weitgehend selbst überlassen, wie sie Evaluation verstehen, in welcher Form und für welche Zwecke sie eingesetzt wird und welche Qualitätsanforderungen sowohl an die Evaluation als auch an die geprüften Maßnahmen gestellt werden. Dies bedingt den Aufbau von Verwaltungseinheiten zur Beauftragung, Überwachung und Durchführung von Evaluationen, die es mit der Ausnahme von Portugal und Rumänien in allen untersuchten europäischen Staaten gibt. Selten sind dies allerdings auf Evaluationen spezialisierte Einheiten, zumeist wird diese Funktion mit anderen Aufgaben (z. B. Accounting, Controlling oder Monitoring) verknüpft und die Evaluation spielt in diesem Kontext eher eine untergeordnete Rolle. Eigenständige staatliche Evaluationsinstitute wie das Deutsche Evaluierungsinstitut DEval im Bereich der deutschen Entwicklungszusammenarbeit gibt es nur äußerst selten.

Zentrales Organ der Legislative ist das Parlament und zur wichtigsten Aufgabe gehört die Kontrolle der Exekutive (insbesondere der Regierung). Diese Kontrollfunktion wird zumeist durch die parteiübergreifende Arbeit der Abgeordneten in Ausschüssen ausgeübt, die neben der Rechtmäßigkeit auch die Effektivität des Regierungshandelns überwachen sollen. Darüber hinaus gibt es unabhängige Organe wie z. B. Rechnungshöfe und *Audit Offices*, welche das Regierungshandeln und die administrative Umsetzung von Gesetzen eigenständig oder im Auftrag der Legislative prüfen können.

Aus diesem Grund liegt es nahe, dass die Parlamentarier und Parlamentarierinnen sowie speziell die parlamentarischen Ausschüsse ein besonderes Interesse an Evaluationen und an einer Nutzung für diesen Zweck haben sollten. Dieses Interesse manifestiert sich z. B. in der Tatsache, dass sich 2014 Parlamentarierinnen und Parlamentarier weltweit in einem „Global Parliamentarians Forum for Evaluation“ (<https://globalparliamentarianforum.wordpress.com/>) zusammengeschlossen haben. U. a. wird hier der institutionelle Aufbau von Monitoring und Evaluationssystemen und die Entwicklung nationaler Evaluationspolitiken unterstützt. Bemerkenswert ist allerdings, dass Europa in diesem Forum mit Ausnahme der Schweiz keine Rolle spielt und es vor allem Parlamentarier und Parlamentarierinnen aus dem globalen Süden sind, die sich hier engagieren und die Verbreitung der Evaluation zu befördern suchen.

Die Ergebnisse des GLOBES bestätigen dies: Außer der Schweiz gibt es kein Land, in dem das Parlament Evaluationsergebnisse systematisch zur Kontrolle des Regierungshandelns einsetzt oder gar selbst Evaluationen beauftragt. Evaluationsbefunde sind nur in drei weiteren Ländern (Dänemark, Deutschland und den Niederlanden) in der parlamentarischen Arbeit und in den Debatten von Bedeutung. Einige weitere Länder verfügen über einen wissenschaftlichen Dienst, der sich allerdings höchst selten mit Evaluationen beschäftigt (Finnland, Irland, Italien und Polen). Deutlich mehr Einfluss auf die Etablierung der Evaluation haben allerdings die zumeist eigenständig agierenden Rechnungshöfe / *National Audit Offices*, die in einer Reihe von Ländern (Belgien, Dänemark, Finnland, Lettland, Niederlande, Portugal, Schweiz und Tschechien) nicht nur wie generell üblich Performance-Audits, sondern auch Policy-Evaluationen zur Wirkungskontrolle durchführen.

Diese Heterogenität und Diversität der legislativen Verankerung spiegelt sich in einer ebenso unübersichtlich wie vielfältigen Anwendungspraxis auf Seiten der Exekutive wider. Der Einsatz hängt dabei stark von den einzelnen Politikfeldern ab: In einigen Ländern – insbesondere in Mitteleuropa – wird die Evaluation fast ausschließlich dort genutzt, wo sie von den Evaluationsrichtlinien der EU verlangt wird. Auf der

anderen Seite ist nur in den Niederlanden der Einsatz von Evaluation in praktisch allen Politikfeldern üblich. Ansonsten gibt es einige Bereiche, in denen Evaluation überall eingesetzt wird (Beschäftigung, Bildung, Entwicklungszusammenarbeit, Forschung und Entwicklung sowie Gesundheit) und andere (Außenpolitik, Innenpolitik, Justiz, Kultur und Landwirtschaft), in denen dies nur selten geschieht.

Dies führt allerdings innerhalb der einzelnen Sektoren nicht zu einer gleichartigen Nutzung in verschiedenen Ländern, im Gegenteil kennzeichnet Europa eine extreme Heterogenität hinsichtlich der Art, des Umfangs, der Häufigkeit und der Intensität der Anwendung von Evaluationen. Generell ist ein regelmäßiger, gleichförmiger und standardisierter Einsatz gegenüber dem punktuellen und den jeweiligen Anforderungen entsprechender, individueller Gebrauch die Ausnahme. Selbst in den Sektoren, in denen es auf internationaler Ebene Vereinbarungen und Vorgaben zum Einsatz von Evaluationen gibt, unterscheidet sich die Praxis auf nationaler Ebene sehr stark. Die Intensität des Evaluationseinsatzes nimmt dabei mit der formalen und legalen Verankerung auf nationaler Ebene zu, wobei allerdings unklar bleibt, ob die Verankerung den höheren Einsatz bedingt oder nicht eher umgekehrt die wachsende Anwendung die Notwendigkeit einer legalen Verankerung befördert (hierzu ausführlicher Stockmann & Meyer, 2020).

Eine häufige Anwendung bedeutet jedoch nicht unbedingt, dass die Evaluationsergebnisse auch für die Weiterentwicklung von Programmen und Projekten sowie im weiteren Sinne für die Politikgestaltung genutzt werden (zu den Besonderheiten der Nutzung von Evaluationen im politischen System siehe Bjørnholt & Larsen, 2014; zu den Einflussfaktoren Vanlandingham, 2011). Letztlich sind der Gebrauch und die Nutzung von Evaluation zur Politikgestaltung ein politischer Prozess, dem in den einzelnen Ländern verschiedene Barrieren und Widerstände entgegenstehen. Häufig fehlt es an systematischen, fest institutionalisierten Wegen zur Verbreitung und Weiterverarbeitung der Evaluationsergebnisse (z. B. in Belgien und Frankreich). An manchen Stellen gibt es eine generelle Abwehrhaltung gegen Evaluationen, die zumeist auf Vorurteilen oder Ängsten basieren (z. B. in Italien). Insbesondere in den Ländern mit einem weitentwickelten formalen Rahmen (Niederlande und Schweiz) wird die zunehmende Routinisierung und bürokratische „Abarbeitung“ der „Evaluationspflicht“ als ein Problem gesehen. Hinzu kommt in manchen Ländern (z. B. Dänemark und Finnland) die fortschreitende Vermengung mit anderen Konzepten und Managementpraktiken, welche die zentrale Bedeutung der Evaluation zur Faktengenerierung in Frage stellt und die distanzierte, neutrale und faktenorientierte Perspektive aufgibt. Dies führt dazu, dass sich die Evaluierungspraxis zunehmend von ihrer wissenschaftlichen Basis entfernt und von den aktuellen Trends in der universitären Forschung entkoppelt wird. Umgekehrt verliert der akademische Bereich teilweise das Interesse an Evaluation und die bereits erreichte, aber vielfach noch sehr schwach verankerte Professionalisierung in Forschung und Lehre wird in Frage gestellt (hierzu ausführlicher Meyer & Stockmann, 2020).

Zusammenfassend lässt sich somit sagen, dass es zur systematischen Nutzung von Evaluationen und ihrer Ergebnisse im politischen System formalisierte und fest institutionalisierte Bearbeitungswege geben muss, die dann aber zu einer bürokratischen Verarbeitung im Sinne eines Pflichtprogramms führen und der Evaluationen impliziten Intention einer kreativen, lernorientierten und diskursiven Weiterentwicklung der Programme und Projekte entgegenstehen. Offensichtlich ist es noch nirgendwo gelungen,

eine systematische Be- und Verarbeitung von Evaluation innerhalb der bestehenden Politikpraxis zu institutionalisieren, die den kontinuierlichen Einsatz und gleichzeitig die angemessene Nutzung zufriedenstellend gewährleisten kann. Dies bedeutet jedoch nicht unbedingt, dass Evaluationen prinzipiell nicht oder nur ungenügend genutzt werden (siehe dazu die Fallbeispiele von Petra Kaps und Regine Paul in diesem Heft).

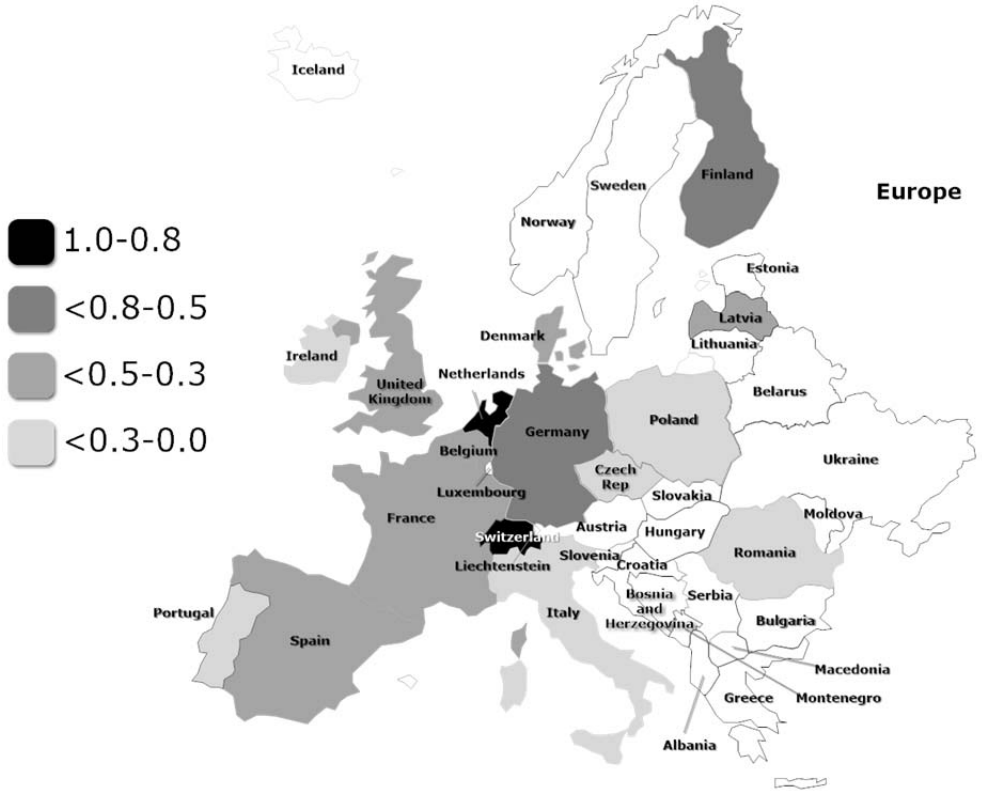
Die Institutionalisierung von Evaluation im politischen System stellt sich also aus mehreren Gründen als Problem dar. Erstens müssen bei der Nutzung von Evaluation als Steuerungsinstrument der Exekutive Ängste, Vorurteile und Widerstände der Akteure überwunden werden, die staatliche Kontrollen und den Verlust eigenständiger Handlungsspielräume fürchten. Zweitens entstehen u. a. hierdurch bedingt innerhalb der verschiedenen Politikfelder und den jeweils zuständigen Ressorts voneinander abweichende Praktiken des Einsatzes von Evaluation, die in spezifische Regelungen münden – häufig zunächst in Form unverbindlicher Empfehlungen und dem Teilen von Erfahrungswerten. Drittens führt diese Regulierung auf sektoraler Ebene, die zumeist durch transnationale Organisationen (insbesondere die EU) zur Homogenisierung der europäischen Politik und zur Absicherung eigener Investitionen gefördert wird, zu einer sehr heterogenen Evaluationspraxis auf nationaler Ebene. Bis auf wenige Ausnahmen ist es deshalb den europäischen Nationalstaaten bisher nicht gelungen, eine einheitliche nationale legale Grundlage für Evaluationen zu entwickeln und diese in ihre jeweilige politische Ordnung als festen Bestandteil der Politikgestaltung aller Ressorts über die Legislative einzubinden.

Die Institutionalisierung der Evaluation im politischen System folgt aus diesem Grund in Europa keineswegs einem gleichmäßigen Verlaufsmuster. Entsprechend der aktuellen Bestandsaufnahme des GLOBE lassen sich grob vier verschiedene Ausmaße der Institutionalisierung identifizieren (*Abbildung 3*), die jedoch nicht als unterschiedliche Stadien eines einheitlichen Entwicklungsprozesses zu bezeichnen sind.

An der Spitze befinden sich die Niederlande und die Schweiz, denen es als einzige Länder (zudem vergleichsweise früh) gelungen ist, auf nationaler Ebene eine legale Grundlage zu schaffen und Evaluation als Standardinstrument zu implementieren. In beiden Staaten werden dadurch Evaluationen regelmäßig eingesetzt und überdurchschnittlich häufig als Entscheidungsgrundlage genutzt. Das hohe Ausmaß an Institutionalisierung bedeutet jedoch nicht, dass alle Widerstände überwunden werden konnten und eine einheitliche, allgemein akzeptierte nationale „Evaluationenkultur“ entstanden ist.

Mit Deutschland und Finnland folgen zwei weitere Staaten, die sich etwas vom Durchschnitt der europäischen Länder absetzen konnten. Hier gibt es zwar keinen allgemeingültigen legalen Rahmen und die Evaluationspraxis konzentriert sich stärker auf bestimmte Politikfelder, im Vergleich zur dritten Gruppe sind aber Evaluationen in deutlich mehr Bereichen gut verankert und werden regelmäßig eingesetzt. Die dritte Gruppe umfasst mit Belgien, Dänemark, Frankreich, Großbritannien und Spanien die meisten westlichen Staaten zu denen bemerkenswerterweise Lettland als einziges der mitteleuropäischen Länder aufgeschlossen hat. Kaum Institutionen und nur wenig Evaluationseinsatz gibt es in Irland, Italien, Polen, Portugal, Rumänien und Tschechien. Es ist natürlich eine interessante Frage, wie sich diese Unterschiede erklären lassen und welche Ursachen hierfür verantwortlich zu machen sind.

Abbildung 3: Stand der Institutionalisierung von Evaluation in Europa 2019



Quelle: Eigene Berechnung; Skala von 0 „Evaluation ist nicht institutionalisiert“ bis 1 „Evaluation ist voll und ganz institutionalisiert“ (Mittelwert von 7 Indikatoren, jeweils 0-1 codiert).

## 4 Einflussfaktoren und ihre Bedeutung

Die beschriebenen Unterschiede und Gemeinsamkeiten hinsichtlich der Institutionalisierung von Evaluation im politischen System sind eine Momentaufnahme und könnten sowohl durch die spezifischen nationalen Entwicklungen als auch durch abweichende Rahmenbedingungen und daraus resultierende Unterschiede zwischen den Ländern bedingt sein. Auf einige dieser potentiellen Einflussfaktoren und ihre Bedeutung zur Erklärung der Varianz zwischen den europäischen Ländern soll im Folgenden näher eingegangen werden.

Als erstes lassen sich die Länder Europas hinsichtlich der *Entwicklung einer nationalen Evaluationstradition* differenzieren. Einige westeuropäische Länder (speziell Deutschland, Frankreich, Großbritannien, die Niederlande, Norwegen, Schweden und die Schweiz) können bereits auf eine lange Evaluationsgeschichte zurückblicken, die bis in die 1960er Jahre zurückreicht. Erste Institutionalisierungsansätze gab es dort bereits in den 1980er Jahren (als Überblick siehe Derlien, 1990; für Deutschland Hellstern & Wollmann, 1984). In anderen Ländern konnte sich Evaluation dagegen erst relativ spät als Steuerungsinstrument etablieren. Dies gilt im Westen für Spanien und



Portugal, deren Demokratisierungsprozesse nach Ende der Diktaturen in den 1980er Jahren abgeschlossen waren und die neuen politischen Strukturen sich vor allem nach dem Beitritt zur Europäischen Union 1986 stabilisieren konnten (siehe dazu Barrios, 2009; Fonseca, 2009). Der Aufbau und die Entwicklung der Evaluation ist hier eng mit dem Beitritt zur Europäischen Union verbunden.

Ähnliches gilt für die mittelosteuropäischen Länder (Lettland, Polen, Rumänien und Tschechien), deren Demokratisierungsprozesse mit der Erosion des Ostblocks in den 1980er Jahren begannen und erst Ende der 1990er Jahre beendet werden konnten. Auch hier prägte der Beitritt zur Europäischen Union 2004 die Institutionalisierung der Evaluation wesentlich mit. Dies begründet sich vor allem durch die Kohäsionspolitik der Europäischen Kommission und die Verknüpfung von Regional- und Infrastrukturförderung mit dem Aufbau eines Evaluationssystems zur Wirkungskontrolle dieser Maßnahmen (vgl. zur „Good Governance“-Politik der EU in Mitteleuropa Börzel, Pamuk & Stahn, 2008; Börzel, 2009, zur Verankerung der Evaluation in der EU Bloemeyer & Eußner, 2020).

Diese *historisch bedingten Unterschiede* prägen vor allem die Art und Weise, wie Evaluationen in den Ländern institutionalisiert worden sind, sowie die Politikfelder, in denen Evaluation betrieben wird. In den Mitteleuropäischen Ländern führte der hohe Bedarf an europäischen Beihilfen zur Strukturförderung zu einer Implementierung von Evaluationsstellen auf nationaler Ebene, die es in dieser Form in den westlichen Pionierstaaten nicht oder nur ansatzweise gegeben hat. Die Evaluationspraxis ist hier deutlich stärker von den europäischen Fördermitteln und damit von deren Richtlinien zur Mittelvergabe abhängig, die eine Implementierung von Evaluationen vorsehen. In anderen Politikfeldern kann sie sich nur allmählich etablieren.

Neben der Europäischen Union gibt es noch weitere *internationale Organisationen und Netzwerke*, welche die nationale Entwicklung der Evaluationspraxis stark geprägt haben. Dies gilt vor allem für die OECD, die insbesondere im Bildungsbereich und bei der Entwicklungszusammenarbeit Einfluss erlangt hat. In der Entwicklungszusammenarbeit z. B. kam es bereits mit der Gründung der OECD 1961 zur Institutionalisierung eines Development Assistance Committees (DAC) zur Koordination der Geberländer, in dessen Rahmen ab den 1980er Jahren auch eine permanente Arbeitsgruppe zur Entwicklung der Evaluation entwicklungspolitischer Programme und Projekte gebildet wurde (DAC Network on Development Evaluation, <http://www.oecd.org/dac/evaluation/>). Dies hat zu einer kontinuierlichen Weiterentwicklung und Institutionalisierung der Evaluation auf nationaler Ebene geführt (u. a. bedingt durch die regelmäßigen *Peer Reviews* der nationalen Evaluationspraxis, siehe u. a. OECD, 2016). Ähnliche Reviews durch die OECD gibt es auch in anderen Politikbereichen (z. B. Nusche, Braun, Halász & Santiago, 2014) und darüber hinaus wurden eine Vielzahl von Handlungsempfehlungen, *Guidelines* und Praxisberichte zu Anwendung und Einsatz von Evaluationen erstellt (als Beispiel hier die Richtlinien für Evaluationen in Krisengebieten OECD, 2012).

Im Unterschied zur EU ist der Verpflichtungsgrad der OECD und anderer internationaler Organisationen geringer, d. h. die Teilnahme und die Einhaltung von Regeln und Vorgaben ist freiwillig. Der Einfluss ist aber in bestimmten Politikfeldern trotzdem hoch und führt zumindest punktuell und mit zeitlichem Abstand zu Veränderungen auf nationaler Ebene. Die Entwicklung der Evaluation wurde und wird durch diese internationale Zusammenarbeit wesentlich geprägt, wobei hierdurch die sektorale Differenzierung auf nationaler Ebene angesichts einer fehlenden allgemeinen Evaluationspolicy

internationaler Organisationen befördert wird. Hinzu kommt noch der Fakt, dass nicht alle Nationalstaaten in Europa in allen Politikfeldern auf internationaler Ebene in gleicher Weise engagiert und in den supranationalen Austausch eingebunden sind. Insgesamt führt der internationale Einfluss deshalb nicht zu einer generellen Stärkung der Evaluation auf nationaler Ebene und beeinflusst die Institutionalisierung im nationalen politischen System lediglich indirekt, ein direkter Einfluss (weder positiv noch negativ) lässt sich mit den GLOBE-Daten nicht belegen. Dies betrifft sowohl die legale Verankerung als auch die Anwendung und Nutzung in der Exekutive.

Ein wichtiger Unterschied zwischen den Ländern Europas ist deren Größe sowohl hinsichtlich der Fläche als auch hinsichtlich der Einwohnerzahl. Besonders die *Bevölkerungsgröße* ist für die Herausbildung eines eigenständigen nationalen Evaluationsmarkts und die Organisation eines Austauschs zwischen den in Evaluationen eingebundenen Personen im Land wichtig. Dies hat unmittelbar Auswirkungen auf die Professionalisierung und die Etablierung professioneller Aus- und Weiterbildungsstrukturen – hier tun sich kleinere Länder deutlich schwerer als die größeren.

Aus Sicht der Anbieter von Evaluationsdienstleistungen bedarf es einer kritischen Masse von Aufträgen, damit sich eine Beratungsfirma auf diese Art von Dienstleistungen spezialisieren kann. Wenn diese im nationalen Kontext nicht vorhanden ist, muss das Unternehmen sich entweder auf dem internationalen Markt etablieren (was angesichts des offenen europäischen Marktes und den Regulierungen zum Schutz des Marktzugangs europäischer Firmen durch die EU vergleichsweise einfach möglich ist) oder ein breiteres Angebot an Dienstleistungen neben der Evaluation auf dem nationalen Markt bereitstellen. Allerdings besteht dabei die Gefahr, dass Anbieter aus anderen (größeren) Staaten als Konkurrenten aufgrund ihrer umfangreicheren Expertise in Evaluation bessere Marktchancen haben.

Die Größe des Nationalstaats hat nicht nur Auswirkungen auf die Angebotsseite, sondern auch auf die Nachfrage, weil bedingt durch das Steueraufkommen und die *Wirtschaftsleistung* in kleinen Ländern in der Regel weniger Programme und Projekte durchgeführt werden bzw. durchgeführt werden können. Bei gleichen Rahmenbedingungen wird es dementsprechend in kleinen Staaten weniger Evaluationen als in größeren geben. Allerdings existieren in Europa kleine Länder mit extrem hoher Wirtschaftskraft (z. B. Luxemburg, die Schweiz oder Norwegen), während einige größere Länder (z. B. Polen, Rumänien oder die Ukraine) über ein deutlich niedrigeres Wohlstandsniveau verfügen.

Bei den 16 untersuchten Ländern schwankte 2019 die Bevölkerungsgröße zwischen 83 Mio. (Deutschland) und knapp 2 Mio. (Lettland), das Bruttoinlandsprodukt zwischen 3,4 Billionen EUR (Deutschland) und etwas über 30 Mrd. EUR (Lettland) (EUROSTAT Online Data TPS00001 und TEC00001). Wie zu erwarten haben beide Größen einen Einfluss auf den Professionalisierungsgrad der Evaluation auf nationaler Ebene, d. h. die Etablierung eigenständiger Studiengänge und Evaluationsgesellschaften, von Fachzeitschriften oder anderen Kommunikationsmedien sowie von Qualitätssicherungsmechanismen wie z. B. Standards, gelingt mit zunehmender Bevölkerungsgröße und Wirtschaftskraft besser. Interessanterweise gilt das jedoch nicht für die Institutionalisierung im politischen System: Hier ergibt sich kein nennenswerter Einfluss von Bevölkerungsgröße und Wirtschaftskraft.<sup>2</sup>

Die europäischen Nationalstaaten unterscheiden sich aber nicht nur durch ihre Größe und Wirtschaftskraft, sondern auch bezüglich der historisch gewachsenen und in

der Vergangenheit vielfachen Verwerfungen ausgesetzten *politischen Systeme*. Hinsichtlich der Herrschaftsform dominieren mittlerweile in Europa parlamentarische Regierungssysteme (Deutschland, Irland, Italien, Finnland, Lettland und Tschechien), von denen allerdings einige als parlamentarische Monarchien (Belgien, Dänemark, Großbritannien, Niederlande und Spanien) oder als semi-präsidentielle Republiken (Frankreich, Polen, Portugal und Rumänien) verfasst sind. Eine Besonderheit stellt die Schweiz dar, die sich von den anderen Republiken durch eine deutlich stärkere Konsensorientierung in Richtung Konkordanzdemokratie mit ausgeprägten basisdemokratischen Elementen (z. B. Volksinitiativen und Referenden) unterscheidet. Im Vergleich zu den stärker zentralistisch organisierten Monarchien und semi-präsidentiellen Republiken ist die Schweiz genauso wie Belgien und Deutschland als Föderalstaat organisiert, d. h. die Teilregionen (in der Schweiz Kantone) verfügen über eigenständige Entscheidungskompetenzen in einer Reihe politischer Fragen.

Weitere wichtige Unterschiede ergeben sich durch das Wahlsystem und die daraus resultierende Verteilung der Parlamentssitze, die Organisation der Gewaltenteilung (insbesondere die Abgrenzung zwischen Legislative und Exekutive) sowie die Rolle und Funktion des Staatsoberhauptes (insbesondere in Relation zum Parlament und zum Regierungschef). All dies hat Auswirkungen auf die formalen Grundlagen, Politikstil und informellen Gepflogenheiten innerhalb der Exekutive, speziell des Regierungshandelns und der administrativen Praxis in der öffentlichen Verwaltung (allgemeinen zu den Unterschieden in der Qualität der Regierungsführung in Europa Berg-Schlosser, 2006; zu den Einflussfaktoren und Veränderungen politischer Steuerung des Staats Mayntz, 2004).

Insgesamt lässt sich allerdings feststellen, dass die Art des politischen Systems und seine prägenden Elemente erstaunlich wenig direkten Einfluss auf die Institutionalisierung von Evaluation haben. Es gibt z. B. keine Hinweise, dass die Institutionalisierung in zentralstaatlichen einfacher und schneller als in föderalen Systemen erfolgt. Im Gegenteil war es gerade die Schweiz, deren Konkordanzdemokratie aufwendige konsensorientierte Entscheidungsprozesse erforderlich macht, welche als bisher einziger europäischer Staat frühzeitig und umfassend Evaluation im politischen System institutionalisieren und kontinuierlich weiterentwickeln konnte (hierzu die ausführlichen Analysen von Sager, Widmer & Balthasar, 2017, siehe auch den Beitrag von Thomas Widmer in diesem Heft). Die anderen Bundesstaaten (Belgien und Deutschland) belegen allerdings, dass nicht die föderale Ordnung per se die Institutionalisierung befördert – einige zentralstaatliche Systeme (Finnland und die Niederlande) haben deutlich stärker Evaluation als politisches Instrument integriert.

Die Befunde des GLOBEs legen also nahe, dass die Institutionalisierung von Evaluation im politischen System weniger durch die strukturellen Rahmenbedingungen und Besonderheiten als durch Entwicklungsprozesse und Reformen in der Regierungspraxis geprägt wird. Besonders in den Pionierländern (Deutschland, Frankreich, Großbritannien und Schweden) sind dabei die *Konzepte des New Public Management und der evidence-based policy* wegweisend gewesen und haben zu einer breiten Nutzung von Evaluationen geführt (allgemein zur Modernisierung des Staates Naschold & Bogumil, 2000, zur Rolle der Evaluation Wollmann, 2003, kritisch Pawson, 2006). In der Tat ist die Korrelation zwischen der Institutionalisierung von Evaluation und den beiden zentralen World Governance Indikatoren zur regulatorischen Qualität und zur Effektivität des Regierens sehr hoch (Stockmann, Meyer & Taube, 2020b, p. 515; zu

den World Governance Indikatoren der Weltbank siehe Kaufmann, Kraay & Mastruzzi, 2010). Es besteht also offensichtlich ein Zusammenhang zwischen guter Regierungsführung und Implementierung von Evaluation, wobei vermutlich die Evaluationspraxis mehr das Ergebnis guter Regierungsführung ist als umgekehrt. Die Institutionalisierung in Form von festen, verbindlichen und allgemeingültigen Regeln ist dagegen offenbar eher die Folge denn die Voraussetzung für eine Nutzung des Instruments der Evaluation im politischen System.

## 5 Schlussfolgerungen und Ausblick

Insgesamt lassen sich gegenwärtig fünf Trends erkennen, die in den Ländern in unterschiedlichem Maße wirksam und für die Institutionalisierung von Evaluation bedeutsam sind.

Erstens prägen in den meisten Ländern immer noch eine *fortschreitende Expansion und Verbreitung* des Instruments der Evaluation die nationale Entwicklung. Dies bedeutet, dass weniger eine Verfestigung und Institutionalisierung denn die Übernahme und Anpassung in neuen Anwendungsgebieten im Vordergrund stehen.

Zweitens gibt es insbesondere in den Politikfeldern, in denen sich die Nutzung von Evaluationen als Standardinstrument etablieren konnte, einen *Trend zur Professionalisierung* im Sinne einer kontinuierlichen Qualitätsentwicklung der Evaluationen als auch einer Diversifikation des Instrumenteneinsatzes. Hier steht die Herausbildung von Institutionen und Regulationen zur Nutzung von Evaluationen im Vordergrund, auch wenn es bisher nur wenige Länder und wenige Handlungsfelder gibt, in denen die Evaluation bereits routinemäßig und klaren Vorgaben folgend eingesetzt wird (dazu für Deutschland die Beiträge in Hense, Böttcher, Kalman & Meyer, 2019; zum weltweiten Professionalisierungstrend siehe Meyer, 2015).

Diesen positiven Entwicklungstrends stehen aber auch negative zur Seite, die in Zukunft mehr Gewicht erlangen könnten. So ist drittens insbesondere in den Pionierländern mit dem fortgeschrittensten Stand der Evaluation eine gewisse Stagnation oder zumindest *nachlassende Dynamik* zu beobachten, die eher mit einer Deprofessionalisierung denn mit fortschreitender Weiterentwicklung einhergeht. Kritische Stimmen beklagen den übertriebenen Einsatz des Instruments in Form einer „Evaluitis“ (Frey, 2007; 2008) und den bescheidenen Nutzwert von Evaluationen (als Überblick zu den Befunden hierzu Johnson, Greenseid, Toal, King, Lawrenz & Volkov, 2009). Diese Tendenz könnte für die Evaluation trotz der in diesen Ländern bestehenden Institutionen gefährlich werden, weil die meisten Institutionen noch vergleichsweise schwach verankert sind (z. B. der gesetzliche Rahmen für die Nutzung fehlt) und deshalb relativ einfach und schnell zurückgenommen werden können (wie z. B. bei der Einrichtung eines nationalen Evaluationsinstituts in Schweden geschehen).

In ähnliche Richtung weist auch der vierte Trend einer *Vermischung und Verfremdung* der Evaluation mit anderen Konzepten wie z. B. dem Auditing oder Controlling, die einer eigenständigen Profilbildung entgegenwirken und ggf. zu einer Auflösung führen können. Ansätze einer solchen Entwicklung lassen sich entsprechend der Befunde des GLOBES in Dänemark, Frankreich und der Schweiz finden.

Schließlich ist noch der generelle Trend weg von faktenbasierter und konsensorientierter Politik hin zu einem seine Wahrheiten selbst produzierenden *Populismus* her-

vorzuheben. Diese allgemeine Entwicklung in der Politik ist nicht allein auf Europa beschränkt, hat aber viele Länder erfasst und zu fundamentalen Umbrüchen im Parteiensystem geführt (Greven, 2016; Bugaric & Kuhelj, 2018; Biorcio, 2019). Bis jetzt sind die Auswirkungen auf die Institutionalisierung von Evaluation noch gering, es ist jedoch zu erwarten, dass ein weiteres Wachstum populistischen Gedankenguts einen ähnlichen Effekt wie die Verbreitung des Konzepts der *evidence-based policy* in den 1980er Jahren haben könnte – nur mit umgekehrtem Vorzeichen, d. h. eine ideologiebasierte Politik wird problemlos auf Fakten produzierende Verfahren wie die Evaluation verzichten können. Einher geht diese Entwicklung zudem mit einer Erosion internationaler Institutionen und einer Renationalisierung, die in Europa bereits zu einer politischen Krise der Europäischen Union geführt hat (Dinan, Nugent & Paterson, 2017). Wie die Befunde belegten, hat die Europäische Union in vielen Ländern eine bedeutende Rolle für die Etablierung und Institutionalisierung gespielt und eine Schwächung der EU könnte sich dementsprechend auch auf nationaler Ebene negativ auswirken.

Insgesamt ist aber nicht zu erwarten, dass die Evaluation als Steuerungsinstrument schnell aus den politischen Systemen Europas verschwinden wird. Vor allem vier Dinge stehen dem entgegen:

- Erstens die ungebrochene Tendenz *transnationaler Organisationen* (insbesondere der EU), Evaluationen von ihren Mitgliedsnationen als Gegenleistung für Förderungsmaßnahmen einzufordern und dafür die Evaluation sowie deren Weiterentwicklung zu fördern.
- Zweitens der mittlerweile in einer Reihe von Sektoren erreichte *Status des Einsatzes von Evaluationen* und dessen Nutzung für die Weiterentwicklung ihrer Instrumente. Dies ist nicht ausschließlich dem Druck der transnationalen Organisationen, sondern auch einer gewissen sektoralen Eigendynamik geschuldet (mit spezifischen nationalen Formungen und Entwicklungstrends).
- Drittens der in manchen Ländern bereits erreichte *Grad der Institutionalisierung* auf nationaler Ebene, nicht nur in Form von Gesetzen und Verordnungen, sondern auch hinsichtlich des regelmäßigen Einsatzes von Evaluation und der geregelten Nutzung von Ergebnissen.
- Und schließlich viertens die *Verfestigung der output- und outcome-orientierten Steuerung* in den politischen Systemen als Folge des *New Public Management*, die kaum zurückgenommen und wieder durch eine reine Inputsteuerung ersetzt werden wird. Als hierfür notwendiges Steuerungsinstrument ist die Evaluation mit dieser Entwicklung verbunden und nicht von ihr trennbar.

Dies bedeutet jedoch auf der anderen Seite nicht, dass Evaluation „fest im Sattel sitzt“ und wie in der Vergangenheit expandieren wird. Dem stehen nicht nur einige allgemeine Entwicklungstrends, sondern auch ein ungelöstes Problem der Institutionalisierung im politischen System entgegen. Die beschriebene „top-down“-Steuerung moderner Gesellschaften mittels bürokratischer Abarbeitung ist nicht, wie von verschiedenen Autoren (z. B. Beck, 1993) erwartet, in eine prinzipielle Krise geraten und ein Wandel von „Government“ zu „Governance“ ist nur partiell eingetreten. Eine kompromissorientierte, partizipative Politikgestaltung konnte sich in Europa nicht verbreiten, in manchen Ländern sind sogar deutlich gegenläufige Tendenzen erkennbar.

Das Beispiel der Schweiz zeigt aber auch, dass sich Evaluation in einem solchen Umfeld deutlich besser entfalten kann als in einer bürokratischen regelgesteuerten und

allein über politische Institution gelenkten Gesellschaft. Dies liegt an einem bisher nicht aufgelösten *Grundsatzkonflikt* zwischen dem Selbstverständnis der Evaluation als Instrument der Aufklärung mit der Aufgabe, zusammenzuführen und die Zivilgesellschaft zu „empowern“, und dem Wunsch der Exekutive, mittels Evaluation das von ihr initiierte gesellschaftliche Handeln hinsichtlich seiner Wirkungen zu überwachen und durch zielorientierte Maßnahmen zu lenken. Diese Ambivalenz prägt gegenwärtig die Institutionalisierung der Evaluation in Europa.

## Anmerkungen

- 1 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Selbstverständlich sind alle Geschlechter gleichermaßen gemeint.
- 2 Getestet wurden neben den absoluten Zahlen auch die Bevölkerungsdichte, das Pro-Kopf-Einkommen, der Human-Development-Index, die Staatsinvestitionen und eine Reihe anderer gängiger demographischer und ökonomischer Kennzahlen. Keiner dieser Indikatoren kann zur Erklärung der beschriebenen Unterschiede hinsichtlich des Institutionalisierungsgrads im politischen System beitragen.

## Literatur

- Ansell, Christopher & Torfing, Jacob (Eds.) (2016). *Handbook on Theories of Governance*. Cheltenham: Edward Elgar.
- Barley, Stephen R. & Tolbert, Pamela S. (1997). Institutionalization and Structuration: Studying the Links between Action and Institution. *Organizational Studies*, 18 (1), 93-117. <https://doi.org/10.1177%2F017084069701800106>.
- Barrios, Harold (2009). Das politische System Spaniens. In Wolfgang Ismayr (Hrsg.), *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage* (S. 713-764). Wiesbaden: VS Verlag.
- Beck, Ulrich (1993). *Die Erfindung des Politischen. Zu einer Theorie reflexiver Modernisierung*. Frankfurt: Suhrkamp.
- Berg-Schlosser, Dirk (2006). The quality of democracies in Europe as measured by current indicators of democratization and good governance. *Journal of Communist Studies and Transition Politics*, 20 (1), 28-55. <https://doi.org/10.1080/13523270410001687091>.
- Biorcio, Roberto (2019). Populism as a challenge for party systems: A comparison between Italy and Spain. In Gregor Fitz, Jürgen Mackert & Brian S. Turner (Eds.), *Populism and the Crisis of Democracy. Vol. 2: Politics, Social Movements and Extremism* (pp. 137-151). London/New York: Routledge.
- Bjørnholt, Bente & Larsen, Flemming (2014). The politics of performance measurement: 'Evaluation use as mediator for politics'. *Evaluation*, 20 (4), 400-411. <https://doi.org/10.1177/1356389014551485>.
- Blomeyer, Roland & Eußner, Ansgar (2020). European Union. In Reinhard Stockmann, Wolfgang Meyer & Lena Taube (Eds.), *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I.* (pp. 463-479). Cham: Springer Nature.
- Börzel, Tanja A., Pamuk, Yasemin & Stahn, Andreas (2008). *Good Governance in the European Union*. Berlin: FU Berlin Berliner Arbeitspapier zur Europäischen Integration 7. Verfügbar unter [https://refubium.fu-berlin.de/bitstream/handle/fub188/19026/2008-7\\_Boerzel\\_Pamuk\\_Stahn.pdf?sequence=1](https://refubium.fu-berlin.de/bitstream/handle/fub188/19026/2008-7_Boerzel_Pamuk_Stahn.pdf?sequence=1) [06.11.19].
- Börzel, Tanja A. (2009). *Transformative Power Europe? The EU Promotion of Good Governance in Areas of Limited Statehood*. Paper prepared for the ERD Workshop "Transforming Political Structures: Security, Institutions, and Regional Integration Mechanisms", Florence, April 16-17, 2009. Verfügbar unter

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.551.2884&rep=rep1&type=pdf> [06.11.2019].
- Bugaric, Bojan & Kuhelj, Alenka (2018). Varieties of Populism in Europe: Is the Rule of Law in Danger? *Hague Journal on the Rule of Law*, 10 (1), 21-33.  
<https://doi.org/10.1007/s40803-018-0075-4>.
- Cousins, J. Bradley, Goh, Swee C., Elliott, Catherine, Aubry, Tim & Gilbert, Nathalie (2014). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, 44, 1-13. <https://doi.org/10.1016/j.evalprogplan.2013.12.001>.
- Derlien, Hans-Ulrich (1990). Genesis and Structure of Evaluation Efforts in Comparative Perspective. In Ray C. Rist (Ed.), *Program Evaluation and the Management of Government. Patterns and Prospects across Eight Nations* (pp. 147-176). New Brunswick/London: Transaction Publishing.
- Dinan, Desmond, Nugent, Neill & Paterson, William E. (Eds.) (2017). *The European Union in Crisis*. Palgrave MacMillan: London.
- EUROSTAT (o. J.). *Europäische Statistik Online Datenbank*. Verfügbar unter: <https://ec.europa.eu/eurostat/de/data/database> [22.04.2020].
- Fonseca, Sara Claro da (2009). Das politische System Portugals. In Wolfgang Ismayr (Hrsg.), *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage* (S. 765-818). Wiesbaden: VS Verlag.
- Frey, Bruno S. (2007). Evaluierungen, Evaluierungen ... Evaluitis. *Perspektiven der Wirtschaftspolitik*, 8 (3), 207-220. <https://doi.org/10.1111/j.1468-2516.2007.00248>.
- Frey, Bruno S. (2008). Evaluitis — eine neue Krankheit. In Hildegard Matthies & Dagmar Simon (Hrsg.), *Wissenschaft unter Beobachtung* (S. 125-140). VS Verlag.
- Furubo, Jan-Eric, Rist, Ray & Sandahl, Rolf (2002). *International Atlas of Evaluation*. New Brunswick, NJ: Transatlantic Publishers.
- Goerres, Achim, Siewert, Markus B. & Wagemann, Claudius (2019). Internationally Comparative Research Designs in the Social Sciences: Fundamental Issues, Case Selection Logics, and Research Limitation. In Hans-Jürgen Andreß, Detlef Fetchenhauer & Heiner Meulemann (Eds.), *Cross-National Comparative Research* (pp. 75-97). Wiesbaden: VS Verlag, Sonderheft 59 der Kölner Zeitschrift für Soziologie und Sozialpsychologie.
- Grevén, Thomas (2016). *The Rise of Right-wing Populism in Europe and the United States. A Comparative Perspective*. Berlin: Friedrich-Ebert-Stiftung.
- Hay, Colin, Lister, Michael & Marsh, David (Eds.) (2006). *The State. Theories and Issues*. London: Red Globe Press, MacMillan International.
- Hellstern, Gerd-Michael & Wollmann, Hellmut (1984). Evaluierung und Evaluierungsforschung – ein Entwicklungsbericht. In dies. (Hrsg.). *Handbuch zur Evaluierungsforschung. Bd. 1* (S. 17-90), Opladen: Westdeutscher Verlag.
- Hense, Jan, Böttcher, Wolfgang, Kalman, Michael & Meyer, Wolfgang (Hrsg.) (2019). *Evaluation: Standards in unterschiedlichen Handlungsfeldern, Einheitliche Qualitätsansprüche trotz heterogener Praxis?* Münster: Waxmann.
- Hroch, Miroslav (2005). *Das Europa der Nationen. Die moderne Nationsbildung im europäischen Vergleich*. Göttingen: Vandenhoeck & Ruprecht.
- Ismayr, Wolfgang (Hrsg.) (2009). *Die politischen Systeme Westeuropas. 4. aktualisierte und überarbeitete Auflage*. Wiesbaden: VS Verlag.
- Ismayr, Wolfgang (Hrsg.) (2010). *Die politischen Systeme Osteuropas. 3. aktualisierte und erweiterte Auflage*. Wiesbaden: VS Verlag.
- Jacob, Steve, Speer, Sandra & Furubo, Jan-Eric (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21 (1), 6-31.  
<https://doi.org/10.1177/1356389014564248>.
- Johnson, Kelli, Greenseid, Lija O., Toal, Stacie A., King, Jean A., Lawrenz, Frances & Volkov, Boris (2009). Research on Evaluation Use. A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30 (3), 377-410. <https://doi.org/10.1177/1098214009341660>.
- Kaps, Petra (2020). Formen der Evaluation experimenteller Politikgestaltung. *dms – der moderne staat*, 13 (1), 81-100.

- Kaufmann, Daniel, Kraay, Aart & Mastruzzi, Massimo (2010). *The Worldwide Governance Indicators Methodology and Analytical Issues*. Washington: World Bank Policy Research Working Paper 5430.
- Lawrence, Thomas B., Winn, Monika I. & Jennings, P. Deveraux (2001). The Temporal Dynamics of Institutionalization. *Academy of Management Review*, 26 (4), 624-644.  
<https://doi.org/10.5465/amr.2001.5393901>.
- Mayntz, Renate (2004). Governance im modernen Staat. In Arthur Benz (Hrsg.). *Governance – Regieren in komplexen Regelsystemen* (S. 65-76). Wiesbaden: VS Verlag.
- Meyer, Wolfgang (2015). Professionalisierung von Evaluation: Ein globaler Blick. In: *Zeitschrift für Evaluation*, 14 (2), 215-246.
- Meyer, Wolfgang (2019a). Professionalisierung zwischen Kohärenz, Koexistenz und Konkurrenz. Das Fallbeispiel der Evaluation. In Nicole Burzan (Hrsg.). *Komplexe Dynamiken globaler und lokaler Entwicklungen. Verhandlungen des 39. Kongresses der Deutschen Gesellschaft für Soziologie in Göttingen 2018*. Verfügbar unter  
[http://publikationen.sozioologie.de/index.php/kongressband\\_2018/issue/view/33\\_\[08.11.2019\]](http://publikationen.sozioologie.de/index.php/kongressband_2018/issue/view/33_[08.11.2019]).
- Meyer, Wolfgang (2019b). Die DeGEval-Standards im weltweiten Vergleich. In Jan Hense, Wolfgang Böttcher, Michael Kalman & Wolfgang Meyer (Hrsg.), *Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?* (S. 277-294) Münster: Waxmann.
- Meyer, Wolfgang & Stockmann, Reinhard (2020). Evaluation in Europe 2019. A comparative analysis. (Manuskript eingereicht zur Publikation in *Evaluation*, im Begutachtungsverfahren).
- Naschold, Frieder & Bogumil, Jörg (2000). *Modernisierung des Staates. New Public Management in deutscher und internationaler Perspektive* (2. Aufl.). Wiesbaden: VS Verlag.
- Nusche, Deborah, Braun, Henry, Halász, Gábor & Santiago, Paulo (2014). *OECD Reviews of Evaluation and Assessment in Evaluation: Netherlands 2014*. Paris: OECD.
- Pattyn, Valérie, Voorst, Stijn van, Mastenbroek, Ellen & Dunlop, Claire A. (2018). Policy Evaluation in Europe. In Edoardo Ongaro & Sandra van Thiel (Eds.), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 577-594), London: Palgrave MacMillan.
- Paul, Regine (2020). Analyse and rule? A conceptual framework for explaining the variable appeals of ex-ante evaluation in policymaking. *dms – der moderne staat*, 13 (1), 124-142.
- Pawson, Ray (2006). *Evidence-based Policy. A realist perspective*. London et al: Sage.
- Peters, B. Guy (2019). *Institutional Theory in Political Science. The New Institutionalism* (4<sup>th</sup> edition). Cheltenham, UK / Northampton, MA, USA: Edward Elgar.
- OECD (2016). *Evaluation Systems in Development Co-operation. 2016 Review*. Paris: OECD.
- OECD (2012). *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results*. Paris: OECD.
- Rosenstein, Barbara (2013). *Status of National Evaluation Policies. Global Mapping Report*. Washington: EvalPartners.
- Sager, Fritz, Widmer, Thomas & Balthasar, Andreas (2017). *Evaluation im politischen System der Schweiz. Entwicklung, Bedeutung und Wechselwirkungen*. Basel: NZZ Libro.
- Stockmann, Reinhard (2008). *Evaluation and Quality Development: Principles of Impact-based Quality Management*. Frankfurt am Main: Peter Lang.
- Stockmann, Reinhard & Meyer, Wolfgang (Hrsg.) (2017a). *Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven*. Münster: Waxmann.
- Stockmann, Reinhard & Meyer, Wolfgang (2017b). Evaluation in Deutschland. Woher sie kommt, wo sie steht, wohin sie geht. *Zeitschrift für Evaluation*, 16 (2), 57-110.
- Stockmann, Reinhard & Meyer, Wolfgang (2020). Die Institutionalisierung der Evaluation in Europa. *Zeitschrift für Evaluation* 1/2020 (im Erscheinen).
- Stockmann, Reinhard & Meyer, Wolfgang (eds.) (2021). *The Institutionalization of Evaluation in the Americas: Evaluation Globe. Vol. II*. Cham: Springer Nature (in Vorbereitung).
- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (Eds.) (2020a). *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I*. Cham: Springer Nature.



- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (2020b). The Institutionalization of Evaluation in Europe: A Synthesis. In: dies. (Eds.), *The Institutionalization of Evaluation in Europe: Evaluation Globe. Vol. I.* (pp. 483-522). Cham: Springer Nature.
- Vanlandingham, Gary R. (2011). Escaping the Dusty Shelf: Legislative Evaluation Offices' Efforts to Promote Utilization. *American Journal of Evaluation*, 32 (1), 85-97.
- Vedung, Evert (2017). *Public Policy and Program Evaluation*. Abingdon/New York: Routledge.
- Widmer, Thomas (2020). Wechselwirkungen von Politik und Evaluation: Befunde aus der Schweiz. *dms – der moderne staat*, 13 (1), 44-60.
- Wollmann, Hellmut (Ed.) (2003). *Evaluation in Public-Sector Reform. Concepts and Practice in International Perspective*. Cheltenham/Northampton: Edward Elgar.
- Yu, Kyoung-Hee (2013). Institutionalization in the Context of Institutional Pluralism: Politics as a Generative Process. *Organizational Studies*, 34 (1), 105-131.

### *Anschriften der Autoren:*

Adj. Prof. Dr. Wolfgang Meyer, Centrum für Evaluation, Universität des Saarlandes, Campus C 5.3, 66123 Saarbrücken, E-Mail: w.meyer@mx.uni-saarland.de.

Prof. Dr. Reinhard Stockmann, Centrum für Evaluation, Universität des Saarlandes, Campus C 5.3, 66123 Saarbrücken, E-Mail: r.stockmann@ceval.de.

Thomas Widmer

## Wechselwirkungen von Politik und Evaluation: Befunde aus der Schweiz

### Zusammenfassung

Der Beitrag stellt ein Forschungsprogramm vor, das sich mit den Wechselwirkungen zwischen der Evaluation und der Politik in der Schweiz befasst. Zuerst weist der Artikel auf ausgewählte Merkmale des schweizerischen politischen Systems hin, die für die weitere Diskussion besonders relevant sind. Danach geht der Beitrag in knapper Form auf die Entwicklung der Evaluation in der Schweiz ein. Im Anschluss wird die übergreifende Fragestellung und der analytische Rahmen des Sinergia-Forschungsprogramms «Policy Evaluation in the Swiss Political System – Roots and Fruits (SynEval)» des Schweizerischen Nationalfonds vorgestellt. Die vier, in diesem Rahmen realisierten Forschungsprojekte zu Evaluations- und Verwaltungskulturen, zu Evaluationsklauseln, zur Bedeutung von Evaluationsbefunden in direktdemokratischen Kampagnen sowie zur Rolle der Evaluation in den schweizerischen Parlamenten werden in knapper Form mit ausgewählten Befunden zum Verhältnis von Evaluation und Politik präsentiert. Der Beitrag schliesst mit einigen weiterführenden Überlegungen zur Forschung über Evaluation und zur Entwicklung der Evaluation in der Schweiz.

**Schlagworte:** Evaluationskultur, Evaluationsklauseln, direkte Demokratie, Parlamente, Evaluation, Schweizer Politik

### Abstract

*Interactions between policy and evaluation: findings from Switzerland*

The contribution presents a research programme that deals with the interaction between evaluation and politics in Switzerland. First, the article highlights selected features of the Swiss political system that are particularly relevant for further discussion. Then the article briefly describes the development of evaluation in Switzerland. Subsequently, the overarching question and the analytical framework of the Sinergia research program "Policy Evaluation in the Swiss Political System - Roots and Fruits (SynEval)" of the Swiss National Science Foundation are presented. The four research projects carried out within this framework 1) on evaluation and administrative cultures, 2) on evaluation clauses, 3) on the importance of evaluation findings in direct-democratic campaigns and 4) on the role of evaluation in Swiss parliaments are summarised concisely with selected findings on the relationship between evaluation and politics. The article concludes with some further reflections on research on evaluation and the development of evaluation in Switzerland.

**Keywords:** Evaluation Culture, Evaluation Clauses, Direct Democracy, Parliaments, Evaluation, Swiss Politics

## 1 Einleitung<sup>1</sup>

Wie gestaltet sich das wechselseitige Verhältnis von Politik auf der einen und Evaluation auf der anderen Seite? Dieser Frage geht der nachfolgende Beitrag anhand des Beispiels der Schweiz und gestützt auf neuere Forschungsbefunde nach. Hierzu sind

einige Präliminarien erforderlich: Zum einen ist die Begrifflichkeit zu klären. Unter *Politik* sollen nachfolgend sowohl strukturelle Rahmenbedingungen der Politik (auch als *polities* bezeichnet), als auch politische Prozesse (auch *politics*) und Politikinhalte (auch *policies*) verstanden werden. Der Begriff *Evaluation* meint hier in Anlehnung an etablierte Definitionen die wissenschaftliche Dienstleistung zur systematischen und transparenten Bewertung eines Gegenstandes, wobei hier Gegenstände der Politik im oben umrissenen Sinne im Zentrum stehen.

Zum anderen ist aber auch die Differenz der beiden Konzepte zu klären, ließe sich doch durchaus überzeugend argumentieren, dass Evaluation (Teil der) Politik und Politik (Teil der) Evaluation ist, wodurch die Debatte zu deren Verhältnis obsolet würde. Stattdessen wird aber im Folgenden von der Prämisse ausgegangen, dass es sich dabei um zwei unterscheidbare Subsysteme handeln soll, die miteinander interagieren und sich gegenseitig prägen können.

Der Beitrag ist wie folgt aufgebaut: Im nachfolgenden Kapitel werde ich zur Kontextualisierung in kurzer Form zentrale Merkmale des schweizerischen politischen Systems skizzieren, die für die nachfolgende Diskussion relevant sind. Im Anschluss folgt im dritten Kapitel eine knappe Darstellung des Entwicklungsstandes der Evaluation in der Schweiz. Das vierte Kapitel umschreibt die Fragestellung und den analytischen Rahmen eines Forschungsprogramms sowie dessen Organisationsstruktur. Es handelt sich dabei um das Sinergia-Forschungsprogramm «Policy Evaluation in the Swiss Political System – Roots and Fruits (SynEval)», das dank der Förderung durch den Schweizerischen Nationalfonds in den Jahren 2013 bis 2016 unter meiner Leitung an fünf Schweizer Universitäten realisiert werden konnte.<sup>2</sup> Kapitel 5 geht sodann auf Inhalte und ausgewählte Befunde des Forschungsprogramms ein. Der Beitrag schliesst mit einem kurzen Fazit.

## 2 Das politische System der Schweiz

Das schweizerische politische System verfügt über eine Reihe von Besonderheiten, die es bei der Diskussion um das Verhältnis von Evaluation und Politik zu beachten gilt. In der gebotenen Kürze möchte ich fünf zentrale Merkmale kurz vorstellen (Widmer, 2017, S. 53-54).

Die *direkte Demokratie* spielt eine zentrale Rolle in der politischen Entscheidungsfindung in der Schweiz. Die Möglichkeit der Bürgerinnen und Bürger, direkt und rechtsverbindlich über politische Sachfragen entscheiden zu können, ist auf allen Staatsebenen stark verankert. Im Gegensatz zu Diskussionen in anderen politischen Systemen, in denen derzeit (wieder) über die Ergänzung der repräsentativen Demokratie mit direkt-demokratischen Elementen diskutiert wird, wird in der Schweiz eher die repräsentative Demokratie als Ergänzung zur direkten Demokratie empfunden. Die repräsentativen Instanzen sind in der Schweiz lediglich soweit legitimiert, wie dies das Volk zulässt – das in der Schweiz auch als «der Souverän» bezeichnet wird.

Die Schweiz ist äusserst kleinteilig organisiert mit etwa 2200 Gemeinden in 26 Kantonen – bei einer ständigen Wohnbevölkerung von lediglich rund 8,5 Millionen Personen. Dabei verfügen die subnationalen Einheiten im schweizerischen *Föderalismus* über außergewöhnlich stark ausgebaute Kompetenzen. Es kommt hinzu, dass die Kantone und die Gemeinden sehr unterschiedlich groß sind. Weil auch die *Dezentralisierung* ausge-

prägt ist, werden auch viele öffentliche Aufgaben auf kantonaler oder kommunaler Ebene erfüllt. Dies stellt aufgrund der Größendifferenzen eine Herausforderung dar.

Der Föderalismus wie auch die Dezentralisierung sind wesentliche Rahmenbedingungen, welche es ermöglichen, das schweizerische Staatswesen auf den jeweiligen Staatsebenen entsprechend den jeweiligen regionalen und lokalen Bedürfnissen auszugestalten. Gemäß dem Subsidiaritätsprinzip sollen die Kompetenzen auf der tiefst möglichen Ebene angesiedelt werden. Aufgrund der großen kulturellen (vier Sprachregionen), sozialen (urbane Zentren versus rurale Peripherie) und geographischen (Flachland versus Berggebiete) *Heterogenität* ergeben sich damit auch erhebliche Unterschiede in der Politikgestaltung – hinsichtlich *polities*, *politics* und *policies*.

Dabei spielt der Schutz von Minderheiten (etwa sprachlicher Art) eine zentrale Rolle. Die Politik ist aufgrund des Minderheitenschutzes, aber auch befördert durch die Einflüsse der direkten Demokratie (Neidhart, 1970), auf eine konsensorientierte Entscheidungsfindung ausgerichtet. Dies zeigt sich etwa an der parteipolitischen Zusammensetzung der Regierungen (auf allen drei Staatsebenen), die typischerweise deutlich mehr als siebzig Prozent der Wählerinnen- und Wählerstimmen auf sich vereinigen. Dieses System der *Konkordanz* löst also die Dualität zwischen Regierung und Opposition auf, wie wir sie von Konkurrenzsystemen kennen.

Schließlich fünftens soll auf die Besonderheit des schweizerischen *Milizsystems* hingewiesen werden. Damit wird ein System bezeichnet, das sich in wesentlichen Teilen auf ein ehrenamtliches respektive nebenamtliches Engagement der Bürgerinnen und Bürger stützt. Außer einer relativ kleinen Zahl von Regierungsämtern auf Bundesebene sowie bei den Kantonen und den größeren Gemeinden (zumeist Städten) werden politische Funktionen ehrenamtlich oder nebenamtlich erfüllt. Auch wenn der Milizcharakter (etwa beim bikameralen Bundesparlament oder auch in den großen Kantonen) immer stärker erodiert (Eberli, Bundi, Frey & Widmer, 2014; Bundi, Eberli & Bütikofer, 2017; Eberli, Bütikofer & Bundi, 2018; 2019; Pilotti, Sciarini, Varone & Cappelletti, 2018; 2019) wird an dieser Tradition zumindest im politischen Diskurs – namentlich von konservativen Kreisen – beharrlich festgehalten.

Neben diesen fünf Merkmalen gäbe es natürlich viele weitere Aspekte zu erwähnen (etwa das Fehlen einer Verfassungsgerichtsbarkeit, das Phänomen der Kollektivregierungen oder die Versammlungsdemokratie, die sich bei zwei Kantonen und auf kommunaler Ebene als persistent erweist). Hierfür sei auf die einschlägige Literatur verwiesen (inter alia Knoepfel, Papadopoulos, Sciarini, Vatter & Häusermann, 2014; Linder & Müller, 2017; Bochler, Hänggli & Häusermann, 2015; Sager, Künzler & Lutz, 2017; Sager & Zollinger, 2011). Zusammenfassend lässt sich festhalten, dass das schweizerische politische System aufgrund der dargelegten Merkmale nicht grundlos als einzig- und eigenartig bezeichnet wurde (Widmer, 2017, S. 53).

### 3 Evaluation in der Schweiz

Die Evaluation hatte in der Schweiz für lange Zeit eine doch eher zögerliche Entwicklung durchlaufen. Vor 1990 spielte die Evaluation kaum eine Rolle in der Schweizer Politik. Dies änderte sich aber deutlich in den 1990er und den 2000er Jahren, in denen die Evaluation in der Schweiz eine dynamische Entwicklung nahm (Widmer, 2017, S. 57-60). Unter anderem hat sich im Jahr 1991 die Bundesversammlung mit der Parla-

mentarischen Verwaltungskontrolle (PVK) einen eigenen Evaluationsdienst geschaffen (Janett, 2004) und seit 2000 enthält die Bundesverfassung (neben den zahlreichen Evaluationsklauseln auf Gesetzesstufe) einen Evaluationsartikel (Art. 170 BV; vgl. unten Kapitel 5.2). Diese Dynamik resultierte darin, dass die Schweiz heute in international vergleichenden Analysen zur Institutionalisierung der Evaluation auf den Spitzenpositionen anzutreffen ist (Jacob, Speer & Furubo, 2015; Stockmann, Meyer & Taube, 2020). Teilweise wird der aktuelle Entwicklungsstand aber auch kritisch kommentiert, namentlich wegen der Inkorporation der «Bürokratie-Technologie» (Widmer, 2017, S. 61) Evaluation in Verwaltungsroutinen, einhergehend mit einem Verlust an wissenschaftlicher Offenheit (Widmer, 2015).

Betrachtet man, wie sich die Evaluationsaktivitäten in der Schweiz vertikal und horizontal verteilen, zeigen sich erhebliche Differenzen: Einerseits lässt sich festhalten, dass die Bundesebene eine deutlich höhere Evaluationsdichte aufweist, als dies die Kantone oder Gemeinden tun. Innerhalb der Kantone wiederum ergeben sich markante Unterschiede zwischen den größeren und kleineren Kantonen (Balthasar & Rieder, 2009; Balthasar & Müller, 2014; Frey, Di Giorgi & Widmer, 2016; 2017). Letztere führen höchstens punktuell Evaluationen durch. Andererseits ist festzustellen, dass es zwischen den Politikfeldern ebenfalls erhebliche Unterschiede in der Intensivität der Evaluationsaktivitäten gibt, und zwar auf allen Staatsebenen. Gewisse Sektoren, namentlich Bildung, Gesundheit und Entwicklungszusammenarbeit, verfügen über eine recht hohe Dichte an Evaluationen, in anderen werden auch heute noch kaum Evaluationen realisiert (siehe zum Beispiel Widmer, 2008b). Wir können somit eine zwar insgesamt stark entwickelte, aber intern ausgeprägt heterogen ausgestaltete Evaluationspraxis konstatieren.

Die Schweiz verfügt mit der Schweizerischen Evaluationsgesellschaft (SEVAL) seit 1996 über eine einschlägige Fachgesellschaft, über verschiedene Aus- und Weiterbildungsangebote für (angehende) Evaluatorinnen und Evaluatoren und über eine Fachzeitschrift (LeGes – Gesetzgebung & Evaluation; Bussmann, 2015). Die Schweizer Evaluationsfachleute haben in den 1990er und 2000er Jahren einen respektablen Grad der Professionalisierung erreichen können (Widmer, 2015; 2017, S. 57-60). Nicht zuletzt besteht in der Schweiz auch eine verhältnismäßig rege Forschungstätigkeit zur Evaluation (Jacob, 2015; Widmer et al., 2016). Auf ein umfangreiches Vorhaben aus diesem Bereich soll nachfolgend genauer eingegangen werden.

## 4 Fragestellung, analytischer Rahmen und Projektstruktur

Schon verhältnismäßig früh brachte Werner Bussmann (1995; 1996) das Argument in die Debatte ein, dass die Schweiz kein guter Nährboden für die Evaluation sei. Aufgrund der Kleinteiligkeit der Staatsstrukturen, aufgrund des Milizsystems mit seiner Bürgernähe und der Rückkopplungsfunktion der direkten Demokratie schätzte er den Bedarf an Evaluation als eher gering ein. Die Nachfrage nach Evaluation sei eher klein, da «policy making in Switzerland is oriented toward gathering support more than toward gathering evidence» (Widmer & Neuenschwander, 2004, p. 391). Zudem wurde auf den über lange Zeit schleppenden Ausbau des Wohlfahrtsstaates in der Schweiz hingewiesen, der die Nachfrage nach Evaluationen namentlich zu sozialstaatlichen Interventionen zusätzlich gehemmt haben soll (Widmer, 2016).

Trotz dieser *prima facie* eher hinderlichen Konditionen hat sich, wie oben gezeigt, die Evaluation in der Schweiz etabliert (siehe auch Horber-Papazian & Baud-Lavigne, 2019). Dieses Paradox veranlasste eine Gruppe von Wissenschaftlerinnen und Wissenschaftlern ein Forschungsprogramm zu lancieren, das sich mit den Wechselwirkungen von politischem System und Evaluation befassen sollte (siehe auch Balthasar, Sager & Widmer, 2017).

Das Forschungsprogramm «Policy Evaluation in the Swiss Political System – Roots and Fruits», kurz SynEval (siehe auch [www.syneval.ch](http://www.syneval.ch)), untersucht die Fragen, ob die Politikevaluation vom politischen System der Schweiz geprägt wird und wie wiederum die Politikevaluation das politische System der Schweiz beeinflusst. Dieses Thema ist sowohl für die Politik- und Verwaltungswissenschaft als auch für die politische Praxis und die öffentliche Verwaltung von besonderer Bedeutung, da bis anhin fundierte Untersuchungen über die Beziehung zwischen spezifischen Eigenschaften von politischen Systemen einerseits und der Praxis und Institutionalisierung von Politikevaluationen andererseits weitgehend fehlten. Aus der Perspektive der Praxis ist dieses Thema besonders relevant, da sich die Evaluation als Instrument der staatlichen Steuerungen etabliert hat und jedes Jahr beträchtliche Ressourcen in Evaluationsaktivitäten investiert werden. Es bestand bereits vor der Initiierung von SynEval ein respektabler Forschungsstand, der jedoch auch zahlreiche Fragen offen ließ (etwa Klöti & Schneider, 1989; Busmann, 1995; 1996; 2005; Busmann, Klöti & Knoepfel, 1997; Widmer, 1996; 2002; 2007; 2008a; 2008b; Widmer, Rüegg & Neuenschwander, 2001; Widmer & Neuenschwander, 2004; Balthasar, 2007; Horber-Papazian & Bützer, 2008; Mader, 2009; Widmer & Beywl, 2009; Frey, 2010; 2012; Frey & Widmer, 2011; 2013; Horber-Papazian & Jacot-Descombes, 2012; Widmer & De Rocchi, 2012; Ledermann, 2012; 2014; für einen Überblick: Widmer et al., 2016).

Das politische System der Schweiz eignet sich ideal zur Untersuchung dieser Fragen. Erstens erlaubt die hohe Heterogenität der kantonalen Systeme Vergleiche innerhalb des politischen Systems der Schweiz. Zweitens verfügt das schweizerische politische System über zwei Eigenschaften, die besonders relevant sind in Bezug auf Politikevaluation: Föderalismus und direkte Demokratie. In beiden Fällen bestehen neben den oben dargelegten Thesen zu einer hemmenden Wirkung auf die Evaluation auch Argumente, die eine gesteigerte Nachfrage nach Politikevaluationen vermuten lassen. So sollen föderalistische Systeme besonders von Evaluationen profitieren, weil sich so von den gewonnenen Erfahrungen in anderen Einheiten profitieren lässt. Weiter soll die Evaluation auch die direkte Demokratie unterstützen, weil sie dazu beitragen kann, den Informationsstand der Stimmbürgerinnen und Stimmbürger zu anstehenden Sachfragen zu verbessern.

Ebenso besteht auch kein Konsens über den Einfluss der Evaluation auf das politische System der Schweiz. Bisher existierte auch keine systematische Analyse der Interdependenzen zwischen den Eigenschaften des politischen Systems der Schweiz und der Politikevaluation, im Besonderen auf subnationaler Ebene. Das Projekt SynEval analysiert deshalb diese Beziehungen, schwergewichtig anhand von Vergleichen auf der kantonalen Ebene.

Im Zentrum des Forschungsprogramms SynEval stehen folgende drei Leitfragen:

- Wie haben sich die Evaluationskulturen im politischen System der Schweiz während der letzten zwei Jahrzehnte entwickelt?
- Wieso haben sie sich auf die beobachtete Art entwickelt?

- Was sind die Konsequenzen dieser Entwicklungen für das politische System der Schweiz?

Der Begriff Evaluationskultur – im Nachfolgenden wird auch von Evaluationsfunktion oder schlicht Evaluation gesprochen – wird hier in einem weiten Sinne verstanden und umfasst zahlreiche Aspekte der Evaluation, einschliesslich der folgenden:

- Angebots- und nachfrageseitige Evaluationskapazitäten,
- Grad der Professionalisierung der Evaluation,
- institutionelle Einbettung der Evaluation einschliesslich dezidierter Einrichtungen und Normen,
- Institutionen (wie Regeln und Organisationen, aber auch Normen und Werthaltungen) und Praktiken der Evaluation,
- Umfang und Ausgestaltung der Evaluationsaktivitäten,
- Umfang und Art von Evaluationsbefunden und
- Evaluationsqualität.

Das Forschungsprogramm SynEval orientierte sich an einem analytischen Rahmen (*Abbildung 1*), welcher die Evaluation (resp. die Evaluationskultur) ins Zentrum stellt und deren Verhältnis zu den drei Aspekten der Politik (*polity*, *politics* und *policy*, vgl. Einleitung) umfasst. Dabei interessieren sowohl die Auswirkungen der Politikaspekte auf die Evaluation wie auch die Auswirkungen der Evaluation auf die Politikaspekte.

*Abbildung 1:* Analytischer Rahmen des Forschungsprogramms SynEval



Quelle: nach Balthasar, Sager & Widmer, 2017, p. 11.

Das Forschungsprogramm SynEval wurde als Forschungsnetzwerk unter meiner Leitung realisiert und umfasste vier Projekte. Die Projekte sind an Universitäten der deutsch- und französischsprachigen Schweiz angesiedelt. Während je ein Projekt an den Universitäten Lausanne, Bern und Zürich realisiert wurde, bestand für ein weiteres Projekt eine Kooperation zwischen den beiden Universitäten Luzern und Genf.

SynEval startete im Januar 2013 und dauerte bis Dezember 2016. Im Rahmen des Forschungsprogramms entstanden unter anderem sieben Promotionen sowie viele Publikationen. Neben den zahlreichen Analysen haben die beteiligten Forschenden auch systematisch bestehende Evaluationsstudien identifiziert und deren Charakteristiken in einer Datenbank erfasst (siehe [www.syneval.ch](http://www.syneval.ch); Frey, Di Giorgi & Widmer, 2016).

Aufgrund des beträchtlichen Umfangs der Forschungsarbeiten muss die nachstehende Präsentation der Analysen und Befunde selektiv bleiben. Für weiterführende Informationen sei auf die zitierte Literatur verwiesen (besonders Sager, Widmer & Balthasar, 2017b).

## 5 Inhalte und ausgewählte Befunde

### 5.1 Projekt 1: Evaluationskultur nach Politikbereich und ihre Einflussfaktoren

Dieses Projekt unter der Leitung von Andreas Balthasar (Universität Luzern) und Frédéric Varone (Universität Genf) befasste sich mit der Frage, wie Merkmale von Politikfeldern und Verwaltungseinheiten die Evaluation und besonders die Evaluationskultur beeinflussen (*Abbildung 1*, Zusammenhang 1).

Bei der Bundesverwaltung werden in letzter Zeit insgesamt durchschnittlich rund achtzig bis hundert Evaluationsstudien pro Jahr realisiert, wobei kein klarer Trend über die Zeit zu erkennen ist (Balthasar & Strotz, 2017, S. 90-91). Die Erhebungen zeigen auf, dass Evaluationen bei zahlreichen Bundesverwaltungsstellen verankert sind, jedoch weiterhin nicht bei allen Amtsstellen. Eine amts- und departementsübergreifende (also über die Ministerien hinweg bedeutende) Evaluationskultur fehlt in der Bundesverwaltung weiterhin, Attitüden und Praktiken sind divers, was sich nur partiell sachlich begründen lässt (Balthasar & Strotz, 2017, S. 92-96). In vertiefenden Fallstudien wird die Entwicklung der Evaluation in drei ausgewählten Amtsstellen (Bundesamt für Gesundheit, Bundesamt für Energie, Eidgenössische Finanzverwaltung) anhand eines Vergleichs mit einer früheren Studie (Widmer, Rüegg & Neuenschwander, 2001) nachgezeichnet (Balthasar & Strotz, 2017, S. 98-113). Es lässt sich festhalten, dass Politikfelder respektive Verwaltungseinheiten durchaus einen Unterschied darstellen für den Umfang der Evaluationsaktivitäten. Diese Differenzen lassen sich aber nur teilweise systematisch auf Eigenschaften der Politikfelder zurückführen.

Auf der kantonalen Ebene (dazu Dolder, Rohrbach & Varone, 2017) lässt sich im Vergleich zur Bundesebene eine zögerlichere, volatilere und weniger persistente Etablierung einer Evaluationskultur beobachten. Dabei gehen die Autoren weiter, als dies in vielen Studien zur Evaluationskultur der Fall ist, und inkludieren in ihr Konzept der Evaluationskultur nicht nur die Dimensionen der Institutionalisierung der Evaluation und der Evaluationspraxis, sondern auch die Einstellungen der Mitarbeitenden der entsprechenden Verwaltungseinheiten zur Evaluation. Die Analysen machen aber deutlich, dass diese Einstellungsdimension auf individueller Ebene zur Erklärung von Differenzen zwischen den Verwaltungseinheiten von untergeordneter Relevanz zu sein scheint. In einem Vergleich der Entwicklung von 2000 bis 2013 in den Politikfeldern Bildung und Gesundheit in den Kantonen Basel-Stadt, Bern und Luzern, der sich auf Leitfadeninterviews, Internet-Recherchen und eine Online-Umfrage stützt (Dolder, Rohrbach & Varone, 2017, S. 148-150), zeigen die Forschenden auf, dass die Institutionalisierung die Evaluation im Bildungsbereich stark beeinflusst, während der Gesundheitsbereich eher durch quantitatives Wachstum der Evaluationsaktivitäten geprägt ist. Die Forschenden stellen weiter fest, «dass die Evaluationskultur zwischen Politikfeldern stärker variiert als zwischen Kantonen» (Dolder, Rohrbach & Varone, 2017, S. 145), was gerade auch für international vergleichende Studien, die in der Regel von einer national einheitlichen Ausgestaltung ausgehen, einen bemerkenswerten Befund darstellt. Aufgrund der vorliegenden Empirie postulieren sie weiter, dass sich in den Kantonen politikfeldspezifische Entwicklungspfade beobachten lassen (Dolder, Rohrbach & Varone, 2017, S. 143). «Namentlich scheint die Bedeutung des Politikfeldes für Veränderungen in der Evaluationskultur



wichtiger zu sein als die verschiedenen kantonalen Kontexte» (Sager, Widmer & Balthasar, 2017a, S. 317). Zudem untersuchte das Projekt die vielfach postulierten Wechselwirkungen zwischen Staatsreformen, die Konzepten des *New Public Management* folgen, einerseits und der Evaluationskultur andererseits (Dolder, 2019).

## 5.2 Projekt 2: Evaluationsklauseln und deren Implementierung in der Schweiz

Das zweite Projekt befasst sich mit einer spezifischen Form der Institutionalisierung der Evaluationskultur, nämlich mit sogenannten Evaluationsklauseln. Darunter werden rechtliche Normen verstanden, welche die Durchführung von Evaluationen einfordern (Bussmann, 2005). Die in der Schweiz bekannteste Evaluationsklausel genießt Verfassungsrang und legt seit 2000 fest, dass die Bundesversammlung (also das Bundesparlament mit den zwei gleichberechtigten Kammern Nationalrat und Ständerat) dafür sorgt, dass die Massnahmen des Bundes auf ihre Wirksamkeit überprüft werden (Art. 170 der Bundesverfassung; Mader, 2005; Widmer & De Rocchi, 2012, S. 120-125). Daneben bestehen aber zahlreiche weitere Evaluationsklauseln in nachgeordnetem Rechtstexten, namentlich Gesetzen und Verordnungen, die in ihrer Summe wohl deutlich relevanter sind als der doch recht offen formulierte Verfassungsartikel (Widmer & De Rocchi, 2012, S. 125).

Das von Katia Horber-Papazian und Luzius Mader (beide Universität Lausanne) geleitete Projekt im Rahmen von SynEval geht der Frage nach, welche Ursachen und Konsequenzen Evaluationsklauseln auf Bundesebene und bei den Kantonen haben. Konkret befasst sich dieses Projekt demnach mit den Effekten von *Polity* und *Politics* auf Evaluationsklauseln sowie den Effekten von Evaluationsklauseln auf Politikinhalte (*Abbildung 1*, Zusammenhänge 2, 3 und 4).

Aufgrund der Erhebungen halten die Autorin und die Autoren zunächst fest, dass die Zahl an Evaluationsklauseln sowohl auf Bundesebene als auch auf kantonaler Ebene in den letzten drei Dekaden kontinuierlich zugenommen habe (Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 158; Wirths & Horber-Papazian, 2016). Vermutlich wird dabei das Wachstum etwas überzeichnet, da die Erhebung auf den Entstehungszeitpunkt der aktuell gültigen Evaluationsklauseln abstellt und bereits außer Kraft gesetzte oder erneuerte Klauseln nicht berücksichtigt.

Insgesamt bestanden 2013 auf Bundesebene 53, und bei den Kantonen insgesamt 269 Evaluationsklauseln (also durchschnittlich gut zehn pro Kanton). Die Zahl der Evaluationsklauseln bei den Kantonen variiert stark, wobei große Kantone (Zürich und Bern) sowie Kantone der Romandie höhere Zahlen aufweisen (Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 160-162). Weiter lassen sich sowohl auf eidgenössischer wie auf kantonaler Ebene deutliche Differenzen in der Zahl der Klauseln zwischen den Politikbereichen konstatieren. Besonders häufig treten Evaluationsklauseln bei den Kantonen in den Politikfeldern Bildung, Gesundheit und Sozialhilfe auf, während auf Bundesebene die Bereiche Staatspolitik und politische Rechte, Sozialversicherung sowie Wirtschaftspolitik relativ häufig auftreten (Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 162-164). Es handelt sich dabei um Politikfelder, denen auf der entsprechenden Staatsebene eine gewisse Bedeutung zukommt. Man kann daraus also schliessen, dass sich die Häufigkeit von Evaluationsklauseln unter anderem an der Kompetenzverteilung orientiert. Selbstverständlich sind diese Feststellungen dadurch zu relativieren, dass die Reichweite und

Bedeutung von Evaluationsklauseln sehr unterschiedlich ausgestaltet sein können, was die Aussagekraft einer nicht-gewichteten Auszählung limitiert.

Evaluationsklauseln sind sehr unterschiedlich ausgestaltet (Wirths, 2017; Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 168-170). Besonders fallen dabei die Differenzen in der normativen Dichte auf, also inwiefern die Klauseln festlegen, was, wann, zu welchem Zweck und mit welchen Evaluationskriterien evaluiert werden soll, wer den Evaluationsbericht wem präsentieren und wer die Evaluation verantworten soll. Die Autorin und die Autoren entwickeln daraus drei typische Kombinationen dieser Merkmale, nämlich Klauseln mit tiefer normativer Dichte und Fokus auf Evaluationskriterium (160 von 322 Fällen), Klauseln mit tiefer normativer Dichte und Fokus auf Evaluationszeitpunkt (90 von 322) sowie Klauseln mit hoher normativer Dichte, die viele der genannten Festlegungen enthalten (72 von 322; Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 168-169). Darüber hinaus kann festgehalten werden, dass Evaluationsklauseln häufig die Zuständigkeit für die Berichterstattung sowie den Evaluationsgegenstand regeln, hingegen selten den Zweck der Evaluation und die evaluationsverantwortliche Behörde (Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 168).

Als Argumente zur Einführung von Evaluationsklauseln werden folgende genannt: Evaluationen sollen als Instrumente der Leistungssteuerung respektive der Kontrolle der Exekutive dienen und sie sollen zur Reduktion von Unsicherheiten beitragen. Evaluationsklauseln werden aber auch als Mittel der Konsensfindung eingesetzt, um kritischen Akteuren die Annahme eines kontroversen Vorschlags zu erleichtern. Weiter beobachten die Forschenden aber auch Fälle, bei denen Evaluationsklauseln dazu eingesetzt werden, um eine kontinuierliche Beobachtung in dynamischen Problemfeldern sicherzustellen oder um Verantwortlichkeit gegenüber bestimmten Adressatengruppen festzuhalten (Wirths, 2016; Wirths, Rosser, Horber-Papazian & Mader, 2017, S. 165-167).

### 5.3 Projekt 3: Direkt-demokratische Kampagnen und Evaluation

Dieses Projekt befasste sich unter der Leitung von Fritz Sager (Universität Bern) mit dem Verhältnis von Evaluation und direkter Demokratie. Es rückt hier also ein zentrales Element des schweizerischen politischen Systems in den Fokus – die Möglichkeit resp. die Verpflichtung, politische Sachentscheide durch die Stimmbürgerinnen und die Stimmbürger treffen zu lassen. Aufgrund der zentralen Bedeutung direkt-demokratischer Entscheide im schweizerischen politischen System (vgl. oben, Kapitel 2) sind diese auch prägend für den öffentlichen politischen Diskurs. Ein wesentlicher Anteil der massenmedialen Berichterstattung befasst sich in der Schweiz mit direkt-demokratischen Entscheiden. Die Frage nach der Rolle von Evaluationsbefunden in direkt-demokratischen Kampagnen stand in dem dritten Projekt von SynEval im Zentrum (vgl. Schlauffer, Stucki & Sager 2018). Somit konzentrierte sich dieses Projekt auf die Wirkung von Evaluation auf politische Prozesse (vgl. oben, Abb. 1, Zusammenhang 6).

Grundsätzlich sind die Projektergebnisse hinsichtlich der gesellschaftlichen Relevanz von Evaluationen ernüchternd. Evaluationsbefunde spielen in direkt-demokratischen Kampagnen eine marginale Rolle. Aufgrund der Inhaltsanalysen der Medienberichterstattung in Tageszeitungen zu Abstimmungskampagnen in den Politikfeldern Bildung und Gesundheit stellen die Autorinnen fest, dass lediglich gut zwei Prozent der insgesamt 11.128 Artikel auf Evaluationsbefunde Bezug nehmen – in der Gesundheitspolitik mit 2,6 Prozent leicht mehr als im Bereich Bildung mit 1,4 Prozent (Stucki & Schlauffer,

2017, S. 286-287). Etwas häufiger erscheinen Evaluationsbefunde in der Abstimmungsbroschüre, in welcher die Behörden die Stimmbürgerinnen und Stimmbürger über die Inhalte der Abstimmungsvorlagen und über die jeweiligen Pro- und Kontraargumente informieren (Schlaufer & Stucki, 2017) sowie wenn sich Expertinnen und Experten in den Medien äußern (Stucki & Schlauffer, 2017, S. 286-288; Stucki, 2016).

In einer vertiefenden Fallstudie wird anhand von Abstimmungen zum Schutz vor Passivrauchen das Zusammenspiel von unterschiedlichen Argumenten im öffentlichen Diskurs aufgezeigt. Dabei zeigt sich die zentrale Rolle von Narrativen bei der Evidenznutzung in der öffentlichen Debatte (Stucki, 2017; 2018). Weiter machen die Analysen deutlich, dass Evaluationsbefunde besonders häufig von Expertinnen und Experten zur Steigerung ihrer Glaubwürdigkeit in die Debatte eingebracht werden, dass diese aber insgesamt eher selten im öffentlichen Diskurs auftreten (Stucki, 2016).

Im Rahmen einer Fallstudie zur Bildungspolitik wurde zudem analysiert, welche Rolle die Nutzung von Evidenz aus Evaluationen für die Diskursqualität hat. Dabei zeigt sich die dominante Rolle von Befunden aus den Erhebungen des «Programme for International Student Assessment» der OECD (PISA; Schlauffer, 2016; 2018b). Die Analysen zeigen weiter, dass Diskurse, welche auf Evaluationen Bezug nehmen, eine höhere Diskursqualität aufweisen als jene, die keine Verweise auf Evaluationen enthalten (Schlaufer, 2018a; Stucki & Schlauffer, 2017, S. 297-302; Stucki & Sager, 2018).

Weiter zeigen die Forschenden in einer weiteren Studie, dass Stimmbürgerinnen und Stimmbürger unter bestimmten Bedingungen durchaus daran interessiert sind, sich über die zur Entscheidung anstehenden Sachverhalte zu informieren und dass dies einen Beitrag zur sachlichen Entscheidungsfindung darstellen kann (Stucki, Pleger & Sager, 2018).

Aus diesen Befunden leiten die Autorinnen nachstehende Folgerung ab: «Wenn sich Evaluierende stärker in den direkt-demokratischen Diskurs einbringen, können Evaluationen vermehrt in die Abstimmungskampagnen hineingebracht werden» (Stucki & Schlauffer, 2017, S. 303-304).

## 5.4 Projekt 4: Schweizer Parlamente und Evaluation

Im vierten Projekt des SynEval-Forschungsprogramms, das von einem Team an der Universität Zürich unter meiner Leitung realisiert wurde, steht die Legislative im Zentrum des Interesses. Während bereits recht viele Studien zur Rolle der Evaluation bei Regierung und Verwaltung bestanden, sind Analysen zur Rolle der Legislative selten. Diese Untersuchung fragt deswegen nach der Bedeutung von Parlamenten bei der Veranlassung, bei der Realisierung und bei der Nutzung von Evaluationen. Es befasst sich also sowohl mit der Genese wie auch mit der Nutzung der Evaluation im institutionellen Rahmen der Parlamente auf Ebene des Bundes und der Kantone, die nach wie vor weitgehend als Milizgremien konstituiert sind (vgl. oben, Abschnitt 2; Eberli, Bundi, Frey & Widmer, 2014; Bundi, Eberli & Bütikofer, 2017; Eberli, Bütikofer & Bundi, 2018; 2019; Pilotti, Sciarini, Varone & Cappelletti, 2018; 2019). Generell lässt sich feststellen, dass sich die Forschung über Evaluation bisher auf die Frage nach der Nutzung respektive – in der neueren Literatur – nach dem Einfluss von Evaluationen ausgerichtet hat, erhielt die Nachfrage nach respektive die Auslösung von Evaluationen bisher weniger Beachtung (Widmer et al., 2016; Hense & Widmer, 2013).

Eine Umfrage unter den Mitgliedern der 26 Kantonsparlamente sowie der beiden Kammern des Bundesparlaments ergab, dass die Abgeordneten eher parlamentarische

Vorstöße als Mittel zur Veranlassung von Evaluationen nutzen (Bundi, Eberli, Frey & Widmer, 2014; Eberli, Bundi, Frey & Widmer, 2014). Seltener tun sie dies unter Verwendung von Evaluationsklauseln. Letztere werden in der Regel im vorparlamentarischen Verfahren von der Exekutive in die Gesetzgebung eingeführt (Eberli & Bundi, 2017; Wirths, Rosser, Horber-Papazian & Mader, 2017). Mitglieder von National- und Ständerat, also den beiden Kammern des Bundesparlaments, verwenden meist die Form einer Interpellation, um eine Evaluation mittels eines parlamentarischen Vorstoßes zu initiieren (Bundi, Varone, Gava & Widmer, 2018). Die Parlamentarierinnen und Parlamentarier fordern Evaluationen vor allem zur Ausübung ihrer Kontrollfunktion gegenüber Exekutive und Verwaltung (Bundi, 2016). Dabei spielen Interessenbindungen der Politikerinnen und Politiker eine nicht unwesentliche Rolle: Sie verlangen häufiger nach Evaluationen in Politikfeldern, in denen sie über Verbindungen zu einschlägigen Interessengruppen verfügen (Varone, Bundi & Gava, 2020). Erfahrene Parlamentsmitglieder sowie jene in Führungspositionen fordern häufiger Evaluationen als ihre Kolleginnen und Kollegen (Bundi, 2016). Sie tun dies vor allem in Politikbereichen, in denen der Legitimationsdruck besonders hoch ist und in denen die Erbringung öffentlicher Dienstleistungen an Dritte delegiert wird (Bundi, 2018b). Damit scheinen für die Parlamentsmitglieder ähnliche Gesichtspunkte handlungsrelevant, wie dies auch bei anderen Akteuren, namentlich der Verwaltung, postuliert wird (Widmer, 2008b).

Während sich frühere Studien (eher am Rande) auch mit der Nutzung von Evaluationen im Parlament befasst haben (z. B. Balthasar, 2007; 2009; Spinatsch, 2008; Frey, 2010; 2012; Frey & Widmer, 2011; 2013; Ledermann 2012; 2014; Balthasar & Müller, 2014; Frey & Ledermann, 2010; 2017; Rissi & Sager, 2013), konzentrierte sich Daniela Eberli (2019) in ihrer Studie ausschliesslich auf diese. Sie zeigt, dass Evaluationen als Teil des Gesetzgebungsprozesses sowohl analytisch (zur Verbesserung des Gesetzes- oder Antragsentwurfs) als auch politisch (zur Unterstützung der eigenen Position) genutzt werden. Sie weist auch auf die beträchtliche Bedeutung hin, die Evaluationsbefunden als Evidenz in parlamentarischen Kommissionen zugeschrieben wird (Eberli, 2018). Ferner zeigen die Ergebnisse, dass die Nutzung von Evaluationen im vorparlamentarischen Prozess (durch die Verwaltung) die Nutzung in der parlamentarischen Phase prägt, also eine gewisse Pfadabhängigkeit in der Evaluationsnutzung zu beobachten ist (Eberli, 2018).

Die im Rahmen dieses Projektes realisierten Analysen machen aber auch deutlich, dass von einer umfassenden, systematischen Nutzung von Evaluation in den Parlamenten bei weitem nicht gesprochen werden kann – was aber wohl, gerade im Falle von Milizparlamenten, aber auch darüber hinaus, eine unangemessen hohe Erwartung darstellt (siehe Widmer, 2009). Vielmehr handelt es sich um eine selektive Nutzung, sowohl im Hinblick darauf, welche Evaluationsergebnisse genutzt werden, als auch welche Parlamentarierinnen und Parlamentarier Evaluationen nutzen (Eberli, 2019).

Zentrales Ergebnis der Analysen bildet aber auch der Befund, dass Evaluationen selbst wesentlich häufiger angefordert werden, als dass ihre Ergebnisse tatsächlich genutzt werden (Bundi, 2018a). Parlamente in der Schweiz nutzen also eher *Evaluationsprozesse* als *Evaluationsbefunde*. Neben einer partiellen Delegation der Aufsichtsfunktion an die evaluierenden Instanzen (man könnte auch von einem Outsourcing sprechen), scheint man auch wesentlich auf die Vorwirkungen und auf den Prozessnutzen (Patton, 1998) von Evaluationen abzielen.

## 6 Fazit

Insgesamt ist die Evaluation in der Schweiz gut etabliert. Institutionen und Praktiken haben sich seit den 1990er Jahren stark (weiter-)entwickelt. Obwohl dies – auch vor dem Hintergrund der Thesen, welche die Eignung der Evaluation für das schweizerische politische System in Frage stellen – bemerkenswert ist, weist die Evaluation in der Schweiz auch bedeutende Schwächen auf. Gerade im Hinblick auf die zukünftige Ausgestaltung der Evaluation, aber auch der Forschung über Evaluation drängen sich aus meiner Sicht folgende weiterführende Überlegungen auf (dazu Sager, Widmer & Balthasar, 2017a):

Die Ausgestaltung der Evaluation ist in der Schweiz nach wie vor ausgesprochen heterogen. Wir treffen regional und sektoral deutlich distinkte Praktiken, Qualitäten, Institutionen und Kulturen der Evaluation an. Diese Vielfalt ist zwar durchaus typisch für die Schweiz und daher nicht besonders überraschend, aber man kann zeigen, dass die angetroffene Heterogenität weder demokratisch legitimiert noch objektiv zu rechtfertigen ist. Vielmehr hat man den Eindruck einer gewissen Zufälligkeit und/oder Willkür, ob, geschweige denn wie in einer gegebenen Situation eine Evaluation durchgeführt werden soll. Eine außenstehende Person würde wohl eine systematischere Herangehensweise empfehlen. Hier könnte die Forschung über Evaluation einen wertvollen Beitrag leisten, da sie eine Reflexion über die relevanten Strukturen und Prozesse ermöglicht und Vorschläge machen kann, wie eine systematischere, zielorientierte und effektive Ausgestaltung der Evaluationspraxis besser unterstützt werden kann.

Der auch international zu beobachtenden Tendenz, die Forschung über Evaluation primär auf die Nutzung (oder den Einfluss) von Evaluationsbefunden auszurichten, ist entgegenzuwirken, indem sich die Forschung noch stärker mit der Erklärung der Evaluationskultur befassen sollte. Wir wissen noch sehr wenig darüber, weshalb, wann, unter welchen Bedingungen, durch wen, sich welche Evaluationskultur entwickelt. In diesem Bereich bestehen in der Schweiz, aber auch international, weiterhin erhebliche Forschungslücken.

Die realisierten Forschungsarbeiten haben sich zwar darum bemüht, sich auch zu wesentlichen Teilen mit der subnationalen Politik zu befassen. Dies ist im Hinblick auf die kantonale Ebene nicht schlecht gelungen, wobei natürlich auch zur Rolle der Evaluationsfunktion bei den Kantonen weiterhin noch viele Fragen offen sind. Weitgehend außen vor blieb jedoch die kommunale Ebene, zu welcher der Wissensstand weiterhin ausgesprochen schwach ist. Es bestehen hierzu nur sehr punktuelle Kenntnisse, die auf keiner systematischen Basis beruhen. Aufgrund der Kleinteiligkeit der kommunalen Strukturen ist zwar davon auszugehen, dass die Evaluationskapazitäten (hinsichtlich Produktion und Nutzung) bei kleineren Gemeinden sehr gering sind. Anders verhält es sich aber bei den (für schweizerische Verhältnisse) größeren Städten, bei denen durchaus Potentiale bestehen würden – wir aber kaum etwas darüber wissen, ob und wenn ja, wie diese auch genutzt werden.

Ein weiterer Aspekt, der im Rahmen der vorgestellten Forschungsprojekte im Rahmen des SynEval-Programms kaum Beachtung fand, sind die internationalen Bezüge. Einerseits wissen wir bisher nur wenig über die Zusammenhänge zwischen nationalen Evaluationskulturen. Es gibt zwar gewisse Überlegungen zu den internationalen Einflüssen, die aber bisher nicht systematisch analysiert wurden. Dies ist eine bedeutende Forschungslücke mit zunehmender Relevanz, besonders für den Kleinstaat Schweiz.

Die Bemühungen um eine Professionalisierung der Evaluation in der Schweiz sind aus zwei Gründen auch mit Skepsis zu betrachten. Obwohl sich die Evaluation weiter-

entwickelt hat, fehlt es bislang an einer angemessenen Anerkennung der eigentlichen Evaluationsexpertise. In vielen Bereichen besteht nach wie vor die falsche Vorstellung, dass die Kenntnis eines bestimmten politischen oder institutionellen Kontextes ausreicht, um eine gute Evaluation durchzuführen. Das Ergebnis ist, dass es immer noch zu zahlreichen, von Laien durchgeführten Evaluationen kommt, die gravierende evaluationsfachliche Schwächen aufweisen. Um der mangelnden Anerkennung der Evaluationsfachlichkeit entgegenzuwirken, hat die Evaluationsgemeinschaft zudem in letzter Zeit eine Strategie der bürokratischen «Routinisierung» und der Abschottung des Marktes verfolgt. Dabei wird jedoch übersehen, dass die Evaluation für eine gesunde Weiterentwicklung zwingend auf einen regen Austausch zwischen Evaluationspraxis und sozialwissenschaftlicher Forschung angewiesen ist.

Der einleitende Hinweis in diesem Fazit auf die Heterogenität der Evaluationsausgestaltung ist nicht so zu verstehen, dass per se eine homogenere Ausgestaltung der Evaluationsfunktion erstrebenswert wäre. Vielmehr plädiere ich hier für ein besseres Verständnis dieser Heterogenität. Dies geschieht nicht nur aus wissenschaftlichen Beweggründen, sondern durchaus auch um uns in die Lage zu versetzen, aufgrund eines besseren Verständnisses der Erklärungsfaktoren einer Evaluationskultur Hinweise für eine bedürfnisgerechte Ausgestaltung der Evaluationsfunktion abzuleiten. Nicht zuletzt besteht doch das Ziel darin, dass die Evaluation ihren Beitrag zur demokratischen Gestaltung einer leistungsfähigen Politik zielführend erbringen kann.

Die hier dargelegten Analysen, Befunde und weiterführenden Überlegungen sind geprägt durch den Kontext des schweizerischen politischen Systems. Auch wenn sie deswegen nicht unbesehen auf andere politische Kontexte übertragen werden sollten, soll dieser Beitrag mit dem Wunsch schliessen, dass die dargelegten Inhalte zu weiterführenden Untersuchungen veranlassen – nicht nur in der Schweiz.

## Anmerkungen

- 1 Ich danke den am Forschungsprogramm SynEval beteiligten Forscherinnen und Forschern für die Zusammenarbeit und dem Schweizerischen Nationalfonds für die finanzielle Förderung. Mein Dank geht auch an die Herausgeber sowie an die zwei anonymen Gutachtenden für die Hinweise zur Verbesserung des Beitrags. Für die verbliebenen Schwächen trage ich alleine die Verantwortung.
- 2 SNF Projektnummer 141893, siehe <http://p3.snf.ch/Project-141893> [23.3.2020]

## Literatur

- Balthasar, Andreas (2007). *Institutionelle Verankerung und Verwendung von Evaluationen: Praxis und Verwendung von Evaluationen in der schweizerischen Bundesverwaltung*. Zürich: Rüegger.
- Balthasar, Andreas (2009). Evaluationen in der Schweiz: Verbreitung und Verwendung. In Thomas Widmer, Wolfgang Beywl & Carlo Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 486-497). Wiesbaden: VS Verlag.
- Balthasar, Andreas & Müller, Franziska (2014). Die Verbreitung evidenzbasierter und gleichstellungssensitiver Informationen in den Entscheidungsprozessen kantonaler Steuer- und Sozialtransferpolitik: eine quantitative Analyse. *Swiss Political Science Review*, 20 (1), 70-95.
- Balthasar, Andreas & Rieder, Stefan (2009). Wo ist evidenzbasierte Politik möglich? Die Verbreitung von Evaluationen auf kantonaler Ebene. In Adrian Vatter, Frédéric Varone & Fritz Sager (Hrsg.), *Demokratie als Leidenschaft* (S. 403-429). Bern: Haupt.

- Balthasar, Andreas, Sager, Fritz & Widmer, Thomas (2017). Einleitung. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 9-20). Zürich: NZZ Libro.
- Balthasar, Andreas & Strotz, Chantal (2017). Verbreitung und Verankerung von Evaluation in der Bundesverwaltung. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 89-117). Zürich: NZZ Libro.
- Bochsler, Daniel, Regula Hänggli & Silja Häusermann (Eds.) (2015). Consensus lost? Disenchanted democracy in Switzerland. Special Issue. *Swiss Political Science Review*, 21 (4).
- Bundi, Pirmin (2016). What do we know about the demand for evaluation? Insights from the parliamentary arena. *American Journal of Evaluation*, 37 (4), 522-541.
- Bundi, Pirmin (2018a). Parliamentarians' strategies for policy evaluations. *Evaluation and Program Planning*, 69, 130-138.
- Bundi, Pirmin (2018b). Varieties of accountability: How attributes of policy fields shape parliamentary oversight. *Governance*, 31 (1), 163-183.
- Bundi, Pirmin, Eberli, Daniela & Bütikofer, Sarah (2017). Between occupation and politics: Legislative professionalization in the Swiss cantons. *Swiss Political Science Review*, 23 (1), 1-20.
- Bundi, Pirmin, Eberli, Daniela, Frey, Kathrin & Widmer, Thomas (2014). *Befragung Parlamente und Evaluationen. Methodenbericht*. Zürich: Institut für Politikwissenschaft, Universität Zürich. Verfügbar unter: [https://www.syneval.ch/images/PDF/Methodenbericht\\_ParEval\\_web.pdf](https://www.syneval.ch/images/PDF/Methodenbericht_ParEval_web.pdf)
- Bundi, Pirmin, Varone, Frédéric, Gava, Roy & Widmer, Thomas (2018). Self-selection and misreporting in legislative surveys. *Political Science Research and Methods*, 6 (4), 771-789.
- Bussmann, Werner (1995). Evaluations and grassroots politics: The case of Switzerland. *Knowledge and Policy*, 8 (3), 85-98.
- Bussmann, Werner (1996). Democracy and evaluation's contribution to negotiation, empowerment and information. *Evaluation*, 2 (3), 307-319.
- Bussmann, Werner (2005). Typen und Terminologie von Evaluationsklauseln. *LeGes – Gesetzgebung & Evaluation*, 16 (1), 97-102.
- Bussmann, Werner (2015). Konstanz, Wandel und Neuentwicklungen: 25 Jahre Evaluation in LeGes. *LeGes – Gesetzgebung & Evaluation*, 26 (1), 107-125.
- Bussmann, Werner, Klöti, Ulrich & Knoepfel, Peter (Hrsg.) (1997). *Einführung in die Politikevaluation*. Basel: Helbing & Lichtenhahn.
- Dolder, Olivier (2019). *Wie beeinflussen sich Evaluation und New Public Management? Analyse am Beispiel von Schweizer Kantonsverwaltungen*. Bern: Haupt.
- Dolder, Olivier, Rohrbach, Walter & Varone, Frédéric (2017). Evaluationskultur auf kantonaler Ebene: politikfeld- oder kantonsspezifische Entwicklungspfade? In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 119-153). Zürich: NZZ Libro.
- Eberli, Daniela (2018). Tracing the use of evaluations in legislative processes in Swiss cantonal parliaments. *Evaluation and Program Planning*, 69, 139-147.
- Eberli, Daniela (2019). *Die Nutzung von Evaluationen in den Schweizer Parlamenten*. Zürich: Seismo.
- Eberli, Daniela & Bundi, Pirmin (2017). Parlament und Evaluation: Guts Meets Brain. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 243-278). Zürich: NZZ Libro.
- Eberli, Daniela, Bundi, Pirmin, Frey, Kathrin & Widmer, Thomas (2014). *Befragung Parlamente und Evaluationen. Ergebnisbericht*. Zürich: Institut für Politikwissenschaft, Universität Zürich. Verfügbar unter: [https://www.syneval.ch/images/PDF/Ergebnisbericht\\_ParEval.pdf](https://www.syneval.ch/images/PDF/Ergebnisbericht_ParEval.pdf).
- Eberli, Daniela, Bütikofer, Sarah & Bundi, Pirmin (2018). La professionalizzazione dei parlamenti cantionali. In Andrea Pilotti & Oscar Mazzoleni (Eds.), *Milizia e professionismo nella politica svizzera* (pp. 75-102). Locarno: Armando Dadò editore.
- Eberli, Daniela, Bütikofer, Sarah & Bundi, Pirmin (2019). La professionalisation dans les parlements cantonaux. In Andrea Pilotti & Oscar Mazzoleni (Eds.), *Le système de milice et la professionnalisation politique en Suisse* (pp. 91-126). Neuchâtel: Livreo-Alphil.

- Frey, Kathrin (2010). Revising road safety policy: The role of systematic evidence in Switzerland. *Governance*, 23 (4), 667-690.
- Frey, Kathrin (2012). *Evidenzbasierte Politikformulierung in der Schweiz*. Baden-Baden: Nomos.
- Frey, Kathrin, Di Giorgi, Vanessa & Widmer, Thomas (2016). Die Evaluationstätigkeit in der Schweiz: Die Syneval-Datenbank. *LeGes – Gesetzgebung & Evaluation*, 27 (3), 505-522.
- Frey, Kathrin, Di Giorgi, Vanessa & Widmer, Thomas (2017). Die Evaluationstätigkeit in der Schweiz: ausgewählte Entwicklungen im Vergleich. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 67-88). Zürich: NZZ Libro.
- Frey, Kathrin & Ledermann, Simone (2010). Introduction: Evidence-based policy: A concept in geographical and substantive expansion. *German Policy Studies*, 6 (2), 1-15.
- Frey, Kathrin & Ledermann, Simone (2017). Nutzung von Evaluationen in Regierung und Verwaltung. In Fritz Sager, Fritz, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 211-242). Zürich: NZZ Libro.
- Frey, Kathrin & Widmer, Thomas (2011). Revising Swiss policies: The influence of efficiency analyses. *American Journal of Evaluation*, 32 (4), 494-517.
- Frey, Kathrin & Widmer, Thomas (2013). Vom Eigenen oder Fremden lernen bei der evidenzbasierten Politikgestaltung? In Jan Hense, Stefan Rädiker, Wolfgang Böttcher, & Thomas Widmer (Hrsg.), *Forschung über Evaluation* (S. 165-185). Münster: Waxmann.
- Hense, Jan & Widmer, Thomas (2013). Ein Überblick zum internationalen Stand der Forschung über Evaluation. In Jan Hense, Stefan Rädiker, Wolfgang Böttcher & Thomas Widmer (Hrsg.), *Forschung über Evaluation* (S. 251-278). Münster: Waxmann.
- Horber-Papazian, Katia & Baud-Lavigne, Marion (2019). Factors contributing to the strong institutionalization of policy evaluation in Switzerland. In Andreas Ladner, Nils Soguel, Yves Emery, Sophie Weerts & Stéphane Nahrath (Eds.), *Swiss public administration. Making the state work successfully* (pp. 355-370). Cham: Palgrave Macmillan.
- Horber-Papazian, Katia & Bützer, Michael (2008). Dissemination of evaluation reports in newspapers: The case of CEPP evaluations in Geneva, Switzerland. In Richard Boyle, Jonathan D. Breul & Peter Dahler-Larsen (Eds.), *Open to the public: Evaluation in the public arena* (pp. 43-66). New Brunswick: Transaction.
- Horber-Papazian, Katia & Jacot-Descombes, Caroline (2012). Is evaluation culture shaped by the Swiss political system and multiculturalism? In Jean-Claude Barbier & Penny Hawkins (Eds.), *Evaluation cultures. Sense-making in complex times* (pp. 89-104). New Brunswick, NJ: Transaction Press.
- Jacob, Steve (2015). La recherche sur l'évaluation en Suisse. In Katia Horber-Papazian (Ed.), *Regards croisés sur l'évaluation en Suisse* (pp. 267-284). Lausanne: Presses polytechniques et universitaires romandes.
- Jacob, Steve, Speer, Sandra & Furubo, Jan-Eric (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21 (1), 6-31.
- Janett, Daniel (2004). Die Evaluationsfunktion des Bundesparlaments – eine Zwischenbilanz. *LeGes – Gesetzgebung & Evaluation*, 15 (2), 137-150.
- Klöti, Ulrich & Schneider, Gerald (1989). *Die Informationsbeschaffung des Gesetzgebers*. Grösch: Rüegger.
- Knoepfel, Peter, Papadopoulos, Yannis, Sciarini, Pascal, Vatter, Adrian & Häusermann, Silja (Hrsg.) (2014). *Handbuch der Schweizer Politik*. Zürich: NZZ.
- Ledermann, Simone (2012). Exploring the necessary conditions for evaluation use in program change. *American Journal of Evaluation*, 33 (2), 159-178.
- Ledermann, Simone (2014). Evidenz und Expertise im vorparlamentarischen Gesetzgebungsprozess. *Swiss Political Science Review*, 20 (3), 453-485.
- Linder, Wolf & Müller, Sean (2017). *Schweizerische Demokratie*. 4. Auflage. Bern: Haupt.
- Mader, Luzius (2005). Artikel 170 der Bundesverfassung: Was wurde erreicht, was ist noch zu tun? *LeGes – Gesetzgebung & Evaluation*, 16 (1), 29-37.
- Mader, Luzius (2009). Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz. In Thomas Widmer, Wolfgang Beywl & Carlo Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 52-63). Wiesbaden: VS Verlag.



- Neidhart, Leonhard (1970). *Plebizit und pluralitäre Demokratie. Eine Analyse der Funktion des schweizerischen Gesetzesreferendums*. Bern: Francke.
- Patton, Michael Quinn (1998). Discovering Process Use. *Evaluation*, 4 (2), 225-233.
- Pilotti, Andrea, Sciarini, Pascal, Varone, Frédéric & Cappelletti, Fabio (2018). L'Assemblea federale: un parlamento di milizia in fase di professionalizzazione. In Andrea Pilotti, & Oscar Mazzoleni (Eds.), *Milizia e professionismo nella politica svizzera* (pp. 45-73). Locarno: Armando Dadò editore.
- Pilotti, Andrea, Sciarini, Pascal, Varone, Frédéric & Cappelletti, Fabio (2019). L'Assemblée fédérale: un parlement de milice en voie de professionnalisation. In Andrea Pilotti & Oscar Mazzoleni (Eds.), *Le système de milice et la professionnalisation politique en Suisse* (pp. 53-89). Neuchâtel: Livreo-Alphil.
- Rissi, Christof & Sager, Fritz (2013). Types of knowledge utilization of Regulatory Impact Assessments: Evidence from Swiss policymaking. *Regulation & Governance*, 7 (3), 348-364.
- Sager, Fritz, Künzler, Johanna & Lutz, Philipp (2017). Das politische System der Schweiz und seine Kontaktpunkte zur Evaluation. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 23-48). Zürich: NZZ Libro.
- Sager, Fritz, Widmer, Thomas & Balthasar, Andreas (2017a). Schlussfolgerungen: Evaluation als Teil des politischen Systems der Schweiz? In Fritz Sager, Thomas Widmer, & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S.313-330). Zürich: NZZ Libro.
- Sager, Fritz, Widmer, Thomas & Balthasar, Andreas (Hrsg.) (2017b). *Politikevaluation im politischen System der Schweiz*. Zürich: NZZ Libro.
- Sager, Fritz & Zollinger, Christine (2011). The Swiss political system in comparative perspective. In: Christine Trampusch & André Mach (Eds.), *Switzerland in Europe* (pp. 27-42). London: Routledge.
- Schlauffer, Caroline (2016). Global evidence in local debates: the Programme for International Student Assessment (PISA) in Swiss direct-democratic debates on school policy. *Policy & Politics*, 44 (4), 547-561.
- Schlauffer, Caroline (2018a). The contribution of evaluations to the discourse quality of newspaper content. *Evaluation and Program Planning*, 69, 157-165.
- Schlauffer, Caroline (2018b). The narrative uses of evidence. *Policy Studies Journal*, 46 (1), 90-118.
- Schlauffer, Caroline & Stucki, Iris (2017). Die Nutzung wissenschaftlicher Evidenz in Abstimmungserläuterungen. *LeGes – Gesetzgebung & Evaluation*, 28 (1), 69-80.
- Schlauffer, Caroline, Stucki, Iris & Sager, Fritz (2018). The political use of evidence and its contribution to democratic discourse. *Public Administration Review*, 78 (4), 645-649.
- Spinatsch, Marcus (2008). Using evaluative knowledge for policy design: A case study of the Swiss addiction policy. In Richard Boyle, Jonathan D. Breul & Peter Dahler-Larsen (Eds.), *Open to the public: Evaluation in the public arena* (pp. 121-133). New Brunswick: Transaction.
- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (Eds.) (2020). *The institutionalisation of evaluation in Europe*. Cham: Palgrave Macmillan.
- Stucki, Iris (2016). The use of evidence in public debates in the media: the case of Swiss direct-democratic campaigns in the health policy sector. *Evidence & Policy*, 12 (4), 487-504.
- Stucki, Iris (2017). Arguing about smoking bans: The role of evidence in the social construction of conflicting policy ideas. *Critical Policy Studies*, 11 (4), 411-432.
- Stucki, Iris (2018). Evidence-based arguments in direct democracy: The case of smoking bans in Switzerland. *Evaluation and Program Planning*, 69, 148-156.
- Stucki, Iris, Pleger, Lyn E. & Sager, Fritz (2018). The making of the informed voter: A split-ballot survey on the use of scientific evidence in direct-democratic campaigns. *Swiss Political Science Review*, 24 (2), 115-139.
- Stucki, Iris & Sager, Fritz (2018). Aristotelian framing: logos, ethos, pathos and the use of evidence in policy frames. *Policy Sciences*, 51 (3), 373-385.
- Stucki, Iris & Schlauffer, Caroline (2017). Die Bedeutung von Evaluationen im direktdemokratischen Diskurs. In Fritz Sager, Thomas Widmer & Andreas Balthasar, Andreas (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 279-310). Zürich: NZZ Libro.
- Varone, Frédéric, Bundi, Pirmin & Gava, Roy (2020). Policy evaluation in parliament: Interest groups as catalysts. *International Review of Administrative Sciences*, 86 (1), 98-114.

- Widmer, Thomas (1996). Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Bern: Haupt.
- Widmer, Thomas (2002). Staatsreformen und Evaluation: Konzeptionelle Grundlagen und Praxis bei den Schweizer Kantonen. *Zeitschrift für Evaluation*, 1 (1), 101-114.
- Widmer, Thomas (2007). Rahmenbedingungen und Praxis der Evaluation im schweizerischen Bundesstaat. *ÖHW – Das öffentliche Haushaltswesen in Österreich*, 48 (1-2), 69-93.
- Widmer, Thomas (2008a). Evaluationsansätze und ihre Effekte. In Hildegard Matthies, & Dagmar Simon (Eds.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. Leviathan Sonderheft*, 24, 267-287.
- Widmer, Thomas (2008b). Evaluation in der Aussenpolitik: Gründe für eine Evaluationslücke. *Zeitschrift für Internationale Beziehungen*, 15 (1). 125-137.
- Widmer, Thomas (2009). The contribution of evidence-based policy to the output-oriented legitimacy of the state. *Evidence & Policy*, 5 (4), 351-372.
- Widmer, Thomas (2015). Développements et perspectives entre profession et discipline. Evaluation quo vadis? In: Katia Horber-Papazian (Ed.), *Regards croisés sur l'évaluation en Suisse* (pp. 285-307). Lausanne: Presses polytechniques et universitaires romandes.
- Widmer, Thomas (2016). Wandel von Staatlichkeit in der Schweiz in der zweiten Hälfte des 20. Jahrhunderts. In Lucien Criblez, Christina Rothen & Thomas Ruoss (Hrsg.), *Staatlichkeit in der Schweiz* (S. 403-418). Zürich: Chronos.
- Widmer, Thomas (2017). Geschichte der Evaluation im schweizerischen politischen System. In Fritz Sager, Thomas Widmer & Andreas Balthasar, Andreas (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 51-66). Zürich: NZZ Libro.
- Widmer, Thomas & Beywl, Wolfgang (2009). Divergente Entwicklungspfade der Evaluation. In Thomas Widmer Wolfgang Beywl & Carlo Fabian (Hrsg.), *Evaluation: Ein systematisches Handbuch* (S. 509-527). Wiesbaden: VS Verlag.
- Widmer, Thomas & De Rocchi, Thomas (2012). *Evaluation. Grundlagen, Ansätze und Anwendungen*. Zürich: Rüegger.
- Widmer, Thomas, Eberli, Daniela, Ackermann, Günter, Balzer, Lars, Bundi, Pirmin, Hirschi, Christian, Kaspar, Tanya, Neuenschwander, Peter, Neuhaus, Björn, Rohrbach, Walter, Schlauffer, Caroline & Strebel, Felix (2016). Forschung über Evaluation in der Schweiz: Stand und Aussichten. *LeGes – Gesetzgebung & Evaluation*, 27 (3), 459-483.
- Widmer, Thomas & Neuenschwander, Peter (2004). Embedding evaluation in the Swiss federal administration. *Evaluation*, 10 (4), 388-409.
- Widmer, Thomas, Rüegg, Erwin & Neuenschwander, Peter (2001). *Stand und Aussichten der Evaluation beim Bund (EvalBund)*. Zürich: Institut für Politikwissenschaft.
- Wirths, Damien (2016). Les clauses d'évaluation sectorielles: outils d'ajustement ou de stratégie politique? Le cas de l'interdiction de la mendicité à Lausanne. *Gestion et management public*, 4 (4), 53-74.
- Wirths, Damien (2017). Procedural institutionalization of the evaluation through legal basis: A new typology of evaluation clauses in Switzerland. *Statute Law Review*, 38 (1), 23-39.
- Wirths, Damien & Horber-Papazian, Katia (2016). Les clauses d'évaluation dans le droit des cantons suisses: leur diffusion, leur contenu et la justification à l'origine de leur adoption. *LeGes – Législation & Evaluation*, 28 (3), 485-502.
- Wirths, Damien, Rosser, Christian, Horber-Papazian, Katia & Mader, Luzius (2017). Über die gesetzliche Verankerung von Evaluation: die Verteilung von Evaluationsklauseln und deren Auswirkungen auf kantonaler Ebene. In Fritz Sager, Thomas Widmer & Andreas Balthasar (Hrsg.), *Politikevaluation im politischen System der Schweiz* (S. 155-187). Zürich: NZZ Libro.

### *Anschrift des Autors:*

Prof. Dr. Thomas Widmer, Institut für Politikwissenschaft, Universität Zürich, Affolternstrasse 56, CH-8050 Zürich, E-Mail: thow@ipz.uzh.ch.

Jörg Faust

# Rigorese Wirkungsevaluierung – Genese, Debatte und Nutzung in der Entwicklungszusammenarbeit

## Zusammenfassung

Der Aufstieg rigoroser Wirkungsevaluierungen (RIE) in der internationalen Entwicklungszusammenarbeit (EZ) geht einher mit einer anhaltenden Kontroverse um experimentelle und quasi-experimentelle Methoden. Vor diesem Hintergrund und auf Basis eines empirisch-analytischen Wissenschaftsverständnis leistet der vorliegende Aufsatz Beiträge zur Erklärung der Genese von RIE im Politikfeld, zu einer konstruktiven Auseinandersetzung mit den Vorteilen und Begrenzungen von RIE aus evaluatorischer Perspektive sowie zur Analyse ihrer Anwendung und Nutzung in der Praxis der internationalen EZ. Im Ergebnis erfolgt ein Plädoyer für eine breite Anwendung und Nutzung von RIE in der EZ, ohne diese aufgrund ihrer Begrenzungen jedoch zum Königsweg der Evaluierung oder gar der empirischen Sozialforschung zu erheben. Vielmehr sollte die Anwendung gegenstandsangemessen, theoriebasiert und um qualitative Methoden ergänzt erfolgen und andere Evaluationsverfahren dort eingesetzt werden, wo RIE nicht geeignet sind.

**Schlagworte:** Evaluation, Entwicklungszusammenarbeit, Wirkungsevaluierung, Experimente, Methoden

## Abstract

*Rigorous Impact Evaluation in Development Cooperation: Origins, Debates and Use*

The rise of rigorous impact evaluation (RIE) in international development cooperation has been accompanied by a controversy around experimental and quasi-experimental approaches. Against this backdrop and from an empirical-analytical perspective, this paper contributes to the ongoing debate in three ways: first, it explains the emergence of RIE in foreign aid from the particularities of the aid effectiveness debate; second, it contributes to an enlightened debate by sketching the advantages and limitations of RIE from an evaluation perspective. Finally, it sketches the current use of RIE in the practical realm of development cooperation. Overall, the paper argues in favour of a broader and more systematic use of RIE in development cooperation without disregarding the methodological, empirical and practical limitations of the approach.

**Keywords:** Impact Evaluation, Development Cooperation, Randomized Control Trials, Methods

## 1 Einleitung<sup>1</sup>

Mit der Vergabe des Alfred-Nobel-Gedächtnispreises für Wirtschaftswissenschaften 2019 an Esther Duflo, Abhijit Banerjee und Michael Kremer hat die Bedeutung von Feldexperimenten und quasi-experimentellen Designs in der Entwicklungsforschung

bzw. den damit befassten Sozialwissenschaften weiter zugenommen. Von der in den letzten Jahren stark angestiegenen Anzahl der Vertreter\*innen experimenteller Feldforschung wird diese nicht selten als empirischer Königsweg der Entwicklungsforschung aufgefasst und deren weitere Verbreitung eingefordert (Banerjee & Duflo, 2011). Gleichzeitig formierte sich insbesondere innerhalb der qualitativ arbeitenden Sozialforschung Widerstand gegen den vermeintlichen Goldstandard und seinen quantitativ-erklärenden Zuschnitt zu Lasten einer eher qualitativ-verstehenden Wissenschaftstradition. Doch auch quantitativ arbeitende, einem empirisch-analytischen Wissenschaftsverständnis verpflichtete Ökonom\*innen wandten sich gegen eine unkritische Verbreitung der experimentellen Forschung, so etwa der Nobelpreisträger Angus Deaton (u. a. Deaton, 2010; Pritchett & Sandefur, 2015).

Ihren Niederschlag in der angewandten Forschung fand diese Debatte insbesondere in der Evaluierung internationaler entwicklungspolitischer Maßnahmen. Auch hier wurde seit den 2000er Jahren eine Zunahme rigoroser Wirkungsevaluierungen eingefordert und auch – wenngleich keinesfalls systematisch – durchgeführt (u. a. Centre for Global Development, 2006). Als rigorose Wirkungsevaluierungen (*rigorous impact evaluation* - RIE) werden im Folgenden Evaluierungen begriffen, die auf die Identifikation der Wirkungen einer Maßnahme auf Zielgruppenebene fokussiert sind und sich hierbei experimenteller *oder* quasi-experimenteller Methoden bedienen. Dabei steht die kausale Zuordnung möglicher Veränderungen bei der Zielgruppe (z. B. Individuen, Haushalte, Schulen, Wirtschaftsakteure) zu einer oder mehrerer zuvor erfolgten Maßnahme im Mittelpunkt.

Im Kern besteht der methodische Ansatz darin, einen Vergleich einer Interventionsgruppe mit einer möglichst ähnlichen Kontroll- oder Vergleichsgruppe durchzuführen. Ein solcher Vergleich ermöglicht „dann eine empirisch fundierte Aussage darüber, wie sich die Zielgruppe einer Intervention sowohl mit als auch ohne Durchführung der entwicklungspolitischen Maßnahme entwickelt hätte“ (Bruder, Faust & Krämer 2019, S. 1).<sup>2</sup> RIE können sowohl am Ende eines Programms als summative Evaluierung als auch im Sinne einer wirkungsorientierten Begleitforschung oder formativen Evaluierung bereits im Verlauf der Umsetzung von Maßnahmen eingesetzt werden (u. a. Pritchett & Sandefur, 2015; Funk, Groß, Leininger & Schiller, 2019).

Rigorose Wirkungsevaluierungen sind gemäß ihrer Befürworter\*innen von hohem Nutzen, können sie doch bei richtiger Anwendung mit einer im Vergleich zu anderen Methoden deutlich höheren Sicherheit die Wirkungen von entwicklungspolitischen Maßnahmen bzw. deren Ausbleiben identifizieren. Entsprechend groß könne ihr Beitrag für eine evidenzbasierte Politikgestaltung in einem Politikfeld sein, das immer noch einer intensiven und nicht selten ideologischen Debatte zwischen Verfechtern externer Hilfe und deren Gegnern ausgesetzt ist. Eine auf RIE basierende Evidenzagenda solle dabei weit über einzelne RIE hinausreichen (White, 2019). Eine solche Evidenzagenda enthält auch die regelgebundene Aggregation von RIE in systematischen Reviews, um die externe Validität zu erhöhen und hierdurch projektübergreifendes Lernen zu ermöglichen. Ebenso sind Evidenzkarten Teil dieser Agenda, die mittels einer systematisch erfolgenden Sammlung und Typologisierung von RIE nach thematischen Fragestellungen das bereits vorhandene Wissen sowie Evidenzlücken gleichsam kartographieren. Schließlich soll die Nutzbarkeit existierender Erkenntnisse über die Einrichtung von Evidenz-Portalen verbessert werden, die das vorhandene Wissen der globalen Gemeinschaft einfach und zweckdienlich zur Verfügung stellen (White, 2019).

Demgegenüber positionierten sich die Kritiker\*innen von RIE mit einer Reihe von Argumenten, welche die Anwendung von RIE zumindest relativieren. Zu diesen Argumenten zählen unter anderem die begrenzte thematische Reichweite bzw. Anwendbarkeit von RIE. Zudem sei die externe Validität von RIE sehr gering. Auch basierten viele RIE auf einer mangelnden theoretischen Fundierung, vernachlässigten die Analyse von Kausalmechanismen, seien unverhältnismäßig teuer und aus ethischen Gründen abzulehnen. Schließlich seien RIE politisch neoliberal aufgeladen und es bestünde oftmals eine Verzerrung zu Gunsten der Kontrollfunktion von Evaluierung, was deren Nützlichkeit für das Lernen in der Praxis der Entwicklungszusammenarbeit (EZ) begrenze.<sup>3</sup>

Angesichts des skizzierten Aufstiegs rigoroser Wirkungsevaluierungen in der EZ und der anhaltend kontroversen Debatte, leistet der vorliegende Aufsatz auf Basis eines empirisch-analytischen Wissenschaftsverständnis Beiträge 1) zur Erklärung der Genese von RIE im Politikfeld, 2) zu einer konstruktiven Auseinandersetzung mit den Vorteilen und Begrenzungen von RIE aus evaluatorischer Perspektive sowie 3) zur Analyse deren Anwendung und Nutzung in der Praxis der internationalen EZ.

Der Beitrag gliedert sich wie folgt: Nach der Einleitung wird der Aufstieg rigoroser Wirkungsevaluierungen aus der Kombination eines funktional abgeleiteten Bedarfs an Evaluierung und drei Katalysatoren einer intensiv geführten Wirksamkeitsdebatte im Politikfeld (Kapitel 2) begründet. In der Auseinandersetzung mit den Vorteilen und Begrenzungen plädiert dieser Beitrag zweitens für eine breitere Anwendung und Nutzung von RIE im Politikfeld und eine Stärkung der Evidenzagenda, ohne diese jedoch zum Goldstandard zu erheben. Vielmehr sollte die Anwendung von RIE gegenstandsangemessen, theoriebasiert und um qualitative Methoden ergänzt erfolgen. Dies öffnet den Raum für andere Evaluationsverfahren in einer Vielzahl von Anwendungsbereichen, in denen RIE nicht durchgeführt werden können bzw. nicht zweckdienlich für ein praxisgetriebenes Erkenntnisinteresse sind (Kapitel 3). Drittens schließlich verweist der Beitrag auf die Heterogenität bzw. Uneinheitlichkeit und Ungleichzeitigkeit der Anwendung, Verbreitung und Nutzung von RIE in der internationalen EZ und problematisiert die Fragmentierung der Gebergemeinschaft bzw. deren Wissensmanagement in diesem Feld (Kapitel 4).

## 2 Zur Genese rigoroser Wirkungsevaluierung in der EZ

Die Diskussion um Relevanz, Angemessenheit und praktische Verankerung rigoroser Wirkungsevaluierung in der EZ ist eng mit der breiter geführten Wirksamkeitsdebatte im Politikfeld verbunden. Um diese Verknüpfung mit der hier im Zentrum stehenden Evaluationsperspektive angemessen herauszuarbeiten, soll zunächst aus funktionaler Sicht der strukturell hohe Bedarf an Evaluation in der EZ begründet werden. Daraufhin wird argumentiert, dass dieser Bedarf an Evaluation durch drei Faktoren aus der breiteren Wirksamkeitsdebatte verstärkt wurde. Das Zusammenspiel des strukturellen Evaluationsbedarfs und der drei Katalysatoren haben im Politikfeld die Forderung nach bzw. Durchführung von rigorosen Wirkungsevaluierungen begünstigt, die in den letzten Jahren zentraler Bestandteil einer breiteren Agenda zur Nutzung rigoroser Evidenz im Politikfeld geworden sind.

## 2.1 Die strukturell hohe Nachfrage nach Evaluation in der EZ

Die Analyse der potenziellen Wirkungen in der EZ und deren direkte Anbindung an die Durchführungspraxis und politische Entscheidungsfindung mittels des Instrumentariums der Evaluierung hat eine lange Tradition. In nur wenigen anderen Politikfeldern wurde in den letzten vier Jahrzehnten so umfassend evaluiert wie in der EZ. Vielfach wurden hierbei auch Strukturen, Organisationen und Prozesse in den Blick genommen, so dass Evaluation nicht nur die Wirkungsanalyse zum Gegenstand hat. Neben der schier unerschöpflichen Anzahl von Evaluierungen ist zudem die Institutionalisierung der Evaluation bemerkenswert. Diese schlägt sich erstens in Evaluationsreferaten, Stabsstellen, Kommissionen oder gar eigenständigen Instituten auf bilateraler Ebene wie auch in multilateralen Organisationen nieder. Zweitens ist die Institutionalisierung der Evaluation durch ein hohes Maß an (internationaler) Standardsetzung und organisationsspezifischer Regulationsdichte gekennzeichnet, was etwa Prinzipien, Kriterien und Durchführungspraktiken der Evaluierung anbelangt (im Überblick OECD, 2016; für Deutschland siehe u. a. Meyer, Bär, Faust, Jan, Silvestrini & Wein, 2019).

Dieser ausgeprägte Institutionalisierungsgrad und die hohe Evaluationsdichte in der EZ lassen sich aus einer funktionalen Perspektive begründen. Als spezifische Form angewandter Forschung soll Evaluierung drei miteinander verschränkte Funktionen erfüllen (Stockmann, 2004, S. 3). Mittels der Erkenntnisfunktion ist Evaluation auf die Schaffung praxisrelevanten, nützlichen Wissens ausgerichtet. Das generierte Wissen soll sodann über die Lernfunktion von den jeweils Verantwortlichen zur Verbesserung von Projekten, Programmen und Strategien eingesetzt werden. Gleichzeitig hat Evaluierung eine wichtige Kontrollfunktion, da ihre Befunde der Rechenschaftslegung über die Verwendung öffentlicher Mittel dienen. In Kombination können Erkenntnis-, Lern-, und Kontrollfunktion schließlich zur Legitimität eines Politikfeldes beitragen, wenn sie über die Bereitstellung der ermittelten Erkenntnisse für Lernen und Rechenschaftslegung sowohl Output- wie auch Inputlegitimität erhöhen.

Erkenntnis-, Lern- und Kontrollfunktion begründen den strukturell hohen Evaluierungsbedarf in der EZ. Die Erkenntnisfunktion von Evaluierung trifft auf ein hohes Maß an Unkenntnis bzw. Unsicherheit als einem Strukturmerkmal des Politikfeldes. Maßnahmen der EZ können als externe Interventionen in „fernen“ Gesellschaften begriffen werden. Diese Interventionen bzw. die sie durchführenden oder veranlassenden Akteure der EZ sind gemeinhin mit einem hohen Maß an Unsicherheit konfrontiert. Ob Brunnenbau, humanitäre Hilfe, Fortbildungen im Gesundheitssektor, Reformberatung auf Regierungsebene oder Korruptionsbekämpfung: Die Kenntnisse über die vor Ort existierenden Strukturen und Prozesse, auf die die Entwicklungsmaßnahmen treffen, sind vergleichsweise gering und zudem sind diese Strukturen und Prozesse oftmals durch ein hohes Maß an Instabilität gekennzeichnet. Entsprechend hoch ist der Bedarf an verlässlicher Evidenz über die Auswirkungen entwicklungspolitischer Maßnahmen und die diesen Wirkungen zugrundeliegenden kausalen Mechanismen.

Zudem sind Entwicklungsmaßnahmen gemeinhin keine einmaligen Vorgänge. Maßnahmen werden oft über etliche Jahre in mehreren Phasen fortgesetzt und – zunehmend kontextspezifisch angepasst – an anderen Orten repliziert. Daher gilt es, Erfahrungen und Erkenntnisse aus der Zusammenarbeit in einem Land oder in einer Region aufzubereiten, um diese in die Implementierung neuer Maßnahmen in anderen Kontexten einfließen zu lassen. Die umfassende geografische Reichweite der EZ und

die Notwendigkeit erfahrungsbasierter Anpassungen über Zeit, Sektoren und räumliche Grenzen hinweg machen Lernen und Wissensmanagement daher zu einem besonders wichtigen Erfolgsfaktor im Politikfeld. Entsprechend groß ist das Potenzial, Evaluation in der EZ als Lerninstrument in der eigenen Organisation, aber auch darüber hinaus organisationsübergreifend einzusetzen.

Schließlich ist die EZ durch ein strukturelles Kontrollproblem gekennzeichnet. Hierbei treten zwischen den unterschiedlichen Akteuren der entwicklungspolitischen „Wertschöpfungskette“ ausgeprägte Prinzipal-Agenten-Probleme sowie eine fehlende Feedback-Schleife zwischen den Financiers der Zusammenarbeit und deren Endbegünstigten auf (u. a. Martens, Mummert & Murrell, 2002). Die Steuerzahlenden, aber oft auch deren parlamentarische Vertreter\*innen verfügen gemeinhin über deutlich weniger Informationen und Kenntnisse über das breite Einsatzspektrum und die potenziell mannigfaltigen Auswirkungen von Entwicklungshilfe als über vergleichbare innenpolitische Sektorpolitiken wie Bildungs- und Gesundheitspolitik oder innere Sicherheit.<sup>4</sup> Gleichzeitig sind die eigentlichen Zielgruppen der EZ - marginalisierte und/oder bedürftige Bevölkerungsgruppen in Entwicklungsländern - kaum dazu in der Lage, ihre Erfahrungen systematisch zu bündeln und an die Steuerzahlenden oder deren Vertreter\*innen rückzukoppeln. Zwischen den Steuerzahlenden und den Zielgruppen sind jedoch eine Vielzahl von besser organisierten und informierten Akteuren positioniert, die nicht ausschließlich die Verwirklichung originär entwicklungspolitischer Ziele verfolgen: Nationale Geberregierungen mit außenpolitischen und außenwirtschaftlichen Interessen; bilaterale und multilaterale Durchführungsorganisationen sowie auch zivilgesellschaftliche Organisationen mit bürokratischen Sonderinteressen und schließlich Staatsorgane in Empfängerländern, die vielfach nicht sonderlich gemeinwohlorientiert handeln (Martens, Mummert & Murrell, 2002; Easterly, 2002; Faust & Michaelowa, 2013). Insgesamt ist die EZ somit ein Politikfeld, in dem ein hoher Bedarf für unabhängige Evaluierung zur Stärkung von Transparenz und Rechenschaftslegung existiert.

## 2.2 Katalysatoren der Wirksamkeitsdebatte

Der zuvor funktionalistisch begründete Bedarf für ein überproportional hohes Maß an Evaluierung in der EZ ist in den letzten beiden Dekaden im Zuge einer kritischen Auseinandersetzung um die Wirkungen der EZ durch drei Faktoren verstärkt worden. Alle drei Faktoren haben hierbei gleichsam als Katalysatoren für die zunehmende Bedeutung von RIE fungiert.

*Ordnungspolitische Kritik:* Ein erster Katalysator ist die ordnungspolitisch begründete Skepsis gegenüber dem Politikfeld bzw. den hierin wirkenden Akteuren aus Staat und nationalen, internationalen und zivilgesellschaftlichen Entwicklungsorganisationen. Bereits in den frühen 1960er Jahren wurde „Entwicklungshilfe“ als ein außenpolitisches Instrument erachtet, das – mit Ausnahme humanitärer Hilfe – in den Dienst außenpolitischer Interessen der Geber gestellt werden solle (Morgenthau, 1962). Seit den 1990er Jahren ist die Zahl an Allokationsanalysen gestiegen, die empirisch belegen, dass die Allokation von Entwicklungsgeldern auch ökonomischen und sicherheitspolitischen Motiven der Geberstaaten bzw. multilateralen Organisationen folgt (u. a. Dreher, Sturm & Vreeland, 2009; mit Blick auf die deutsche EZ u. a. Faust & Ziaja, 2012). Unabhängig davon, ob diese Berücksichtigung außenwirtschaftlicher oder sicherheits-

politischer Motive als legitim erachtet wird, kann deren Berücksichtigung negative Auswirkungen auf die Wirksamkeit entwicklungspolitischer Zielsetzungen haben. Denn die Ressourcen werden dann nicht mehr mit der gleichen Konsequenz in diejenigen Länder bzw. Projekte und Programme fließen, deren Aussicht auf entwicklungspolitische Wirksamkeit am höchsten oder zumindest aus entwicklungspolitischen Gründen angemessen ist (Dreher, Fuchs, Lang & Langlotz, 2018).

Neben der der Wirksamkeit abträglichen Verquickung von außenpolitischen mit entwicklungspolitischen Motiven bei der Vergabe, fokussiert ein weiterer ordnungspolitischer Kritikpunkt auf die Planungseuphorie einer über Jahrzehnte angewachsenen und gleichzeitig fragmentierten Entwicklungsbürokratie auf nationaler wie multilateraler Ebene (Easterly, 2002; 2006). Die Kritik lautet dabei, dass fehlerhafte Anreizstrukturen innerhalb dieser Organisationen den Fokus auf die Verausgabung von Mitteln, Projektproliferation und Abgrenzung gegenüber anderen Organisationen begünstige. Hingegen fehle eine marktkonforme Wettbewerbsstruktur, die wirksame Maßnahmen belohne (Easterly, 2002). Dadurch seien Mittelbeschaffung und Mittelverausgabung zentrale Erfolgsgrößen der Entwicklungsorganisationen geworden, nicht jedoch die Wirksamkeit der implementierten Maßnahmen; ein erst allmählich im Wandel begriffenes Strukturmerkmal der internationalen EZ.

*Miko-Makro-Paradoxon:* Komplementär zu der oben genannten ordnungspolitischen Kritik besteht der zweite Katalysator für die zunehmende Bedeutung von RIE im sogenannten Mikro-Makro-Paradoxon (u. a. Faust & Leiderer, 2008). Dieses besteht darin, dass auf der Makro-Ebene des statistischen Ländervergleichs mit Aggregatdaten starke Indizien für eine allenfalls eingeschränkte Wirksamkeit der EZ bestehen. Zugleich vermelden Entwicklungsorganisationen auf der Mikro-Ebene konkreter Evaluationen jedoch beachtliche Erfolgsquoten.<sup>5</sup>

So ist auf der Ebene der statistischen Aggregatdatenforschung nach wie vor umstritten, ob EZ Wachstum fördere und Armut vermindere. Gleichzeitig liegen auf dieser Ebene eine Vielzahl an Befunden vor, die nahelegen, dass die Vergabe von EZ-Ressourcen auch politisch und ökonomisch motiviert ist und dass Entwicklungsressourcen zumindest in autoritär regierten Ländern Patronage und autoritäre Strukturen stärken (Hodler & Raschky, 2014; Dutta, Leeson & Williamson, 2013). Im Vergleich zu diesen kritischen Befunden sind Ergebnisse aus Projektevaluierungen deutlich positiver; Erfolgsquoten von über 80% sind die Regel. Aufgelöst wird dieses Mikro-Makro-Paradoxon einerseits durch die Kritik an den Projektevaluationen. Diese seien nicht in hinreichendem Maße unabhängig, in der methodischen und empirischen Qualität unzureichend und zudem meist nicht auf die Identifikation potenzieller, nicht erwünschter negativer Effekte ausgerichtet (Center for Global Development, 2006). Andererseits mehrte sich auch die Kritik an der makro-quantitativen Forschung. Deren Ergebnisse seien geplagt durch oftmals schlechte Datenqualität, notorische Endogenitätsprobleme und ein für praktische Handlungsempfehlungen oftmals zu hohes Abstraktionsniveau.<sup>6</sup> Auch aus dieser Perspektive bieten sich unabhängige und methodisch verlässlichere Wirkungsevaluierungen auf Projekt- und Programmebene als ein relevanter Beitrag zur Auflösung des Paradoxons an.

*Weltinnenpolitik:* Der dritte Katalysator für die zunehmende Bedeutung rigoroser Wirkungsevaluierung lässt sich mit der Metapher der „Weltinnenpolitik“ umschreiben, die auf die ansteigenden Interdependenzen zwischen den Staaten und der Verquickung von



Außenbeziehungen und Innenpolitik verweist. In der Entwicklungspolitik manifestieren sich diese Interdependenzen in der wachsenden innergesellschaftlichen Relevanz von Entwicklungsherausforderungen der Entwicklungs- und Schwellenländer in Geberländern. Konflikt- und ökologisch induzierte Migrationsströme, Klimawandel, Terrorismus, ökologische Ressourcenknappheit: Drängende Herausforderungen in Entwicklungs- und Schwellenländern betreffen heute viel unmittelbarer die Gesellschaften der Geber von EZ. Damit einher geht auch ein breiteres Interesse an deren Wirksamkeit zumal die innenpolitischen Auseinandersetzungen mit dem Erstarken des Rechtspopulismus an Intensität gewonnen haben. Kurzum: Die Wirksamkeit der EZ ist nicht mehr nur bürokratisches und akademisches Interessengebiet in einem normativ aufgeladenen, aber politisch etwas randständigen Politikfeld; sie wird angesichts globaler Entwicklungsherausforderungen wie Klimawandel, Gesundheitsschutz, Biodiversität und Gewaltkonflikten zunehmend Gegenstand einer auch innenpolitisch motivierten Debatte.

### 3 Rigorese Wirkungsevaluierung als Instrument der Wirkungsmessung

Die Forderungen nach anspruchsvollen Evaluierungen zur Wirksamkeit der EZ können aus funktionaler Perspektive wie auch aus der Wirksamkeitsdebatte gut abgeleitet werden. Doch stellt sich aus evaluatorischer Perspektive die Frage, inwieweit RIE mit ihren experimentellen und quasi-experimentellen Verfahren der hierfür geeignete Evaluierungstypus sind.

Können Eltern durch die Vergabe konditionierter Haushaltszuschüsse dazu motiviert werden, ihre Töchter in die Schule zu schicken und sollten diese Haushaltszuschüsse eher der Mutter oder dem Vater gegeben werden? Sollte die chronisch anzutreffende Misswirtschaft beim Straßenbau in ruralen Gebieten eher durch lokale Partizipation der Bevölkerung reduziert werden oder stellt eine strenge Kontrolle durch den Rechnungshof eine bessere Alternative dar? Verhindert die Vergabe von Kurzzeitzjobs an Bewohner\*innen von Flüchtlingslagern das Potenzial von Gewaltkonflikten? Führt die Beratung kommunaler Verwaltungen bei der Raumplanung zu verbessertem Katastrophenschutz und verbesserter Anpassung an den Klimawandel? Hat die Wahlbeobachtung durch internationale Akteure einen Einfluss auf das Wahlergebnis in semidemokratischen Regimen und wird die Opposition hierdurch gestärkt?

All dies sind Fragen nach den Wirkungen von Entwicklungsmaßnahmen, für die es unterschiedliche aber gut begründete Ansichten geben kann. Einen vergleichsweise eindeutigen Befund darüber, ob und wie stark eine (durchschnittliche) Wirkung einer Intervention auf die Zielgruppe eingetreten ist, können zumindest für den Einzelfall rigorese Wirkungsevaluierungen geben. Wenn etwa – wie in der internationalen Wahlbeobachtung üblich – Wahlbeobachter ihre zu beobachtenden Wahldistrikte via Zufallsstichprobe zugeteilt bekommen und ebenso zufallsbasiert eine Kontrollgruppe ohne Wahlbeobachter gezogen wird, dann kann der Einfluss der Beobachter auf das offizielle Wahlergebnis in den Distrikten ermittelt werden (Hyde, 2007). Ähnliches kann mit Blick auf lokale Infrastrukturvorhaben geschehen. Gemäß einer Zufallsziehung können unterschiedliche Gruppen mit Dörfern gebildet werden, in denen lokale Stra-

ßenbauprojekte geplant sind: Solche, in der parallel zur Maßnahme lokale Partizipationsmechanismen der sozialen Kontrolle eingeführt werden, solche, in der eine Untersuchung durch den Rechnungshof angekündigt wird, und eine Kontrollgruppe, also eine Gruppe von Dörfern, in denen keine der Maßnahmen eingeführt werden. Sind die Gruppengrößen hinreichend groß, so kann der durchschnittliche Qualitätsunterschied der gebauten Straßen, deren Materialgüte und deren Kosten auf die unterschiedlichen Interventionen zurückgeführt werden (Olken, 2007).

Auch quasi-experimentelle Designs können bei überzeugenden Identifikationsstrategien einen ähnlich verlässlichen Effekt einer Intervention wie randomisierte Studien identifizieren, wenn die Randomisierung etwa aus administrativ-politischen, zeitlichen, ethischen oder finanziellen Gründen nicht möglich ist. So können in einer Vielzahl von Fällen etwa Matching-Verfahren eingesetzt werden, bei denen nach entsprechenden Erhebungen Interventions- und Kontrollgruppe verglichen und mittels statistischer Verfahren, verzerrende Störgrößen eliminiert werden: Sei es bei der Untersuchung von Reformberatung auf lokaler Ebene (Leppert, Hohfeld, Lech & Wencker, 2018), der Untersuchung von konditionierten Haushaltszuschüssen an bedürftige Familien (Rawlings & Rubio, 2005; Molina, Millán, Barham, Macours, Maluccio & Stampini, 2019) oder auch der Entsendung junger Freiwilliger in Entwicklungsländer (Polak, Guffler & Scheinert, 2018).

Der entscheidende Vorteil experimenteller und quasi-experimenteller Verfahren ist dabei, dass sie den Einfluss typischer Störgrößen eliminieren. Der einfache Vorher-Nachher-Vergleich kann den potenziellen Effekt einer Maßnahme in den vielfach besonders dynamischen Kontexten der EZ kaum identifizieren, wenn auf Zielindikatoren viele inkonstante Drittvariablen wirken. Ebenso schwerwiegend sind gerade in der EZ oftmals anzutreffende Selektionsverzerrungen. Ist es tatsächlich die Trainingsmaßnahme, die Erfolge bei der Zielgruppe hervorruft oder haben sich nicht gerade besonders motivierte und qualifizierte Akteure in die Interventionsgruppe hineinselektiert? Auf aggregierter Ebene wählen Akteure der EZ oftmals besonders leistungsfähige oder auch besonders marginalisierte oder bedürftige räumliche Einheiten oder Organisationen wie Schulen, Gemeinden oder Provinzverwaltungen für ihre Maßnahmen aus. Dieser Bias erfolgt oft, weil sich durch die Wahl qualifizierter Einheiten eine höhere Erfolgswahrscheinlichkeit der Intervention erhofft wird oder auch weil die Wahl besonders prekärer Verhältnisse für die Intervention aus normativen Gründen geboten erscheint. Zur Folge hat eine solch systematische Auswahl jedoch schwerwiegende Endogenitätsprobleme bei der Wirkungsmessung. Der Vorteil experimenteller und quasi-experimenteller Designs besteht daher in der Eliminierung dieser Störgrößen und einer vergleichsweise eindeutigen Zuordnung von Ursache und Wirkung.

Doch sind RIE mit ihren experimentellen und quasi-experimentellen Komponenten deshalb Goldstandard der Evaluation oder gar methodischer Königsweg der Sozialwissenschaften und wie weit tragen die Argumente, die Kritiker von RIE gemeinhin anführen? Die am häufigsten betrachteten Kritikpunkte lassen sich in unterschiedliche Kategorien einordnen. Es sind erstens Kritikpunkte, die sich auf das theoretisch-analytische Fundament von RIE beziehen und diesen die Vernachlässigung von Theoriebasiiertheit und mangelnde Untersuchung von Kausalmechanismen vorwerfen (Deaton, 2010; Deaton & Cartwright, 2018; Donovan, 2018). Es sind zum zweiten normative Kritikpunkte, die den Fokus auf die ethische Unangemessenheit bzw. die Instrumentalisierung von RIE für eine Ökonomisierung der EZ und deren Fokus auf die Kontrollfunktion von Evaluierung legen (Baele, 2013; Donovan, 2018; Bédécarrats,

Guérin & Roubaud, 2019). Drittens schließlich sind es Kritikpunkte, die sich auf Aggregationsfähigkeit und die thematische Reichweite von RIE beziehen (Deaton, 2010; Pritchett & Sandefur, 2015; Kvangraven, 2019).

*Ad 1) theoretische Fundierung und Kausalmechanismen:* Eine oftmals anzutreffende Kritik an RIE führt an, dass experimentelle und quasi-experimentelle Designs nicht oder nur geringfügig theoriegeleitet und zudem lediglich auf die Identifikation von Kausalzusammenhängen ausgerichtet seien, nicht jedoch auf die der Kausalität zu Grunde liegenden Mechanismen (u. a. Deaton, 2010; Deaton & Cartwright, 2018). Diese Kritik kann fallspezifisch zutreffen, richtet sich aber nur auf sehr rudimentäre experimentelle bzw. quasi-experimentelle Untersuchungen. Qualitativ hochwertige RIE, so der mittlerweile breite Konsens auch unter den Befürworter\*innen des Ansatzes, sind theoriegeleitete Untersuchungen, die eine theoriebasierte und hypothesengestützte Interventionslogik eines Programms nutzen, erarbeiten bzw. rekonstruieren und diese auch mit der relevanten breiteren sozialwissenschaftlichen Auseinandersetzung verknüpfen (White, 2009). Dass Theoretiker\*innen und Empiriker\*innen nicht einfach zusammenfinden (Deaton, 2010) und die theoretische Anleitung und Einbettung empirischer Studien kein einfaches Unterfangen ist, gilt denn auch nicht nur für RIE. Insofern ist ein sporadisch auftretendes Missverhältnis von Theorie und Empirie denn auch kein spezifisches Defizit von RIE, sondern eine Herausforderung für qualitativ hochwertige Evaluation bzw. Sozialforschung insgesamt.

Weiterhin muss sich die Theoriebasiertheit von RIE nicht nur auf die Identifizierung von Kausalzusammenhängen beziehen, sondern kann auch Fragen nach den Ursachen der identifizierten Zusammenhänge nachgehen. Sich der Öffnung solch schwarzer Boxen zu widmen, kann nicht nur mittels quantitativer Mediatorenanalyse geschehen. Vielmehr sind hierfür auch qualitative Komponenten des *Process Tracing* geeignet, die sowohl vor oder auch nach der experimentellen/quasi-experimentellen Erhebung eingebaut werden und auch ethnographische Elemente enthalten können. Hierbei wird aus der aktuellen Diskussion um die Untersuchung von Kausalmechanismen in der Evaluation (Schmitt, 2020) deutlich, dass mittlerweile auch RIE eine Vielzahl von Möglichkeiten nutzen, quantitative und qualitative Komponenten der Datenerhebung miteinander zu kombinieren (Oakley, Strange, Bonell, Allen & Stephenson, 2006; White, 2013; Bell & Peck, 2016; Jimenez et al., 2018).

„RCTs can play a role in building scientific knowledge and useful predictions but they can only do so as part of a cumulative program, combining with other methods, including conceptual and theoretical development, to discover not ‘what works’, but ‘why things work’.” (Deaton & Cartwright, 2018, p. 2)

Sicherlich ist mit dem Bedeutungsaufschwung von RIE und der Verbreitung experimenteller und quasi-experimenteller Verfahren auch eine zunehmende Zahl solcher Untersuchungen einhergegangen, die weder eine sonderlich ausgeprägte Theoriebasiertheit noch zufriedenstellende Erklärungsangebote hinsichtlich der wirkenden Kausalmechanismen erkennen lassen (Deaton, 2010). Eine grundsätzliche Kritik am Instrument der RIE lässt sich hierdurch jedoch nicht ableiten, da es sich theoriebasiert anwenden lässt und qualitative Komponenten zur Identifizierung von Kausalmechanismen integriert werden können. Insofern hatten die genannten Kritikpunkte aufgrund ihrer fallspezifischen Berechtigung durchaus die positive Auswirkung, Multi-Methoden-Ansätze und Theoriebasiertheit bei der Durchführung von RIE zu befördern.

*Ad 2) Ethische Unangemessenheit, Ökonomisierung und Fokussierung auf die Kontrollfunktion:* Weitere Kritikpunkte sind eher normativen Ursprungs. Hierbei sind zunächst die ethischen Probleme zu nennen, die bei einer Randomisierung von Entwicklungsinterventionen entstehen können. Vor allem wenn Interventionen das entwicklungspolitische *Do-No-Harm*-Gebot zu verletzen drohen und mit der Randomisierung negative Konsequenzen für die Interventions- oder Kontrollgruppe verbunden sind, sollten Feldexperimente aus normativ-ethischen Gründen unterbleiben (u. a. Barret & Carter, 2010; Baele, 2013). Es ist daher einerseits notwendig, die Durchführung eines geplanten Experiments vor Beginn durch eine unabhängige Ethikkommission bewerten zu lassen.

Andererseits droht die intensive Diskussion um die ethischen Standards von Feldexperimenten in der EZ in eine Schieflage zu geraten. Die makro-quantitative Wirksamkeitsdebatte hat nicht unerhebliche Zweifel an der entwicklungspolitischen Wirksamkeit genährt und viele Projektevaluierungen lassen Bedenken aufkommen, ob deren Wirksamkeitsbefunde methodisch hinreichend abgesichert sind (u. a. Noltze, Euler & Verspohl, 2018). Angesichts dieser Ausgangssituation ist es verwunderlich, dass an nicht randomisierten Interventionen bislang nur wenig Kritik an fehlenden ethischen Standards aufgekommen ist. Bislang wird das Gros an Maßnahmen mit ihren potenziellen Konsequenzen auf etwa Steuer-, Wahl-, Bildungs- oder Konfliktverhalten ohne größere ethische Prüfungen akzeptiert, obwohl über die zu erwartenden Wirkungen Unsicherheit herrscht. Angesichts dieser Ausgangslage scheint die auch ethisch begründbare Forderung nach mehr RIE zwecks Reduzierung dieser Unsicherheit nachvollziehbar (Center for Global Development, 2006). Hierbei kann zudem darauf verwiesen werden, dass sich die ethische Kritik an RIE fast ausnahmslos auf die Randomisierung in Feldexperimenten beschränkt. Quasi-experimentelle Verfahren hingegen bieten aus ethischer Perspektive ein breiteres Einsatzfeld und können wohl begründete ethische Bedenken bei einzelnen Feldexperimenten insofern oftmals gut ersetzen.

Ein zweiter normativer Kritikpunkt postuliert, dass RIE vornehmlich die Kontrollfunktion von Evaluierung zu bedienen suchen und einer Ökonomisierung des Politikfeldes bzw. dessen Anpassung an neoliberale Ideen Vorschub leisten (Bédécarrats Guérin & Roubaud, 2019; Kvangraven, 2019). Allerdings haben beide Argumente aus einer evaluatorisch-analytischen Perspektive wenig Substanz. Sicherlich können RIE auch der Funktion der Rechenschaftslegung dienen, indem sie in Form einer Ex-Post-Evaluierung die Wirkungen einer Entwicklungsmaßnahme erfassen und damit auch als kritische Evidenz über deren Fortführung oder Replizierung herangezogen werden. Gleichwohl sind gerade Feldexperimente – insbesondere solche mit mehreren Interventionsarmen – oftmals so angelegt, dass sie als Begleitforschung großer Sozialprogramme auf das Lernen zu Beginn der Implementierung ausgerichtet sind. So kann die mehrjährig andauernde Ausbreitung und Anpassung großer Sozialprogramme von den Ergebnissen begleitender RIE profitieren. Ebenso sind auch summative RIE nicht nur auf die Kontrollfunktion von Evaluierung ausgelegt, sondern deren Evidenz kann für die Weiterentwicklung und Anpassung der Instrumente dienlich sein. Es mag daher zwar zutreffen, dass RIE oftmals auch aus einer *top-down*-Logik als Kontrollinstrument eingesetzt wurden und nicht auf strukturiertes Lernen und Wissensmanagement ausgerichtet waren (Pritchett, Samji & Hammer, 2013). Gleichwohl zeigen eine Vielzahl von RIE auch, wie diese zu Lern- und Anpassungszwecken genutzt werden können.

Ähnliches gilt prinzipiell auch für den Vorwurf neoliberaler Ökonomisierung der EZ mittels des Instruments von RIE. Es ist möglich, dass kritische Elemente aus RIE in der politischen Auseinandersetzung von Gegnern der EZ aufgenommen und ins Feld geführt werden. Ebenso ist es möglich, dass die Ergebnisse aus RIE für Argumente genutzt werden, Anpassungen und Effizienzsteigerungen zu fordern. In Abhängigkeit von der produzierten Evidenz, können die Ergebnisse aus RIE jedoch auch als Plädoyer für Forderungen wirken, die gegen libertäre Politiken gerichtet sind und etwa auf eine Erhöhung von EZ-Zahlungen zielen. So sind maßgebliche Vertreter\*innen von RIE unter den Entwicklungsökonom\*innen weder als Befürworter\*innen neoliberaler Strategien aufgetreten, noch haben sie für eine Abschaffung der EZ plädiert. Was jedoch von den Verfechter\*innen der rigorosen Evidenzagenda zugestanden wird, ist, dass diese Agenda eng verwandt ist mit Kernkonzepten des *New Public Managements*, die den öffentlichen Sektor stärker in Analogie zu marktwirtschaftlichen, individualisierten Anreizsystemen gestalten wollen (White, 2019). Ob die Ausgestaltung von *New-Public-Management*-Systemen jedoch ausschließlich von einer libertären Idee von Marktwirtschaft getrieben wird, ist umstritten.

*Ad 3) Externe Validität und thematische Reichweite:* Ein letztes Bündel an Kritikpunkten befasst sich mit der Reichweite und Aggregationsmöglichkeiten von RIE und offenbart im Unterschied zu den vorherigen Kritikpunkten strukturelle Begrenzungen rigoroser Wirkungsevaluierung. Zunächst ist hierbei das Problem geringer externer Validität zu nennen. Die berechtigte Kritik lautet hierbei, dass die Ergebnisse von RIE letztlich stark kontextgebundene Fallstudien mit einer hohen internen aber geringen externen Validität seien, aus denen eben keine Empfehlungen für andere Kontexte bzw. Populationen gezogen werden können (Prichett & Sandefur, 2015). Gerade für die Ableitung von kontextübergreifenden Empfehlungen ist jedoch ein hohes Maß an externer Validität von großer Bedeutung. Dani Rodrik (2009) beschreibt treffend, dass ebenso wie eine herkömmlich quantitative Studie zusätzliche Argumente außerhalb ihres empirischen Kerns benötige, um ihre interne Validität zu begründen, eine experimentelle Untersuchung zusätzlicher empirischer Argumente zur Begründung ihrer externen Validität bedürfe. Diese Herausforderung kann punktuell durch die Wiederholung des Experiments oder dessen paralleler Durchführung in unterschiedlichen Kontexten behoben werden. Gleichwohl ist der bloße Vergleich der Ergebnisse für eine robuste Aussage über die externe, kontextunabhängige Validität nicht ausreichend.

Hierzu können systematische Reviews genutzt werden, die speziell auf die Aggregation von Ergebnissen aus experimentellen und quasi-experimentellen Studien ausgerichtet sind und mittlerweile wichtiger Bestandteil der von Howard White (2019) postulierten Evidenzagenda in der EZ sind. Systematische Reviews in der sozialwissenschaftlichen Entwicklungsforschung sind ein vergleichsweise neues Phänomen und folgen in ihrer Methodik dabei den Verfahren aus Medizin und Gesundheitsökonomik. Sie finden und bewerten zunächst nach einem ex ante bestimmten Protokoll möglichst alle existierenden Studien, die das erwünschte Entwicklungsziel zum Gegenstand haben (Mallett, Hagen-Zanker, Slater & Duvendack, 2012). Sodann werden die Daten derjenigen Studien aggregiert, die strenge Qualitätskriterien erfüllen, um hierdurch einen durchschnittlichen, kontextübergreifenden Durchschnittseffekt der Intervention und dessen statistische Signifikanz zu ermitteln. Insofern sind Systematische Reviews ein geeignetes Mittel, um die existierende Evidenz aus einzelnen RIE zu einem spezifischen Outcome zusammenzu-

führen und damit das Problem geringer externer Validität zu reduzieren (Mallett, Hagen-Zanker, Slater & Duvendack, 2012; White, 2018). Sie stellen damit eine begrüßenswerte Ergänzung zu einzelnen RIE und makroquantitativen Studien dar.

Gleichwohl werden Systematische Reviews im Unterschied zur Medizin das Problem geringer externer Validität absehbar kaum lösen. Zu gering ist oftmals die Fallzahl an rigorosen Wirkungsevaluierungen, die für einen Systematischen Review in Frage kommen, zu unterschiedlich sind die zu Grunde liegenden Interventionen und zu vielfältig die sozialen Kontexte, in denen keine RIE stattgefunden hat (Mallett, Hagen-Zanker, Slater & Duvendack, 2012). Zudem setzt sich auch bei Vertreter\*innen der rigorosen Evidenzagenda dabei zunehmend die Erkenntnis durch, dass theoriegeleitete und um qualitative Methodenelemente ergänzte Systematische Reviews einen Erkenntnismehrwert bringen (Jimenez et al., 2018; White, 2018). Insofern scheint sich zu bestätigen, dass traditionelle, fallübergreifende quantitative Studien mit Herausforderungen interner Validität für die Politikberatung ebenso erkenntnisreich sind wie einzelne rigorose Wirkungsevaluierungen oder auch systematische Reviews mit ihren Beschränkungen externer Validität (Pritchett & Sandefur, 2015). Erneut besteht der Königsweg damit in einer Kombination unterschiedlicher, auf die spezifische Fragestellung zugeschnittener Verfahren aus dem Baukasten der empirisch-analytischen Sozialforschung, um ein hohes Maß an interner und externer Validität zu erzielen.

Hinsichtlich der thematischen Reichweite von RIE hat sich gezeigt, dass diese nicht auf wenige Sektoren beschränkt bleiben müssen. Ursprünglich stark auf die sozialen Sektoren wie Gesundheit, Bildung sowie auf Armutsbekämpfung beschränkt, hat sich in den letzten Jahren gezeigt, dass RIE auch in den allermeisten anderen Sektoren der Entwicklungszusammenarbeit angewendet werden kann. RIE können im Governance-Bereich durchgeführt werden (Funk, Groß, Leininger & Schiller, 2019), sind auch in der (angelsächsischen) Politikwissenschaft mittlerweile verankert (Humphreys & Weinstein 2009; Hyde, 2015) und finden ebenso bei Umwelt- und Klimaanpassungsthemen sowie auch in der humanitären Hilfe Anwendung. Gerade mittels quasi-experimenteller Untersuchungsdesigns und unter Verwendung neuer Datentypen wie etwa Satellitendaten sind auch sektorübergreifende bzw. multisektorale Evaluierungen möglich (Leppert, Hohfeld, Lech & Wencker, 2018).

Wenn auch deutlich häufiger anwendbar als noch vielfach in der Entwicklungsspraxis angenommen, sind RIE jedoch hinsichtlich ihrer Evaluierungsgegenstände strukturell begrenzt, da eine Vielzahl von EZ-Interventionen mit experimentellen- und quasi-experimentellen Methoden kaum zu evaluieren sind. Hierzu zählen insbesondere viele Maßnahmen auf der Meso- oder Makroebene, die sich der Reformberatung und Reformimplementierung auf regionaler oder nationaler Ebene widmen. Die begrenzte Fallzahl – ein oder wenige Ministerien, Behörden, NGOs, etc. – der adressierten Einheiten macht dann ein (quasi-)experimentelles Design unmöglich. Zudem sind nicht alle entwicklungspolitisch relevanten Evaluationsfragen auf Wirkungen gerichtet. Institutionelle Evaluierungen einzelner Organisationen, Prozessevaluationen, Portfolioanalysen und (zumindest teilweise) auch Strategie- und Instrumentenevaluationen sind nicht oder nur begrenzt mit RIE zu bearbeiten. Diese strukturelle Begrenzung des Instruments sollte jedoch nicht dazu führen, RIE „Kleinteiligkeit“ und mangelnde Fähigkeit zur Thematisierung größerer theoretischer und systemischer Fragen zu unterstellen, wie dies einige Kritiker\*innen tun (Kvangraven, 2019; Bédécarrats Guérin & Roubaud, 2019). RIE sind in einer Vielzahl von Fällen gut geeignet, die Wirksamkeit kon-

kreter Entwicklungsmaßnahmen zu identifizieren. Sind sie theoriegeleitet, dann lassen sich ihre Befunde in vielen Fällen auch an breitere Debatten in Entwicklungsforschung und Sozialwissenschaften koppeln; so etwa in der Arbeitsmarktforschung oder in der Diskussion um direktdemokratische Partizipationsmechanismen. Nur weil viele RIE auf der Ebene der Individuen ansetzen, müssen sie nicht notwendigerweise einer Verantwortungsverschiebung hin zum Individuum Vorschub leisten, können eine Bewertung von Politiken auf der Makroebene gleichwohl nicht ersetzen.

## 4 Anwendung und Nutzung rigoroser Wirkungsbeurteilung in der EZ

Angesichts der Bedeutung von RIE ist neben der Frage nach deren wissenschaftlich-evaluatorischer Eignung zu klären, wie es um die praktische Durchführung und Nutzung von RIE in der internationalen EZ bestellt ist. Ein erster Anhaltspunkt für die Durchführung ist das größte Repositorium für RIE in der angewandten Entwicklungsforschung, das gegenwärtig von der International Initiative for Impact Evaluation (3ie) betrieben wird. Das Repositorium umfasst gegenwärtig über 3.700 einzelne RIE sowie 722 Systematische Reviews ([www.3ie.org](http://www.3ie.org)), die entlang unterschiedlicher Kriterien geordnet werden können. Die tatsächlich existierende Zahl von durchgeführten RIE in der internationalen EZ dürfte allerdings deutlich größer sein und war in den letzten beiden Jahrzehnten durch kontinuierlich hohe Steigerungsraten gekennzeichnet (Manning, Goldmann & Hernandez, 2020). Die überwiegende Mehrzahl dieser RIE wird aus Mitteln der EZ (ko-)finanziert; sei es direkt aus den Projekten und Programmen oder durch die Finanzierung aus größeren Fonds oder Förderprogrammen. Insgesamt ist die Finanzierung und Nutzung von RIE in der Entwicklungsforschung daher immer noch stark durch die Geberorganisationen getrieben. Gleichzeitig ist aber auch ein Trend erkennbar, wonach insbesondere in Schwellenländern wie Mexiko, Indien, Kolumbien oder Indonesien die Beauftragung von RIE als summative Evaluierungen oder Begleitforschung durch staatliche Entscheidungsträger zunimmt und hiermit eigene Sozialprogramme analysiert werden (Manning, Goldmann & Hernandez, 2020). Teils werden aber auch diese Initiativen von Organisationen der EZ (ko-)finanziert.

In der internationalen EZ ist die Finanzierung, Anwendung und Nutzung von RIE im internationalen System durch ein hohes Maß an Heterogenität gekennzeichnet und nicht in einen ordnungspolitischen Rahmen eingebettet. Am stärksten ausgeprägt und systematisiert ist die Anwendung und Nutzung von RIE in einigen internationalen Finanzorganisationen und Entwicklungsfonds. Hierbei zählt die Weltbank als Vorreiter, da sie über ein großes Programm zur Durchführung von RIE verfügt (Development Impact Evaluation DIME). In Kooperation mit Partnerländern und ko-finanziert von anderen Gebern, werden dort eine Vielzahl von RIE und Systematischen Reviews durchgeführt und veröffentlicht. Auch in einigen anderen internationalen Entwicklungsbanken und Entwicklungsfonds ist die Anwendung von RIE verbreitet, allerdings in geringerem Umfang und weniger systematisch institutionalisiert als in der Weltbank (u. a. Interamerican Development Bank, 2017).

Parallel zu den Aktivitäten der internationalen Organisationen der EZ gibt es mittlerweile eine Vielzahl internationaler Nichtregierungsorganisationen (NGOs), die sich

als Expertenorganisationen im Schnittfeld zwischen Wissenschaft und Praxis etabliert haben.<sup>7</sup> Organisationen wie das am Massachusetts Institute of Technology (MIT) angesiedelte J-PAL (Abdul Latif Jameel Poverty Action Lab), IPA (Innovations for Poverty Action) oder 3ie haben sich dabei auf die Durchführung von RIE, deren Verbreitung und – im Falle von 3ie – auch als Finanzierungsplattform für RIE spezialisiert. Wenn gleich ebenfalls in überwiegendem Maße durch Mittel der internationalen EZ finanziert, setzen sich diese Organisationen zum Ziel, die Anwendung und Nutzung von RIE durch Entwicklungs- und Schwellenländer zu fördern. Insbesondere 3ie verfolgt zudem das Ziel, die Evidenz aus RIE transparent aufzubereiten, zu systematisieren und als kostenloses, globales Kollektivgut zur Verfügung zu stellen (siehe [www.3ie.org](http://www.3ie.org)).

Im Unterschied zu den sichtbaren Aktivitäten auf der Ebene der Finanzorganisationen und denjenigen von NGOs im Schnittfeld zwischen Wissenschaft, Evaluierung und Praxis ist die Nutzung von RIE auf bilateraler Ebene unübersichtlich. Vornehmlich unter den angelsächsischen und partiell auch bei den nordischen Gebern werden RIE finanziert und genutzt, sind aber keinesfalls dominantes Instrument der Evaluierung. Trotz der vergleichsweise breiteren Akzeptanz und Nutzung von RIE ist aber auch in Ländern wie den USA oder Großbritannien über Zeit und Sektoren kaum eine systematische Verankerung und Normierung von RIE in der EZ erkennbar (Manning & White, 2014). Eine eindeutige Zuordnung der Evaluierungspraxis nach Geberländern wird auch dadurch erschwert, weil innerhalb dieser verschiedene Durchführungsorganisationen mit unterschiedlichen Evaluierungspraktiken existieren können. In anderen bedeutenden Geberländern wie Deutschland, Frankreich und Japan war die Beauftragung von RIE durch die Akteure der EZ bislang eher die Ausnahme als die Regel. In Deutschland etwa ist während der letzten Dekade die Expertise in Universitäten und anderen Forschungseinrichtungen zu RIE stetig angewachsen und mittlerweile auch jenseits der Entwicklungsökonomie verbreitet. Gleichwohl wurden RIE seitens der finanziellen und technischen Zusammenarbeit bislang nur sporadisch beauftragt, wobei in der deutschen EZ gegenwärtig eine Trendwende hin zu einer breiteren Akzeptanz und Nutzung von RIE erkennbar ist (Bruder, Faust & Krämer, 2019; Florian, Krisch, Till & Hermanns, 2019).

Trotz der international heterogenen Landschaft, was Anwendung und Nutzung von RIE anbelangt, hat sich durch die Dynamik in der Entwicklungsforschung eine praxisorientierte, methodisch wie inhaltlich geprägte epistemische Gemeinschaft (Haas, 1992) mit stark instrumentellem Fokus auf das Policy-Instrument von RIE herausgebildet. Insofern verfügt diese Gemeinschaft auch über Merkmale einer *instrumental constituency* (Simons & Voss, 2018). Sie hat Überlappungen mit der primär universitär orientierten Wissenschaft mit ihrem Fokus auf möglichst anspruchsvollen Feldexperimenten, ist insgesamt aber pragmatischer und versucht als Scharnier zwischen Wissenschaft und Implementierungspraxis eine stärker anwendungsbezogene Nutzung von RIE für praktische Zwecke voranzutreiben. Die Protagonisten dieser epistemischen Gemeinschaft arbeiten denn auch nur teils an Universitäten und Hochschulen, sondern oftmals in Evaluierungseinrichtungen bilateraler und multilateraler EZ-Organisationen, in den genannten wissenschaftsnahen NGOs oder partiell auch in staatlichen Evaluierungs- und Analyseeinheiten von Entwicklungs- und Schwellenländern. Auch ist diese Gemeinschaft in nationalen und internationalen Evaluierungsverbänden vertreten. Doch aufgrund der anhaltenden – auch epistemologischen – Auseinandersetzungen um die Angemessenheit und Anwendbarkeit von RIE sind die Verbände bislang nicht als



Förderatoren von RIE aufgetreten. Ähnliches gilt auch für das Evaluierungsnetzwerk innerhalb der OECD, in dem alle bilateralen Geber der OECD sowie auch multilaterale Organisationen (als Beobachter) vertreten sind. Wenngleich dieses Netzwerk in der Vergangenheit als ein bedeutsamer Standardsetzer in der internationalen Entwicklungsevaluierung aufgetreten ist, waren die Ansichten innerhalb der Gebergemeinschaft zur Anwendung von RIE bislang zu heterogen, als dass dieses Netzwerk eine substanzielle Rolle als Förderator von RIE hätte einnehmen können.

Insofern hat sich die Debatte um RIE bislang als Barriere für eine systematischere und strukturiertere Anwendung von RIE auf internationaler Ebene erwiesen, da sie auf der Ebene von Ministerien, multilateralen Organisationen, bilateralen Durchführungsorganisationen und auch zivilgesellschaftlichen Akteuren ganz unterschiedlich motivierte Vorbehalte provoziert hat; von epistemologischen Auseinandersetzungen zwischen einem empirisch-analytischen und einem konstruktivistisch-postkolonialen Wissenschaftsverständnis bis hin zu Vorbehalten, was Anwendbarkeit und Kontrolle von RIE in der konkreten Praxis anbelangt. Mit einigen relevanten Ausnahmen insbesondere in multilateralen Organisationen, den USA und Großbritannien erfolgte die Anwendung meist dezentralisiert und war oft mehr von wissenschaftlichen Dispositionen und praktischen Erkenntnisinteressen von Projekt- oder Programmleitungen abhängig als von einer gezielten Umsetzung von Evidenzstrategien der Entwicklungsorganisationen. Damit bleibt aber auch der praktische Wert von RIE bislang weit hinter ihrem Potenzial zurück. Um dieses künftig auszuschöpfen, ist es notwendig, dass Entwicklungsorganisationen Strategien, Anreizsysteme und finanzielle Ressourcen für eine situationsangemessene Nutzung von RIE etablieren bzw. bereitstellen sowie die hieraus gewonnene Evidenz systematischer und nicht nur punktuell nutzen und zudem auch der Allgemeinheit im Sinne eines Kollektivgutes zur Orientierung und Aggregation zur Verfügung stellen.

## 5 Fazit

Rigorese Wirkungsevaluierungen mit ihren experimentellen und quasi-experimentellen Komponenten haben in den letzten beiden Dekaden an Bedeutung in der Evaluierung entwicklungspolitischer Maßnahmen gewonnen. Eine größer werdende Zahl an Ökonom\*innen und mittlerweile auch Wissenschaftler\*innen anderer Sozialwissenschaften sind entsprechend ausgebildet worden und plädieren für eine weitere Verbreitung (quasi-)experimenteller Analysen in der Entwicklungsforschung. Der Vorteil besteht dabei im Vergleich zu anderen Methoden darin, intendierte und nicht-intendierter Wirkungen von Maßnahmen besser identifizieren zu können. In der EZ bedienen RIE daher zum einen den funktional begründbaren, hohen Bedarf an Evaluierung. Zum anderen bedienen sie die Forderung nach zuverlässiger Evidenz über die Wirkungen der EZ aus einer zunehmend kritischen Wirkungsdebatte.

Doch trotz dieser nachvollziehbaren Genese von RIE in der Entwicklungszusammenarbeit hält die Debatte um deren Vor- und Nachteile an. Aus empirisch-analytischer Sicht spricht allerdings wenig gegen eine Verbreitung von RIE in Entwicklungsforschung und Entwicklungszusammenarbeit. Insgesamt haben sich „die Bedenken hinsichtlich grundsätzlicher oder sektoraler Einschränkungen der Anwendbarkeit solcher Verfahren, unüberwindbarer ethischer Probleme oder nicht ausreichend theorieba-

sierter Vorgehensweisen als unbegründet, lösbar oder nur für eingeschränkte Bereiche als relevant erwiesen“ (Bruder, Faust & Krämer, 2019, S. 2). RIE können vielmehr in unterschiedlichen Sektoren theoriebasiert, mit anderen Methoden kombiniert und ethisch vertretbar eingesetzt werden. Sie müssen weder einer neoliberalen Ideologie verpflichtet sein, einen Kontrollbias haben noch kleinteilige akademische Reparaturarbeiten ohne Bezug zu größeren Debatten sein. Oft liegt denn der Ursprung der Kritik auch in den nur begrenzt miteinander versöhnbaren Schulen eines empirisch-analytischen bzw. verstehend-hermeneutischen Wissenschaftsverständnisses.

Gleichwohl sind RIE nicht für alle wirkungsbezogenen Fragestellungen der EZ geeignet und sollten auch nicht präferiertes Verfahren der Evaluation bei jedweder entwicklungspolitischen Maßnahme sein. Schließlich kann auch die Aggregation von RIE in Systematischen Reviews nicht strukturelle Begrenzungen für die externe Validität auflösen. Auch hier ist eine Kombination unterschiedlicher (quantitativer und qualitativer) Methoden die zielführende Variante und sollte stärker genutzt werden. Es gilt somit auch für RIE und die Evaluation der EZ ein zentraler Grundsatz anwendungsorientierter Forschung, wonach die Methode dem praktischen Erkenntnisinteresse zu folgen hat und nicht umgekehrt. RIE sind dann besonders zielführend, wenn die potenziellen Wirkungen einer Intervention im Fokus des Erkenntnisinteresses stehen und diese Intervention viele Akteure in ähnlicher Weise erreichen soll (Bruder, Faust & Krämer, 2019, S. 2). Kurzum: RIE sind weder Goldstandard der Evaluation noch Königsweg der Entwicklungsforschung, können aber sehr wohl breit und erkenntnisgewinnend eingesetzt werden.

Was schließlich die Durchführung und Nutzung in der Praxis der internationalen EZ anbelangt, so ist diese insgesamt über die letzten zwei Dekaden deutlich gestiegen. Gleichzeitig ist die Uneinheitlichkeit und Ungleichzeitigkeit der Anwendung, Verbreitung und Nutzung von RIE weiterhin gängige Praxis in der internationalen EZ. Obwohl sich mittlerweile eine stark praxis- und anwendungsorientierte epistemische RIE-Gemeinschaft herausgebildet hat, gibt es bis auf wenige Ausnahmen kaum organisationsweite und schon gar nicht organisationsübergreifende Strategien und Anreizsysteme zur Anwendung und Nutzung von RIE für übergreifendes Lernen und Rechenschaftslegung. Es liegt dabei nahe, dass die intensive Auseinandersetzung um die analytische und normative Eignung sowie die Nutzungsmöglichkeiten von RIE ein Hindernis für die strukturierte Anwendung und Nutzung von RIE auf internationaler Ebene gewesen ist. Allerdings ist die Heterogenität mit Blick auf die Anwendung und Nutzung von RIE auf internationaler Ebene nur begrenzt eine Besonderheit von RIE. Denn trotz des im Vergleich zu anderen Politikfeldern erreichten hohen Maßes an Austausch und Standardisierung in der Entwicklungsevaluierung ist die Heterogenität der Anwendung und Nutzung von RIE auch ein Strukturmerkmal eines allenfalls begrenzt funktionalen Wissensmanagements der internationalen EZ.

## Anmerkungen

- 1 Für wertvolle Anregungen zu einer früheren Version dieses Beitrags danke ich Martin Bruder, Marion Krämer, Julia Leininger, Elisabeth Schneider, Holger Straßheim sowie den Gutachter\*innen dieses Beitrags.
- 2 Wie auch in experimentellen Verfahren in Medizin, Gesundheitsökonomie oder Erziehungswissenschaften basieren entwicklungspolitische Feldexperimente (Randomized Controlled Trials - RCTs) auf

der zufallsbasierten Auswahl von Interventions- und Kontrollgruppe. Dabei wird eine Entwicklungsmaßnahme bei einer Gruppe zufällig aus einer Grundgesamtheit gewählter Akteure durchgeführt (Interventionsgruppe) und eine ebenso zufällig gewählte Gruppe dient als Kontrollgruppe. Bei hinreichender Größe können beide Gruppen in ihren beobachtbaren wie nicht beobachtbaren Eigenschaften als gleich erwartet werden. Die Differenz der Zielgröße zwischen beiden Gruppen nach der Durchführung der Maßnahme schätzt einen unverzerrten Nettoeffekt der Intervention. Neben RCTs können auch quasiexperimentelle Verfahren ähnlich verlässlich sein. Diese Verfahren führen zwar kein Experiment zu Beginn einer Intervention durch, bedienen sich aber bestimmter statistischer Verfahren, um hierdurch hinreichend große Vergleichsgruppen zu bilden und Störgrößen systematisch ausschließen (Bruder, Faust & Krämer, 2019, S. 1; einführend siehe u. a. Caspari & Barbu, 2008). Für eine anspruchsvolle Einführung in die Methoden und Anwendungsfelder rigoroser Wirkungsevaluierung siehe Frölich & Sperling, 2018.

- 3 Aus der umfangreichen kritischen Befassung mit RIE siehe u. a. Acemoglu 2009; Deaton, 2010; Barrett & Carter, 2010; Pritchett, Samji & Hammer, 2013; Kvangraven, 2019; Donovan, 2018; Bédécarrats, Guérin & Roubaud, 2019.
- 4 Zu den Kenntnissen der deutschen Bevölkerung über die Entwicklungszusammenarbeit und die Nachhaltigen Entwicklungsziele der Agenda 2030 siehe Schneider & Gleser, 2018.
- 5 So verweist beispielsweise der letzte Evaluierungsbericht der KfW aus dem Jahr 2019 auf eine Erfolgsquote von 77% im Zeitraum von 2017-2018 (KfW Entwicklungsbank, 2019, S. 64). Die Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) gibt in ihrem letzten Evaluierungsbericht von 2017 an, dass auf einer fünfstufigen Bewertungsskala 98% der im Berichtszeitraum (2016-2017) evaluierten Maßnahmen mit der Note „eher erfolgreich“ oder besser bewertet wurden (GIZ, 2018, S. 16). Auch im internationalen Bereich werden gemeinhin Erfolgsquoten von mehr als 75% berichtet.
- 6 Endogenitätsprobleme bestehen insbesondere dann, wenn aus einer identifizierten Korrelation zweier Variablen nicht auf einen kausalen Zusammenhang zwischen beiden Variablen geschlossen werden kann, weil nicht plausibel auszuschließen ist, dass eine (unbeobachtete) dritte Variable beide gemeinsam beeinflusst oder die Kausalität auch in die umgekehrte Richtung verläuft.
- 7 Neben den großen internationalen Akteuren wie DIME, J-Pal, IPA und 3ie existiert eine Vielzahl weiterer Organisationen wie GRADE (Grupo de Analisis para el Desarrollo), CEGA (Center for Effective Global Action) oder C4ED (Center for Evaluation and Development). Letztere haben sich zumeist auf die Durchführung von RIE im Auftrag internationaler Entwicklungsorganisationen spezialisiert und haben ihren Sitz sowohl in den OECD-Ländern als auch in Entwicklungsländern. Schließlich ist noch die auf die Durchführung systematischer Reviews spezialisierte Campbell Collaboration zu nennen, die in den letzten Jahren ihre Aktivitäten auf die Entwicklungsforschung ausgeweitet hat (White, 2019).

## Literatur

- Acemoglu, Daron (2009). Theory, General Equilibrium, Political Economy and Empirics in Development Economics. *Journal of Economic Perspectives*, 24 (3), 17-32.
- Baele, Stéphane J. (2013). The ethics of New Development Economics: is the Experimental Approach to Development Economics morally wrong? *The Journal of Philosophical Economics* 7 (1), 2-42.
- Banerjee, Abhijit V. & Duflo, Esther (2011). *Poor Economics: rethinking poverty & the ways to end it*. Random House.
- Barrett, Christopher & Carter, Michael R. (2010). The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Applied Economic Perspectives and Policy*, 32 (4), 515-548.
- Bédécarrats, Florent, Guérin, Isabelle & Roubaud, Francois (2019). All that glitters is not gold. The political economy of randomized evaluations in development. *Development & Change*, 50 (3), 735-762.
- Bell, Stephen & Peck, Laura (2016). On the “How” of Social Experiments: Experimental Designs for Getting Inside the Black Box. *New Directions for Evaluation*, 152, 97-107.
- Bruder, Martin, Faust, Jörg & Krämer, Marion (2019). *Rigorese Wirkungsevaluierung in der deutschen Entwicklungszusammenarbeit*. Policy-Brief 05/2019. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit.

- Caspari, Alexandra & Barbu, Ragnhild (2008). *Wirkungsevaluierungen: Zum Stand der internationalen Diskussion und dessen Relevanz für Evaluierungen der deutschen Entwicklungszusammenarbeit*. Evaluation Working Papers. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.
- Center for Global Development (2006). *When Will We ever learn? Improving Lives through Impact Evaluation*. Washington.
- Deaton, Angus (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48 (2), 424-55.
- Deaton, Angus & Cartwright, Nancy (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.
- Donovan, Kevin P. (2018). The rise of the randomistas: on the experimental turn in international aid, *Economy and Society*, 47 (1), 27-58.
- Dreher, Axel, Fuchs, Andreas, Lang, Valentin & Langlotz, Sarah (2018). Migration: Schaffen wir das? *Frankfurter Allgemeine Zeitung* vom 17.09.2018.
- Dreher, Axel, Sturm, Jan Egbert & Vreeland, James (2009). Global horse trading: IMF loans for votes in the United Nations Security Council. *European Economic Review*, 53 (7), 742-757.
- Dutta, Nabamita, Leeson, Peter & Williamson, Cornelia (2013). The Amplification Effect: Foreign Aid's Impact on Political Institutions. *Kyklos*, 66 (2), 208-228.
- Easterly, William (2002). The cartel of good intentions: the problem of bureaucracy in foreign aid. *The Journal of Policy Reform*, 5 (4), 223-250.
- Easterly, William (2006). *The White Man's Burden. Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York.
- Faust, Jörg & Leiderer, Stefan (2008). Zur Effektivität und politischen Ökonomie der Entwicklungszusammenarbeit. *Politische Vierteljahresschrift*, 49 (1), 129-152.
- Faust, Jörg & Michaelowa, Katherina (Hrsg.) (2013). *Politische Ökonomie der Entwicklungszusammenarbeit*. Nomos: Baden-Baden.
- Faust, Jörg & Ziaja, Sebastian (2012). *German aid allocation and partner country selection: development-orientation, self-interests and path dependency*. Discussion Paper 7/2012. Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).
- Florian, Michael, Krisch, Franziska, Till, Tatjana & Hermanns, Sophie (2019). *Rigorous Impact Evaluation. A Corporate Strategic Review of Causal Analysis during Implementation*. Bonn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.
- Frölich, Markus & Sperlich, Stefan (2018). *Impact Evaluation*. CUP.
- Funk, Evelyn, Groß, Lisa, Leininger, Julia & Schiller, Armin (2019). *Erkenntnisse aus der wirkungsorientierten Begleitforschung: Potential und Grenzen der rigorosen Wirkungsanalyse von Governance-Programmen*. Discussion Paper 13/2019. Bonn: Deutsches Institut für Entwicklungspolitik.
- GIZ Gesellschaft für Internationale Zusammenarbeit (2018). *Evaluierungsbericht 2017 – Wissen was wirkt*. Eschborn.
- Hodler, Roland & Raschky, Paul (2014). Regional Favoratism. *The Quarterly Journal of Economics*, 129 (2), 995-1033.
- Haas, Peter M. (1992). Introduction: Epistemic Communities and International Policy Coordination. *International Organization*, 46 (2), 1-35.
- Humphreys, Macartan & Weinstein, Jeremy (2009). Field experiments and the political economy of development. *Annual Review of Political Science*, 12, 367-376.
- Hyde, Susan (2007). The Observer Effect in International Politics: Evidence from a Natural Experiment. *World Politics*, 60 (1), 37-63.
- Hyde, Susan (2015). Experiments in International Relations: Lab, Survey, and Field. *Annual Review of Political Science*, 18, 403-424.
- Interamerican Development Bank (2017). *IDB's Impact Evaluations – Production Use and Influence*. Washington: Office of Evaluation and Oversight.
- Jimenez, Emmanuel, Waddington, Hugh, Goel, Neeta, Prost, Audrey, Pullin, Andrew, White, Howard, Lahiri, Shaon & Narain, Anmol (2018). Mixing and matching: using qualitative methods to

- improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes. *Journal of Development Effectiveness*, 10 (4), 400-421.
- KfW Entwicklungsbank (2019). 15. *Evaluierungsbericht 2017-2018 – Zu größerer Wirkung in kleineren Städten*. Frankfurt: Kreditanstalt für Wiederaufbau.
- Kvangraven, Ingrid H. (2019). *Impoverished economics? Unpacking the economics Nobel Prize*. Open Democracy Net, 18.10.2019.
- Leppert, Gerald, Hohfeld, Lena, Lech, Lena & Wencker, Thomas (2018). *Impact, Diffusion and Scaling-Up of a Comprehensive Land-Use Planning Approach in the Philippines. From Development Cooperation to National Policies*. Bonn: German Institute for Development Evaluation (DEval).
- Mallett, Richard, Hagen-Zanker, Jessica, Slater, Rachel & Duvendack, Maren (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness*, 4 (3), 445-455
- Manning, Richard, Goldman, Ian & Hernandez Licona, Gonzalo (2020). *The impact of impact evaluation: Are impact evaluation and impact evaluation synthesis contributing to evidence generation and use in low- and middle-income countries?*. WIDER Working Paper 2020/20. Helsinki: UNU-WIDER.
- Manning, Richard & White, Howard (2014). Measuring results in development: The role of impact evaluation in agency-wide performance measurement systems. *Journal of Development Effectiveness*, 6 (4), 337-349.
- Martens, Bertin, Mummert, Uwe & Murrell, Peter (2002). *The Institutional Economics of Foreign Aid*. CUP.
- Meyer, Wolfgang, Bär, Thomas, Faust, Jörg, Jan, Susanne v., Silvestrini, Stefan & Wein, Stefanie (2019). Die DeGEval-Standards in der deutschen bilateralen Entwicklungszusammenarbeit. In: Jan Ulrich Hense, Wolfgang Böttcher, Michael Kalman & Wolfgang Meyer (Hrsg.), *Evaluation: Standards in unterschiedlichen Handlungsfeldern - Einheitliche Qualitätsansprüche trotz heterogener Praxis?* (S. 165-182). Münster: Waxmann.
- Molina Millán, Teresa, Barham, Tania, Macours, Karen, Maluccio, John A. & Stampini, Marco (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence. *The World Bank Research Observer*, 34 (1), 119-159.
- Morgenthau, Hans (1962). A Political Theory of Foreign. *American Political Science Review*, 56 (2), 301-309.
- Noltze, Martin, Euler, Michael & Verspohl, Ida (2018). *Meta-Evaluierung von Nachhaltigkeit in der deutschen Entwicklungszusammenarbeit*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Oakley, Ann, Strange, Vicki, Bonell, Chris, Allen, Elisabeth & Stephenson, Judith (2006). Process evaluation in randomised control trials of complex interventions. *British Medical Journal*, 332, 413-416.
- Olken, Benjamin A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115 (2), 200-249.
- Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) (2016). *Evaluation Systems in Development Co-operation – A Review*. Paris.
- Polak, Jan T., Guffler, Kerstin & Scheinert, Lara (2017). *Weltwärts-Freiwillige und ihr Engagement in Deutschland*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Pritchett, Lant, Samji, Salimah & Hammer, Jeffrey (2013). *It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space*. Working Paper 322. Washington: Center for Global Development.
- Pritchett, Lant & Sandefur, Justin (2015). Learning from Experiments When Context Matters. *American Economic Review*, 105 (5), 471-75.
- Rawlings, Laura B. & Rubio, Gloria M. (2005). Evaluating the Impact of Conditional Cash Transfer Programs. *The World Bank Research Observer*, 20 (1), 29-55.
- Rodrik, Dani (2009). *The New Development Economics: We Shall Experiment, but How Shall We Learn?*. Harvard Kennedy School Working Paper No. RWP08-055.

- Schmitt, Johannes (2020). *The Causal Mechanism Claim in Evaluation: Does the Prophecy fulfill?* (mimeo).
- Schneider Sebastian H. & Gleser, Solveig H. (2018). *Meinungsmonitor Entwicklungspolitik 2018: Einstellungen zu Entwicklungszusammenarbeit und nachhaltiger Entwicklung*. Bonn: Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval).
- Simons, Arno & Voss, Jan-Peter (2018). Instrument Constituencies: The politics of policy formulation. *Policy and Society*, 37 (1), 14-35.
- Stockmann, Reinhardt (2004). *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren*. Centrum für Evaluation CEval Arbeitspapier Nr. 9. Saarbrücken.
- White, Howard (2009). Theory-Based Impact Evaluation: Principles and Practice. *Journal of Development Effectiveness*, 3 (1), 271-284.
- White, Howard (2013). The use of mixed methods in randomized control trials. *New directions for evaluations*, 138, 61-73.
- White, Howard (2018). Theory-based systematic reviews. *Journal of Development Effectiveness*, 10 (1), 17-38.
- White, Howard (2019). The twenty-first century experimenting society: the four waves of the evidence revolution. *Palgrave Communications*, 5 (47). Online verfügbar unter: <https://www.nature.com/articles/s41599-019-0253-6> [30.9.2019].

*Anschrift des Autors:*

Prof. Dr. Jörg Faust, DEval Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, Fritz-Schäffer-Str. 26, 53113 Bonn, E-Mail: [joerg.faust@deval.org](mailto:joerg.faust@deval.org).

Petra Kaps

## Formen der Evaluierung experimenteller Politikgestaltung

### Zusammenfassung

Die Reformen der Arbeitsmarktpolitik der letzten 20 Jahre wurden von einer Reihe politischer Experimente begleitet, die jeweils anlassbezogen evaluiert wurden. An drei Beispielen – den Kombilohn-Experimenten, Experimenten zur Pauschalierung einmaliger Leistungen in der Sozialhilfe und am Wettbewerb zwischen Optionskommunen und Arbeitsagenturen in Arbeitsgemeinschaften mit den Kommunen – werden Evaluierungsarrangements als Kombinationen von Evaluierungsgegenständen, Evaluierungsaufträgen und Evaluierungspraktiken untersucht. Dabei wird nach Unterschieden und Gemeinsamkeiten zwischen diesen Arrangements, nach „varieties of evaluation“ (Schwab, 2017, S. 201), gefragt. Und es wird der Zusammenhang zwischen den institutionellen Rahmenbedingungen innerhalb des politisch-administrativen Systems und der Gestaltung dieser Evaluierungsarrangements untersucht. Dabei wird auch skizziert, welcher „politische Gehalt von Evaluierungsprozessen“ (ebd., S. 212) im Vergleich der Beispiele sichtbar wird.

**Schlagerworte:** Evaluationsforschung; politische Experimente; Arbeitsmarktpolitik; Verwaltungsforschung; Advokaten-Koalitionen

### Abstract

*Forms of evaluation in cases of experimental policy making*

The reform of labour market policies over the past 20 years has been accompanied by a series of political experiments. Each of these experiments has been evaluated. For three of these evaluation arrangements are studied as combinations of evaluation objects, assignments and practices. Cases are the evaluation of the wage subsidy experiments, of experiments with flat-rate payments for irregularly required benefits in social assistance and of the competition between two administrative models of jobcenters. Questions are asked about “varieties of evaluation” (Schwab, 2017, S. 201) and the connection between the institutional framework and the design of evaluation arrangements. In addition, “politics of evaluation” (Taylor & Ballloch, 2005) visible in the comparison of the three examples are discussed.

**Keywords:** politics of evaluation; policy experiments; labour market policy; public administration; advocacy coalitions

## 1 Einleitung<sup>1</sup>

Die Reformen der Arbeitsmarktpolitik der letzten 20 Jahre wurden in einer Reihe von Themenfeldern von politischen Experimenten eingeleitet oder durch solche Experimente überhaupt erst ermöglicht. Diese Experimente wurden anlassbezogen evaluiert und zumindest formal als Instrument zur Evidenzbasierung von politischen Entscheidungen eingesetzt.

Am Beispiel von drei Evaluationen experimenteller Politikgestaltung unterschiedlicher instrumenteller und institutioneller Tragweite im Feld der Arbeitsmarktpolitik – der Evaluation verschiedener Kombilohn-Experimente (1998-2005), der Experimente zur Pauschalierung einmaliger Leistungen in der Sozialhilfe (1999-2004) und des Experiments zur Erprobung alternativer Modelle der Eingliederung von Arbeitsuchenden im Wettbewerb zwischen Optionskommunen und Arbeitsagenturen in Arbeitsgemeinschaften mit Kommunen (2005-2008) – werden hier Antworten auf die folgenden Fragen skizziert:

- Lassen sich für diese in ihrer institutionellen Reichweite unterscheidbaren Beispiele experimenteller Politik distinkte Evaluierungsarrangements identifizieren?
- Lässt sich dabei ein Zusammenhang zwischen den institutionellen Rahmenbedingungen und der Gestaltung von Evaluierungsarrangements erkennen?
- Welche politischen Deutungskämpfe wurden im Prozess der Evaluierung geführt?

Damit möchte der Artikel im Anschluss an Oliver Schwab (2017) Evaluierungsarrangements und deren Beeinflussung durch ihren politisch-administrativen Kontext systematisch untersuchen und damit einen Beitrag zum Verständnis von „varieties of evaluation“ (Schwab, 2017) leisten.

Evaluierungsarrangements werden hier verstanden als Kombinationen von Evaluierungsgegenständen, Evaluierungsaufträgen und Evaluierungspraktiken, die durch institutionell eingebettete Advokaten-Koalitionen (Sabatier, 1988) gestaltet werden. Es wird davon ausgegangen, dass diese Arrangements unterscheidbar sind und sich die Unterschiede zwischen ihnen, die hier als „varieties of evaluation“ verstanden werden, sowohl aus unterschiedlichen institutionellen Rahmenbedingungen als auch aus unterschiedlichen Interessen der sie prägenden Advokaten-Koalitionen erklären lassen.

## 2 Drei Beispiele experimenteller Arbeitsmarktpolitik und ihrer Evaluierung

### 2.1 Kombilohn-Experimente (1998 bis 2005)

In den 1990er Jahren hatte sich angesichts anhaltend hoher Arbeitslosigkeit eine intensive Debatte darum entwickelt, wie neue Beschäftigungsmöglichkeiten auch unterhalb tariflicher Entlohnung geschaffen werden könnten.

Mikroökonomisch orientierte Marktliberale argumentierten, die Zuverdienst-Regeln in der Sozial- und Arbeitslosenhilfe seien beschäftigungsbehindernd ausgestaltet, weshalb von diesen Leistungen abhängige Arbeitslose in einer Anreizfalle säßen (Boss, 1994; Siebert, 1996). Diese hindere die Menschen daran, existierende Niedriglohnbeschäftigung auszuüben. Deshalb müsse ein sogenannter Lohnabstand gewährleisten, dass auch mit gering entlohnter Tätigkeit mehr Einkommen erzielt werde als ausschließlich mit sozialen Sicherungsleistungen. Dieser Abstand könne entweder durch eine allgemeine Reduktion der Fürsorgeleistungen (Siebert, 1996) oder durch eine gezielte Subventionierung gering entlohnter Tätigkeiten (Spermann, 1996) erzielt werden.

Vertreter einer makroökonomischen Perspektive argumentierten hingegen, dass nicht ausreichend Beschäftigungsmöglichkeiten für gering qualifizierte Personen zur



Verfügung stünden. Deshalb sollten Löhne für Einfach Tätigkeiten flächendeckend subventioniert werden, damit die Lohnkosten für Unternehmen sinken und diese mehr einfache Arbeitsplätze anbieten. Dies sollte entweder über eine niedrigere Abgabenlast für gering entlohnte Beschäftigung (Scharpf, 1997) oder über zielgruppenorientierte Kombilöhne bzw. Lohnkostenzuschüsse erfolgen (Gerster & Deubel, 2001).<sup>2</sup>

Die CDU/CSU wollte im Wahlkampf 1994 „die Sozialhilfe bei denen, für die eine mögliche Erwerbstätigkeit zumutbar ist, ab[zu]senken und ein erzielttes Einkommen dafür nur teilweise auf die Sozialhilfe an[zu]rechnen“ (CDU-Bundesgeschäftsstelle, 1994, S. 17 f.). Die SPD hingegen wollte mit degressiv gestaffelten Lohnkostenzuschüssen an Arbeitgeber neue Beschäftigung für Langzeitarbeitslose am ersten Arbeitsmarkt schaffen, ohne die Tarifstrukturen oder die sozialen Sicherungssysteme zu verändern (SPD-Parteivorstand, 1994, S. 15). Im Jahr 1996 wurden mit dem „Gesetz zur Reform des Sozialhilferechts“ Lohnkostenzuschüsse an Arbeitgeber und Eingliederungszuschüsse an Arbeitnehmende als neue Instrumente in das Bundessozialhilfegesetz (BSHG) eingeführt, die aber zunächst kaum genutzt wurden.

Deshalb beschlossen die konservativ-liberalen Regierungsfractionen und die oppositionelle SPD 1998, eine Experimentierklausel in §18 Abs. 5 BSHG einzuführen, die den Sozialhilfeträgern erlaubte, „im Einzelfall oder durch eine Modellmaßnahme, (...) von den die Dauer oder die Höhe des [Eingliederungs-]Zuschusses begrenzenden Regelungen (...) abzuweichen“ (Deutscher Bundestag, 1998, S. 10). Die Ergebnisse dieser Experimente sollten durch eine Auswertung unter Beteiligung der Länder gesichert werden (ebd., S. 7). Diese Experimentierklausel hatte instrumentellen Charakter. Sie sollte mit der Öffnung der Regelungen zum Instrument des Eingliederungszuschusses in der Sozialhilfe neue Arbeitsanreize setzen.

Verschiedene Bundesländer entwickelten daraufhin eigene Kombilohn-Programme zur Förderung der Beschäftigung von Geringqualifizierten und Langzeitarbeitslosen, allerdings teilweise nicht oder nicht nur für Sozialhilfeempfangende, sondern auch für Langzeitarbeitslose, die Arbeitslosenhilfe erhielten und über die damaligen Arbeitsämter gefördert wurden. Auch einzelne Kommunen entwickelten eigene Fördermodelle über die Sozialhilfe. Bei aller Unterschiedlichkeit im Detail (Blancke, 2004, S. 112 ff.; Kaltenborn, 2001, S. 30 f.) setzten alle Experimente bei der Subventionierung von Niedriglohntätigkeiten an, einzelne alternativ dazu oder darüber hinaus auch an den Regeln zur Anrechnung von Erwerbseinkommen auf die Grundsicherungsleistungen.

Nach dem Regierungswechsel 1998 ließ das vom Bundeskanzleramt initiierte „Bündnis für Arbeit, Ausbildung und Wettbewerbsfähigkeit“ seine „Arbeitsgruppe Benchmarking“ parallel zu den laufenden Experimenten unter relativ großer Aufmerksamkeit einer breiten Fachöffentlichkeit eine Studie zu den Möglichkeiten der Förderung von Niedriglohnbeschäftigung erstellen, für die mehrere Forschungsinstitute mit Modellrechnungen Effekte einer allgemeinen Senkung der Lohnnebenkosten für gering qualifizierte Tätigkeiten schätzten (Bender & Rudolph, 1999; Schupp, Volz, Wagner & Zwiener, 1999). Die Ergebnisse dieser Simulationsrechnungen waren uneindeutig (Fels, Heinze, Pfarr & Streeck, 1999, S. 57 ff.).

Das Bündnis für Arbeit, Ausbildung und Wettbewerbsfähigkeit beschloss daraufhin, verschiedene Kombilohn-Modelle auf Bundesebene zu testen (Andres, 2000, S. 44 f.). Unter den vorhandenen Konzepten fiel die Wahl auf das Mainzer Modell und das SGI-Modell des Saarlandes – nach Ansicht von Susanne Blancke aufgrund der besonders aktiven und frühzeitigen Lobbyarbeit der jeweiligen Landesregierungen beim

Bundeskanzleramt (Blancke, 2004, S. 136 ff.).<sup>3</sup> Unter dem Namen „Chancen und Anreize zur Aufnahme sozialversicherungspflichtiger Tätigkeiten (CAST)“ wurden diese beiden Modelle ab Mitte 2000 in mehreren Regionen erprobt und parallel die Experimentierbedingungen in den beiden laufenden Landes-Modellprojekten im Sinne des CAST-Programmes verändert. Auch die Evaluationen mussten sich diesen Veränderungen anpassen.

Nahezu alle dezentralen Kombilohn-Experimente wurden zwar dezentral evaluiert, wegen des dezentralen Evaluierungsarrangements entstanden aber trotz vergleichbarer Evaluationsgegenstände weder ein einheitliches methodisches Evaluationsdesign noch einheitliche Evaluationskriterien. Die Mehrzahl der Evaluierungen nutzte mikroökonometrische Analysen mit einem Kontrollgruppendesign oder Vorher-Nachher-Vergleiche, die teilweise mit einer Nutzerbefragung oder einer Implementationsanalyse ergänzt wurden. Dabei waren die Evaluatoren davon abhängig, welche Daten die Arbeitsämter oder Sozialämter zur Verfügung stellen und welche sie selbst generieren konnten.

Auch wenn am Ende unter dem Dach der zentralen Evaluierung des Bundesmodellprojekts die Evaluationsergebnisse aus den dezentralen Projekten zusammengestellt wurden (Kaltenborn, 2001), kann von einem einheitlichen, an systematisch vergleichbaren Kriterien orientierten Evaluationsdesign für die verschiedenen erprobten Ansätze keine Rede sein. Eher könnte man argumentieren, dass die Belastbarkeit der Evaluationsergebnisse im folgenden politischen Prozess auch aufgrund der eingesetzten Evaluationsmethodik in unterscheidbarem Maß anerkannt wurde. Im Unterschied zu den dezentralen Evaluierungen wurde das Bundesprogramm CAST zentral von einem Konsortium aus dem Forschungsinstitut der Bundesagentur für Arbeit (IAB) und anderen Forschungs- und Befragungsinstituten evaluiert. Dieses Konsortium konnte, neben eigens durchgeführten Nutzer- und Betriebsbefragungen auf zentral in der Bundesagentur für Arbeit (BA) gehaltene bzw. speziell für das CAST-Programm generierte Prozessdatensätze zugreifen und deshalb zur Evaluation des zentralen Bundesprogramms einen komplexen Methodenmix einsetzen (Kaltenborn, Krug, Rudolph, Weinkopf & Wiedemann, 2005, S. 13).

Die umfangreichen Evaluationen aller Landesprogramme und des Bundesprogramms ergaben nur selten positive Netto-Effekte der Förderung von Kombilöhnen für die Integration in den Arbeitsmarkt, teilweise wurden auch deutlich negative Effekte ermittelt. In allen Experimenten zeigte sich, dass fehlende Arbeitsanreize bzw. die „Faulheit der Arbeitslosen“ nicht die zentrale Ursache hoher Arbeitslosigkeit darstellten. Stattdessen offenbarten sich andere Kernprobleme: Fehlende Arbeitsplätze im Bereich von Einfachtätigkeiten, mangelnde Beschäftigungsfähigkeit eines Teils der Arbeitslosen und der hohe vermittelnde Aufwand, um Angebot und Nachfrage im Bereich von Einfachtätigkeiten zusammenzubringen. Für diese Probleme boten Kombilöhne allein keine Lösung.

Das SGI-Modell wurde wegen geringer Nutzungszahlen vorzeitig abgebrochen, das Mainzer Modell wurde ebenfalls vorzeitig abgebrochen (Kaltenborn, Krug, Rudolph, Weinkopf & Wiedemann, 2005, S. 12), denn die Neuregulierung des Niedriglohnssektors war im Jahr 2002 zur Verhandlungsmasse um das „Zweite Gesetz für moderne Dienstleistungen am Arbeitsmarkt“ geworden. Damit war auch die umfangreiche und methodisch anspruchsvolle Evaluierung des zentralen Bundesprogramms, obwohl umfangreiche Daten zur Verfügung standen, durch Diskontinuität der Experimentier-

bedingungen erschwert. Die Evaluationsergebnisse spielten wegen paralleler politischer Aktivitäten zur Einführung von Mini- und Midi-Jobs keine Rolle für den folgenden politischen Entscheidungsprozess.

Sowohl im politischen Prozess der Entwicklung der verschiedenen Kombilohn-Experimente als auch bei der Auswahl der im Bundesmodellprojekt zu erprobenden Varianten zeigten sich verschiedene Advokaten-Koalitionen (Sabatier, 1988), in denen sich Akteure mit jeweils geteilten *belief systems* unter dem Schirm innovativen Experimentierens zusammenfanden: unter anderem Landespolitiker, die eigene Modelle entwickelten und Wirtschaftswissenschaftler, die der Landespolitik eigene Modelle andienten und diese dann für die Länder evaluierten. Sie lassen sich in eine mikroökonomisch und eine makroökonomisch orientierte Advokaten-Koalition unterscheiden. Während die ersten argumentierten, fehlende Arbeitsanreize seien das Kernproblem für die Reintegration von Langzeitarbeitslosen und entsprechend einen Fokus auf verbesserte Arbeitsanreize legten, argumentierten die Vertreter der zweiten Position eher, es mangle an Beschäftigungsmöglichkeiten im Niedriglohnbereich oder an passenden institutionellen Rahmenbedingungen. Entsprechend entwickelten sie eher Kombilohn-Konzepte, die auf das Angebot von Einfacharbeitsplätzen oder deren Rahmenbedingungen Einfluss nehmen sollten.

In den politischen Deutungskämpfen beurteilte jede Advokaten-Koalition ihre eigenen Projekte als erfolgreich oder begründete geringe Erfolge mit fehlerhafter Umsetzung ihrer Vorschläge durch die Arbeits- oder Sozialverwaltung. Immerhin aber drangen über die Zeit die zentralen Erkenntnisse der Evaluierungen der verschiedenen Kombilohn-Experimente als Wissensbestände in beide Advokaten-Koalitionen und in die breite Fach-Community vor, so dass in den letzten Jahren eine weitere Verstärkung von Arbeitsanreizen nicht mehr als zentraler Lösungsansatz zur Bekämpfung von Langzeitarbeitslosigkeit gilt. Das breite Spektrum der erprobten Kombilohn-Konzepte hat dazu sicher beigetragen. Auffällig ist aber zugleich, dass eine zentrale Erkenntnis der Kombilohn-Experimente über die Zeit verschüttet wurde und erst in der wesentlich späteren Evaluierung des Bundes-ESF-Programms zur Bekämpfung der Langzeitarbeitslosigkeit wiederentdeckt werden musste: Dass die Erschließung und Sicherung von Arbeitsplätzen für Langzeitarbeitslose eine aktive Gestaltung der konkreten individuellen Arbeitsbedingungen im Betrieb durch bewerberorientierte Vermittlung und durch ein begleitendes Coaching benötigt (Boockmann et al., 2019).

## 2.2 Experimente zur Pauschalierung der Sozialhilfe (1998 bis 2004)

In den 1980er und 1990er Jahren wurden diverse Vorschläge zur fundamentalen Änderung des Grundsicherungssystems vorgelegt (Kaltenborn, 2000; Lessenich, 2009). Alle Vorschläge, die auf Einführung eines bedingungslosen Grundeinkommens oder einer negativen Einkommensteuer zielten und mit einer Ausgabensteigerung in den öffentlichen Haushalten verbunden waren, hatten aus haushaltspolitischen Gründen keine Umsetzungschance. Angesichts der Kostenentwicklung und einer ausbleibenden Neugestaltung der Regelsätze der Sozialhilfe entwickelte sich parallel zu diesen Forderungen eine Debatte darum, wie innerhalb der bestehenden Sozialhilfe durch eine Pauschalierung bestimmter einmaliger Leistungen über Verwaltungsvereinfachung Kosten gesenkt werden könnten.

Der Landkreistag Baden-Württemberg forderte, wie mehrere konservativ regierte Bundesländer, schließlich, alle „laufenden Leistungen der Hilfe zum Lebensunterhalt (Regelsatz, Mehrbedarf, Kosten der Unterkunft, Kosten der Heizung, Nebenkosten der Unterkunft und die einmaligen Beihilfen für Bekleidung, Hausrat, persönliche Anlässe usw.) zu einem Gesamtbetrag zusammenzufassen, der in der Regel monatlich ausbezahlt wird. Mit diesen monatlichen Gesamtpauschalen“, so die Argumentation, „sollten die Sozialhilfeempfängerhaushalte auskommen, so wie jeder Arbeiter, Angestellte oder Beamte mit dem ihm monatlich nach Abzug der Steuern und Sozialversicherungsabgaben verbleibenden Betrag den Lebensunterhalt für sich und seine Familie zu bestreiten hat“ (Klinger, 2000, S. 127 f.). Die Vertreter dieser Advokaten-Koalition argumentierten eher paternalistisch, haushaltspolitisch und leistungsorientiert im Sinne konservativer fürsorgischer Ansätze.

Die von SPD und Bündnis 90/Die Grünen geführte Bundesregierung verabredete im Koalitionsvertrag 1998, „für einen Modellversuch zur Pauschalierung von Sozialhilfeleistungen (...) unter Beachtung des Prinzips der Bedarfsdeckung und der Freiwilligkeit die gesetzlichen Grundlagen“ zu schaffen und parallel dazu „ein Konzept für eine bedarfsorientierte soziale Grundsicherung [zu] entwickeln, das schrittweise eingeführt werden“ sollte (SPD & Bündnis 90/Die Grünen, 1998, S. 30). Teile beider Parteien wie auch Wohlfahrtsverbände und Sozialrechtler standen der Pauschalierung aus soziallibertärer Haltung skeptisch gegenüber, weil sie darüber eine schleichende Leistungskürzung befürchteten, auch wenn ein Potential für größere Handlungsautonomie der Leistungsberechtigten gesehen wurde, wenn nicht mehr für jede Kleinigkeit ein zusätzlicher Antrag zu stellen wäre.

Nach einigem Hin und Her über das Wie einigten sich Regierung und Opposition, eine entsprechende Experimentierklausel in § 101a Bundessozialhilfegesetz einzuführen. Beobachter argumentierten, das damalige Scheitern, eine Regelsatzverordnung für das BSHG zu erlassen, sei der Auslöser für diesen Weg gewesen (Armborst, 2003, S. 1152). Die soziallibertäre Advokaten-Koalition wurde mit dem Argument zur Zustimmung bewegt, dass eine Pauschalierung der Sozialhilfe die Handlungsautonomie der Leistungsberechtigten erhöhen würde und die eingesparten Verwaltungskosten in eine bessere Ausstiegsberatung investiert werden könnten.

Mit dieser Experimentierklausel sollte „zur Weiterentwicklung der Sozialhilfe [...] die Pauschalierung weiterer Leistungen [...] erprobt werden“. Die Länder sollten die Kommunen als Träger der Sozialhilfe dazu per „Rechtsverordnung ermächtigen, in Modellvorhaben solche Leistungen der Sozialhilfe pauschaliert zu erbringen, für die Beträge nicht schon durch [das BSHG] festgesetzt oder auf Grund [des BSHG] festzusetzen sind. Die Pauschalbeträge“ waren dabei „für einen bestimmten Bedarf festzusetzen und“ mussten „dem Grundsatz der Bedarfsdeckung gerecht werden. Die Modellvorhaben“ sollten so ausgewertet werden, dass „sie eine bundesweite Bewertung zulassen“. Dazu hatten die Kommunen, die zuständigen Landesministerien und das Bundesministerium für Arbeit und Soziales (BMAS) zusammenzuarbeiten (Deutscher Bundestag, 1999, S. 5).

Die Experimentierklausel hatte materiellen Charakter und war, anders als die Kombilohn-Experimente, ausschließlich auf die Sozialhilfe konzentriert. Sie sollte mit der Pauschalierung von Teilen des materiellen Existenzminimums einerseits den Aufwand einer „Massenverwaltung“ (Armborst, 2003, S. 1151) kleinteiliger einmaliger Leistungen reduzieren und je nach politischer Position andererseits das „Einzelan-

spruchsdenken“ (Klinger, 2000, S. 127) der Leistungsberechtigten zurückdrängen oder zu „mehr Autonomie für die Sozialhilfeempfänger“ (Deutscher Bundestag, 1999, S. 7) beitragen.

Zwölf Bundesländer erließen Verordnungen zur Umsetzung solcher Pauschalierungsexperimente (Deutscher Bundestag, 2001, S. 1). Diese überließen die Ausgestaltung der Pauschalierung und die Festlegung des Teilnehmerkreises weitgehend den Kommunen und regelten selbst nicht Näheres zur konkreten Ausgestaltung der Modellprojekte. In einigen Ländern sollten über die lokalen Voraussetzungen Einvernehmen mit dem zuständigen Landesministerium hergestellt bzw. lokale Beiräte in die Ausgestaltung der Experimente einbezogen werden (GV NRW 2000, S. 250, Gesetzblatt Baden-Württemberg 2000, Nr. 9, S. 433 ff. und Hessisches GVBl. I, S. 528).

Die Landesverordnungen eröffneten einen weiten lokalen Handlungsspielraum, den rund 40 Kommunen bundesweit sehr unterschiedlich nutzten (Felder, 2003). Die kommunalen Experimente wurden von teils heftigen, allerdings auf eine kleine Fach-Community begrenzten, sozialpolitischen und juristischen Debatten um die Rechtmäßigkeit der lokal getesteten Pauschalen, die Pflicht zur Teilnahme für die Leistungsberechtigten, die Informationspflichten der Kommunen und grundsätzlich um den Bedarfsdeckungsgrundsatz begleitet (Rothkegel, 2002; Maaß, 2001, S. 118 ff.). Mehrere lokale Experimente landeten vor Gericht. Das Bundesverwaltungsgericht hob Ende 2004 ein Urteil des Bayrischen Verwaltungsgerichts zur Pauschalierung von Bekleidungsbeihilfen in München auf. Die Stadt hatte „den Betroffenen den Text der Ausführungsbestimmungen (...) nicht bekannt gegeben“. Dem Rechtsstaatsprinzip sei aber „nur dann Genüge getan, wenn der Betroffene unmittelbar Kenntnis von den Bestimmungen selbst nehmen kann. Nur dann kann er diese auf ihre Rechtmäßigkeit und Anwendbarkeit überprüfen und sich des Inhalts der durch sie für ihn begründeten Rechte und Pflichten vergewissern“, urteilte das Bundesverwaltungsgericht (2004, S. 12).

Nur zwei der zwölf Länder mit Landesverordnung ließen die Experimente in ihrem Land evaluieren (Nothbaum, Lübker & Kämper, 2004; Univation e.V. & Dr. Beywl & Associates, 2005). Vereinzelt wurden daneben lokale Evaluationen durchgeführt. Für diese dezentralen Evaluierungen ist nur teilweise rekonstruierbar, mit welchen Methoden und durch wen sie durchgeführt wurden. Die bundesweite Meta-Evaluation, mit der das BMAS Mummert Consulting GmbH beauftragt hatte, wurde nie veröffentlicht.

Die dezentralen Evaluationen hatten alle mit der Schwierigkeit zu kämpfen, dass für die Sozialhilfe keine einheitlichen Forschungsdaten verfügbar waren. Die Kommunen nutzten unterschiedliche Prozesssoftware für die Administration der Sozialhilfe, wenn sie denn Ende der 1990er Jahre überhaupt schon auf digitale Aktenhaltung umgestellt hatten. Je nach Struktur der Software mussten die Evaluatoren deshalb die Daten für ihre Untersuchungen weitgehend selbst generieren oder mit jeder Kommune einzeln vereinbaren, welche Daten für die Evaluation nutzbar gemacht werden konnten. Unter anderem deshalb wurden hier keine mikroökonomischen Kontrollgruppen-Designs eingesetzt. Zudem fehlt es, vor allem wegen der Nichtveröffentlichung der Meta-Evaluation, an einem Überblick darüber, für welche Teilleistungen der Sozialhilfe Pauschalierungen in welcher Höhe erprobt wurden, welche Teilnahmebedingungen an den Pauschalierungsexperimenten herrschten und mit welchen Methoden die einzelnen Evaluationen gearbeitet haben.

Das zersplitterte dezentrale Evaluierungsarrangement führte dazu, dass nicht alle Experimente evaluiert wurden und dass weder ein einheitliches Evaluationsdesign noch

einheitliche Evaluationskriterien entwickelt wurden. Die vom Bund beauftragte Meta-Evaluation setzte, anders als die zentrale Evaluation des CAST-Programms (und soweit das überhaupt rekonstruierbar ist), lediglich auf dezentralen Evaluierungen auf und führte nicht zu einem einheitlichen, an systematisch vergleichbaren Kriterien orientierten Evaluationsdesign. Die Frage, ob eine Pauschalierung bestimmter Sozialhilfeleistungen die Handlungsautonomie der Leistungsberechtigten stärkte, wurde nicht untersucht. Der Aufwand für die Antragsbearbeitung wurde nach Angaben aus einer Befragung in den Sozialämtern zwar reduziert, die Sachbearbeitenden beschrieben aber, dass sie dennoch nicht mehr Zeit hatten, die Leistungsberechtigten zu beraten (Felder, 2003, S. 9). Dies war aber ein zentrales Argument für die Pauschalierungs-Experimente gewesen. Die verfügbaren Informationen aus der Meta-Evaluation (Mummert Consulting, 2003; Felder, 2003) lassen nicht erkennen, ob die Zufriedenheits- und Machbarkeits-einschätzungen der Nutzer und der Sachbearbeitenden in einem Zusammenhang mit der Art und Höhe der lokalen Pauschalen in den 40 Kommunen standen, sie wurden lediglich über alle experimentierenden Kommunen aggregiert. Damit berücksichtigte die Meta-Evaluation die Grundidee der experimentellen Politigestaltung im Evaluationsdesign nicht, so dass ihre Ergebnisse auch wegen methodischer Schwächen wenig Aussagekraft hatten.

Trotz der fehlenden Transparenz über die Evaluationspraxis und die Evaluationsergebnisse wurde in den politischen Deutungskämpfen schon während des Evaluierungszeitraums mit ausgewählten Ergebnissen argumentiert. Soweit es sich aus der Analyse dieser Deutungskämpfe rekonstruieren lässt, hatten die Evaluationen ergeben, dass grundsätzlich die Pauschalierung von Leistungen der Sozialhilfe machbar sei. Allerdings lässt sich nicht erkennen, ob das für alle Arten erprobter Pauschalierungen gleichermaßen galt – was aber angesichts der kritischen gerichtlichen Einschätzungen zu den verfassungsrechtlichen Grenzen derartiger Experimente mit dem Existenzsicherungsniveau notwendig gewesen wäre. So wurde zwar berichtet, dass insgesamt etwa die Hälfte der teilnehmenden Leistungsberechtigten mit der lokalen Pauschalierungsregel zufrieden waren, aber unklar blieb, ob es einen Zusammenhang zwischen Zustimmung und Art oder Ausmaß der Pauschalierung oder veränderten Verwaltungsprozessen gab. Unter den Menschen mit Kindern und denen, die schon länger mit Leistungsbezug lebten, lag die Zufriedenheit zudem deutlich niedriger. Und immerhin 55 Prozent aller Befragten hatten angegeben, dass sie aus den Pauschalen in keinem Monat Rücklagen bilden konnten, um bei Bedarf notwendige Anschaffungen zu tätigen, was aber ein wesentlicher Sinn der Pauschalierung war. Zudem hatten die Kommunen keine Kosten gespart, einige von ihnen mit den Pauschalen sogar Mehrkosten verursacht (Mummert Consulting, 2003; Felder, 2003).

Während die Kombilohn-Experimente zumindest teilweise über die Arbeitsverwaltung administriert wurden, waren die Pauschalierungs-Experimente ausschließlich in Sozialämtern angesiedelt. Deshalb standen den Evaluationen unterschiedlich gute Daten zur Verfügung, weshalb unterschiedliche Evaluationsmethoden eingesetzt werden konnten bzw. mussten. Im Bereich der kommunalen Sozialverwaltung waren Evaluationen bis dahin zudem, mit Ausnahmen im Bereich kommunaler Beschäftigungspolitik (Kaps, 2006), noch weniger etabliert als in der Arbeitsverwaltung.

Die negativen Ergebnisse zur für die Mehrheit der befragten Sozialhilfehaushalte nicht ausreichenden Höhe der Pauschalen wurden aus dem parallel zu den Pauschalierungs-Experimenten begonnenen Prozess zur Neubestimmung der Regelsätze weit-

gehend herausgehalten. Die Vertreter der Bundes-, Landes- und Kommunalverwaltung argumentierten nun vor allem haushaltspolitisch, die Pauschalierung sei machbar und würde Verwaltungskosten sparen, ohne den Leistungsberechtigten materiell zu schaden.<sup>4</sup> Dass die Leistungsberechtigten aus den Pauschalen keine Ansparungen für größere Investitionen tätigen konnten, hätte hingegen zu einer Vergrößerung der Pauschalen und damit zu einem Anstieg der Kosten der Grundsicherung führen müssen, wie die Vertreter der nun deutlich verkleinerten soziallibertären Advokaten-Koalition forderten (Spindler, 2004; Martens, 2006).

Da die Evaluationsergebnisse nicht veröffentlicht wurden, konnte die letztere Advokaten-Koalition nicht öffentlichkeitswirksam darauf zurückgreifen. Zudem fanden diese Ergebnisse ohne Veröffentlichung keinen Eingang in den Wissensbestand der Fach-Community. Für den Auftraggeber der Meta-Evaluierung war es wegen des stark dezentralen Evaluationsarrangements und der gleichgerichteten haushaltspolitischen Interessen der Verwaltung aller Ebenen leicht, die Ergebnisse nur so weit zu veröffentlichen, wie es in die damals aktuelle Strategie zur Abschaffung der Arbeitslosenhilfe passte.

An diesem Beispiel lässt sich zeigen, dass, wie von Schwab (2017, S. 214) vermutet, eine „Evaluierung in Mehrebenenkonstellationen“, die „stark von Steuerungsinteressen [des Auftraggebers] motiviert ist“, „Auswirkungen auf die Ausgestaltung der Evaluierungsarrangements“ hat.

Zugleich wurde im Vergleich zum Kombilohn-Beispiel deutlich, dass eine breite Evaluations-Community und eine Institutionalisierung der Evaluationsfunktion (im ersten Beispiel beim IAB und beim Bündnis für Arbeit) einem politischen Interesse eines Auftraggebers, die Ergebnisse von Evaluationen intransparent zu halten und in politischen Deutungskämpfen selektiv für die eigenen Ziele zu nutzen, entgegenwirken kann. Hingegen gewinnt ein zentraler Auftraggeber in dezentralen Evaluierungsarrangements ohne institutionalisierte Evaluationsfunktion und entsprechende Community einen hohen Handlungsspielraum, die Ergebnisse von Evaluationen selektiv für eigene Steuerungsinteressen zu nutzen.

### 2.3 Experimente zum Wettbewerb zwischen zwei Modellen öffentlicher Aufgabenerbringung (2005 bis 2008)

Auch die Modernisierung der Leistungsverwaltung der Sozialhilfe und der Arbeitslosenhilfe wurde damals breit diskutiert. Dabei wurde als Ursache für den steigenden Anteil von Langzeitarbeitslosen und Sozialhilfeempfängern der so genannte Drehtüreffekt zwischen dem Einsatz der arbeitsmarktpolitischen Förderinstrumente von Sozialhilfe- und Arbeitsverwaltung verantwortlich gemacht, die Ineffizienz der doppelten Verwaltungsstrukturen für bedürftige Arbeitslose kritisiert und die unterschiedlichen Regeln von Arbeitslosen- und Sozialhilfe zum Fordern und Fördern zum zentralen institutionellen Hindernis für wirksame Arbeitsmarktintegrationen erklärt (Sell, 1999; Hoffmann, 2000).

Schon seit den 1980er Jahren ging es darum, die Zusammenarbeit zwischen Sozial- und Arbeitsämtern zu verbessern, um das gegliederte System kompatibler zu gestalten und die Fördermöglichkeiten beider Systeme für alle Leistungsberechtigten zugänglich zu machen. Dazu hatten die Bundesanstalt für Arbeit und die Kommunalen Spitzen-

verbände mehrfach Vereinbarungen zur Verbesserung der Verwaltungskooperation geschlossen. Seit 1993 hatten beide Seiten nach § 12 Arbeitsförderungsgesetz (AFG) zusammenzuwirken, aber die praktische Umsetzung zog sich hin.

In der Debatte um das „Gesetz zur Reform der Sozialhilfe“ 1995 wurden dann drei Advokaten-Koalitionen zu diesem Thema sichtbar: BA, Deutscher Städtetag, SPD, Bündnis 90/Die Grünen sowie die meisten Gewerkschaften wollten die Leistungsverwaltungen durch interne Verwaltungsmodernisierung verbessern und deren Wirksamkeit durch verbesserte Kooperation erhöhen. Der Deutsche Landkreistags, Deutsche Städte- und Gemeindebund sowie die CDU/CSU wollten die Aufgaben der BA für Langzeitarbeitslose kommunalisieren, weil die Kommunen „näher am Menschen“ seien als die lokalen Arbeitsämter. Arbeitgeberverbände, FDP und marktliberale Ökonomen wollten die Arbeitsvermittlung privatisieren und damit langfristig der BA allein die Aufgabe der Zahlung der materiellen Versicherungsleistungen zuschreiben oder sie gleich ganz auflösen und ihre sonstigen Aufgaben kommunalisieren.

Nach dem Regierungswechsel 1998 dynamisierte sich diese Debatte. Treiber waren die nun oppositionelle CDU und die Bertelsmann-Stiftung. Die CDU forderte sowohl eine bessere Zusammenarbeit beider Seiten als auch die Kommunalisierung der Aufgaben (Fink, 1998). Die Bertelsmann-Stiftung initiierte mehrere Projekte zur Verwaltungsmodernisierung in Kommunen nach dem Neuen Steuerungsmodell (Banner, 2008), unter anderem das Projekt „Beschäftigungsförderung in Kommunen (BiK)“, in dem neue Formen der Zusammenarbeit von Sozial- und Arbeitsämtern erprobt wurden (Empter & Frick, 1999), und das „Netzwerk Kommunen der Zukunft“, in dem Kommunen verschiedene Aspekte des Neuen Steuerungsmodells erprobten (Bertelsmann-Stiftung, Hans-Böckler-Stiftung & KGSt, 2002). Daneben tagte unter dem Dach der Bertelsmann-Stiftung ein Arbeitskreis „Reform der Arbeitslosenhilfe und Sozialhilfe“, in dem Vertreter der Exekutiven von Bund und Ländern mit Vertretern der BA, der kommunalen Spitzenverbände, von Gewerkschaften und Arbeitgeberverbänden, ausgewählten Praktikern der kommunalen Beschäftigungsförderung und Experten der Bertelsmann-Stiftung und der *con\_sens* GmbH zusammentrafen, um eine Reform von Arbeitslosen- und Sozialhilfe vorzubereiten (Schmachtenberg, 2003; Bertelsmann-Stiftung, 2002; Pautz, 2008).

Die Bundesregierung förderte dann zunächst unter großer Aufmerksamkeit der Fachöffentlichkeit Modellprojekte zur Verbesserung der Zusammenarbeit von Arbeits- und Sozialämtern (MoZArT-Projekte, Hess, Schröder, Smid & Reis, 2004). Roland Koch (CDU), damals Hessischer Ministerpräsident, dynamisierte die Debatte, indem er – inspiriert von Erfahrungen des Hessischen Partnerlandes Wisconsin in den USA mit dem *workfare*-Programm „Wisconsin works“ – lautstark die „Zusammenlegung“ von Arbeitslosen- und Sozialhilfe und die Kommunalisierung der dann neuen Grundsicherungsleistung forderte. Die CDU übernahm diese Position und forderte ein kommunal administriertes „Sozialgeld“ analog der Sozialhilfe (CDU-Bundesgeschäftsstelle, 2001, S. 115). Im Zuge des so genannten Vermittlungsskandals der BA wurden dann im Bundestagswahlkampf 2002 fundamentale Veränderungen des Grundsicherungssystems gefordert. Während die CDU das Grundsicherungssystem kommunalisieren wollte und darin von marktliberalen Ökonomen unterstützt wurde (Sinn, Holzner, Meister, Ochel & Werding, 2002), wollte die FDP die BA privatisieren (FDP-Bundesgeschäftsstelle, 2002, S. 6 f.).

Nun veröffentlichte der Arbeitskreis „Reform der Arbeitslosenhilfe und Sozialhilfe“ der Bertelsmann-Stiftung ein Positionspapier zur Neugestaltung des Grundsiche-



runge systems. „Kernstück der Reform von Arbeitslosen- und Sozialhilfe“, so formulierte dieses Papier, müsse „die Schaffung eines einheitlichen Systems aktivierender und passiver Leistungen für die Zielgruppe der erwerbsfähigen Erwerbslosen sein. Dieses einheitliche, steuerfinanzierte System aktivierender und passiver Leistungen soll zudem sicherstellen, dass die Leistungsgewährung ‚aus einer Hand‘ erfolgt und nicht in verschiedene behördliche Strukturen bzw. Zuständigkeiten aufgeteilt ist“ (Bertelsmann-Stiftung, 2002, S. 4). In diesem Papier wurde erstmals die enge Advokaten-Koalition zwischen verantwortlichen Experten der Ministerialverwaltungen von Bund und Ländern, den Kommunalen Spitzenverbänden und Beratungsunternehmen sichtbar, die später eine Reihe von Aufträgen in diesem Bereich generierten. Der politische Einfluss der hier vernetzten Akteure wird unter anderem daran erkennbar, dass sie mehrere Papiere in die Hartz-Kommission einspeisten (Hartmann, 2002; Kruse, 2002).

In dem Positionspapier wurde diskutiert, ein einheitliches Grundsicherungssystem in bundesunmittelbarer Trägerschaft durch die BA, in kommunaler Selbstverwaltung oder in Bundesauftragsverwaltung durch die Kommunen umzusetzen. Sichtbar wurde, dass alle drei Organisationsmodelle zur Leistungsgewährung „aus einer Hand“ massive Verschiebungen der Aufgaben- und Finanzverteilung zwischen Bund, Ländern und Kommunen bedeuteten, die Mehrheiten in Bundestag und Bundesrat oder einer Veränderung des Grundgesetzes bedurften. Lösungsvorschläge für diese institutionelle Herausforderung hatte der Arbeitskreis allerdings nicht entwickelt (Bertelsmann-Stiftung, 2002).

Weder die Hartz-Kommission noch die Gemeindefinanzreform-Kommission legten einen konsensfähigen Organisationsvorschlag zur „Zusammenführung“ von Arbeitslosen- und Sozialhilfe vor (Kaps, 2008). Immerhin wurden in den Verhandlungen der letzteren die mit den verbliebenen Organisationsvarianten – Zentralisierung bei der BA in Bundesauftragsverwaltung oder Kommunalisierung in Selbstverwaltung – verbundenen finanziellen Folgen als der zentrale haushaltspolitische Streitpunkt sichtbar (AG Arbeitslosenhilfe/Sozialhilfe der Kommission zur Reform der Gemeindefinanzen, 2003).

Die rot-grüne Bundesregierung entschied sich in ihrem Entwurf für das „Vierte Gesetz für moderne Dienstleistungen am Arbeitsmarkt“ für die Zentralisierungsvariante mit der BA als Leistungsträgerin (Deutscher Bundestag, 2003). Die CDU/CSU-Opposition hielt mit dem Entwurf für ein „Existenzgrundlagen-Gesetz“ dagegen, das die Kommunalisierung der Aufgaben vorsah (Bundesrat 2003).

Letztlich einigten sich Bund und Länder im Vermittlungsausschuss auf einen Vorschlag des Deutschen Städtetags und des Deutschen Städte- und Gemeindebundes zu einer geteilten Leistungsträgerschaft als Ausweg aus dem haushaltspolitischen Verhandlungspatt (Wienand, 2003). Die Kommunen sollten einen Teil der Leistungen des neuen SGB II in eigene Trägerschaft übernehmen. Roland Koch forderte im Gegenzug für Kommunen ein Optionsrecht auf eigenverantwortliche Trägerschaft. „Nachts um halb Drei“ wurde schließlich „ein nicht ausformulierter Kompromissvorschlag von den Beteiligten zu dem vermeintlichen Königsweg erhoben“. Genau das aber hätten „eigenem Bekunden nach alle Beteiligten vor Beginn des Vermittlungsverfahrens unbedingt vermeiden wollen“ (Henneke, 2004, S. 7).

Der folgende Entwurf zum „Kommunalen Optionsgesetz“ sah vor, die optierenden Kommunen über eine Organleihe mit den Leistungen zu beauftragen und sie damit formal der Bundesagentur für Arbeit zuzuordnen. Dagegen argumentierten alle Kom-

munalen Spitzenverbände, die kommunale Selbstverwaltung würde gefährdet. Sie kündigten an, unter dieser Bedingung würden kaum Kommunen eine solche Option ziehen. Damit hätte die BA die Hauptlast der Einführung des SGB II tragen müssen, die das zu diesem Zeitpunkt aber nicht wollte, um den Prozess der Verwaltungsreform in den Arbeitsagenturen nicht zu gefährden (Deutscher Bundestag, 2004, S. 10 ff.).

Im Vermittlungsausschuss einigten sich beide Seiten schließlich darauf, eine begrenzte Zahl von Jobcentern in kommunaler Trägerschaft aller Leistungen des SGB II für einen begrenzten Zeitraum zuzulassen, wobei „die Erprobung (...) insbesondere auf alternative Modelle der Eingliederung von Arbeitsuchenden im Wettbewerb zu den Eingliederungsmaßnahmen der Agenturen für Arbeit ausgerichtet“ werden sollte (§ 6a SGB II i. d. F. von 2005). Letztlich wurde so aus einem Konflikt zwischen rot-grüner Bundesregierung und den konservativ regierten Ländern einerseits und zwischen Landkreisen und kreisfreien Städten andererseits um die Trägerschaft und die Finanzierung der neuen Grundsicherungsleistung ein „konstitutionelles Experiment“ zwischen zwei Organisationsmodellen öffentlicher Leistungsverwaltung, das „existenzielle Eigeninteressen der beteiligten Akteure auf allen Ebenen“ berührte (Kaps, 2009, S. 612 f.).

Die nach § 6b SGB II zugelassenen kommunalen Träger (sog. Optionskommunen) und die nach § 44b SGB II gegründeten Arbeitsgemeinschaften aus Arbeitsagentur und kommunalem Träger (ARGen) begannen Anfang 2005 zu arbeiten. Im Sommer 2005 startete die in § 6c SGB II (a. F.) verankerte und groß angelegte zentrale Evaluierung der beiden Modelle der Aufgabenwahrnehmung im Auftrag des BMAS, an der in vier Konsortien insgesamt zwölf Forschungsinstitute beteiligt waren. Das BMAS ließ sich dabei von einem weiteren Forschungsinstitut administrativ unterstützen und wissenschaftlich beraten (Institut für Sozialforschung und Gesellschaftspolitik, 2006).

Die Evaluierung des Wettbewerbs zwischen den beiden Modellen der Leistungsträgerschaft war von Beginn an von politischen Deutungskämpfen begleitet, die deutlich vom Wettbewerb der Modelle geprägt waren. Der Deutsche Landkreistag (DLT) beauftragte im Sommer 2005 im Vorgriff auf die Ergebnisse der gesetzlichen Evaluierung nach § 6c SGB II ein anderes Institut mit einer eigenen Evaluationsstudie für seine Mitgliedskommunen, mit der er auf Vorrat Argumente sammelte, um sich bei Bedarf gegen die Ergebnisse der vom BMAS in Abstimmung mit den Ländern eingeleiteten gesetzlichen Evaluation stellen zu können (Hesse, Götz, Rüdiger & Schubert, 2008). Und er beauftragte die Bertelsmann-Stiftung und später die con\_sens GmbH damit, ein Benchmarking für die Optionskommunen zu entwickeln, mit dem er immer wieder an die Öffentlichkeit ging und für das Optionsmodell warb (Bertelsmann-Stiftung, 2007; Hartmann, Ahlrichs & Böttcher, 2008). Die Optionskommunen stellten zudem wiederholt die Kompetenz und Neutralität der BA in Frage, die gesetzlich mit der gesamten Statistik zum SGB II beauftragt ist, weil sie Nachteile im Wettbewerb der Modelle fürchteten, wenn die BA als Konkurrent die Hoheit für die Datengenerierung habe. Zur Umsetzung des Statistik-Auftrags der BA wurde deshalb ein Expertenkreis Statistik gebildet, in dem BA, BMAS, zuständige oberste Landesbehörden, kommunale Spitzenverbände, Statistisches Bundesamt und die Statistischen Landesämter vertreten waren und in dem zentrale Verhandlungen um Indikatoren und Kennzahlensets zum Controlling und Benchmarking im SGB II geführt wurden.

Die BA hielt sich mit Öffentlichkeitsarbeit zu den Arbeitsgemeinschaften eher zurück, wirkte allerdings nach innen in deren Handlungsspielräume deutlich steuernd hinein. Deshalb bildeten sich Landesarbeitsgemeinschaften der ARGE-Geschäftsführer

und ein „Bundesnetzwerk Arge SGB II“, um beim Bund die Interessen der ARGEn als eigenständige Grundsicherungsträger (auch gegen die Position der BA-Zentrale) wahrzunehmen und so ihre Chancen im Wettbewerb mit den Optionskommunen zu wahren (Bundesnetzwerk Arge SGB II, 2008).

Alle bis Mitte 2008 laufenden Evaluationsstudien zum Wettbewerb zwischen Arbeitsgemeinschaften und Optionskommunen wurden veröffentlicht. Die gesetzliche Evaluierung wurde durch mehrere Fachtagungen begleitet, in denen die Evaluierenden ihre Zwischenergebnisse vor Optionskommunen und Arbeitsgemeinschaften, Ländervertretern, kommunalen Spitzenverbänden und der BA zur Diskussion stellten. Die Evaluation ergab, dass die Unterschiede in der Umsetzung des SGB II zwischen Arbeitsgemeinschaften und Optionskommunen nicht gravierend waren. Insgesamt hatten Leistungsbererechtigte in ARGEn eine um rund vier Prozentpunkte höhere Wahrscheinlichkeit, innerhalb von zwölf Monaten nicht mehr auf SGB II-Leistungen angewiesen zu sein als in Optionskommunen. Bei der Nachhaltigkeit der Integrationen schnitten die ARGEn um rund drei Prozentpunkte besser ab als die Optionskommunen (ZEW, IAQ & TNS emnid, 2008, S. 132). Die ARGEn waren etwas erfolgreicher, Männer in Arbeit zu vermitteln (ebd., S. 137 ff.), die Optionskommunen hingegen waren etwas erfolgreicher, Frauen in Arbeit zu vermitteln (ebd., S. 211 ff.). Eine eindeutige Erklärung für diese geringfügigen Unterschiede fand die Evaluation nicht. Die umfangreichen Informationen aus der zentralen Evaluation flossen nur teilweise in die Wissensbestände der Fach-Community ein.

Als das Bundesverfassungsgericht Ende 2007 die Arbeitsgemeinschaften für nicht vereinbar mit dem Grundgesetz erklärte und bis Ende Dezember 2010 eine grundlegende Neuorganisation der Grundsicherung für Arbeitsuchende forderte, änderte dies die institutionellen Rahmenbedingungen des Modellwettbewerbes grundlegend. Das Gericht sah zwei Lösungsmöglichkeiten (Bundesverfassungsgericht 2007, Abs. 172): Entweder der Bund führt das Gesetz nach Artikel 87 Grundgesetz mit einer eigenständigen Bundesbehörde aus<sup>5</sup> oder der Gesamtvollzug wird nach Artikel 83 Grundgesetz Länderangelegenheit. Eine geteilte Leistungsträgerschaft war demnach nicht möglich. Während die Evaluierung noch lief, entwickelten BMAS und BA das Modell eines „Kooperativen Jobcenters“, eine Bund-Länder-Arbeitsgruppe entwickelte ein Modell für „Zentren für Arbeit & Soziales“ (Kaps, 2008, S. 414).

Der zentrale Ergebnisbericht wurde zwar veröffentlicht, der Bundestag beschäftigte sich aber nicht damit und die Evaluierungsergebnisse fanden bei der Neugestaltung der Organisationsmodelle kaum Berücksichtigung – auch nicht die der Parallel-Evaluierung im Auftrag des DLT. Wesentlicher wurde nun, juristisch wasserdichte Lösungen für die Organisation der Jobcenter zu entwickeln, die auch den Normen des Grundgesetzes entsprachen.

Über einen längeren Zeitraum schien eine Einigung nicht möglich. Erst nach einem Regierungswechsel und weiteren zähen Verhandlungen zwischen den Advokaten-Koalitionen der Kommunalisierungsbefürworter und der Zentralisierungsbefürworter wurde mit dem „Gesetz zur Weiterentwicklung der Organisation der Grundsicherung für Arbeitsuchende“ im August 2010 die heutige Organisation der Grundsicherung für Erwerbsuchende installiert (Deutscher Bundestag 2010). Die Ergebnisse der Evaluierung spielten dabei keine Rolle.

Im Vergleich dieser Evaluierung mit den zuvor erörterten Beispielen sticht hervor, dass hier der Evaluationsauftrag zentralisiert war, mit erheblichen Ressourcen und der Beauftragung einer breiten Fach-Community einherging und von Beginn an unter brei-

ter Beobachtung durch eine große Zahl interessierter Akteure stand, die immer wieder auch ihre Positionen in die Medienöffentlichkeit trugen.

### 3 Vergleich der Evaluierungsarrangements und ihrer Rahmenbedingungen

#### 3.1 Vergleich zentraler institutionellen Rahmenbedingungen der Evaluierungen

Vergleicht man die institutionellen Rahmenbedingungen der Evaluierungen dieser drei Beispiele experimenteller Politikgestaltung, so lassen sich einige grundlegende Unterschiede konstatieren.

Während die beiden ersten Beispiele im Bereich der kommunalen Selbstverwaltung angesiedelt waren und durch Landesrecht konkretisiert wurden, war das dritte Beispiel auf der Ebene des Bundes angesiedelt. In allen drei Beispielen sah das Bundesgesetz einen Evaluierungsauftrag vor. Je stärker dezentralisiert der Evaluierungsgegenstand angesiedelt war, desto dezentraler war auch der gesetzliche Evaluierungsauftrag gestaltet. Im ersten Beispiel sollten die kommunal erprobten Ansätze „unter Beteiligung des [jeweiligen] Landes“ ausgewertet werden. Man könnte hier von einem zweidimensionalen *bottom-up*-Auftrag sprechen. Im zweiten Beispiel sollten die Ergebnisse unter Zusammenarbeit von Kommunen, Ländern und dem zuständigen Bundesministerium so ausgewertet werden, dass eine bundesweite Bewertung möglich würde (dreidimensionaler *bottom-up*-Auftrag). Im dritten Beispiel hatte das zuständige Bundesministerium den Auftrag, die Umsetzung zu untersuchen und „die Länder bei der Entwicklung der Untersuchungsansätze und der Auswertung der Untersuchung zu beteiligen“ (*top-down*-Auftrag). Zudem war nur in diesem Beispiel auch eine Berichtspflicht an Bundestag und Bundesrat gesetzlich festgeschrieben. Mit den in Abschnitt 2 vorgestellten Ergebnissen würde Schwabs These gestärkt, dass Mehrebenenkonstellationen die Abstimmung von Evaluierungsaktivitäten und die Entwicklung eines einheitlichen konzeptionellen Ansatzes erschweren (Schwab, 2017, S. 214).

Auch der Grad der „Institutionalisierung der Evaluierungsfunktion“ (Konzendorf, 2009) im Politikfeld hatte einen relevanten Einfluss auf das Evaluationsarrangement. Während im Bereich der Arbeitsmarktpolitik (Beispiele eins und drei) das IAB einen gesetzlichen Auftrag zu Arbeitsmarkt- bzw. Wirkungsforschung hat, gibt es keinen vergleichbaren Auftrag im Bereich der Sozialhilfepolitik, zu dem das zweite Beispiel zählt. Im BSHG galten zwar Regeln zur Generierung einer bundeseinheitlichen Sozialhilfestatistik, aber eine Regel zu wissenschaftlicher Begleitforschung existierte bis 2002 nicht. Erst dann wurde eine Rechtsgrundlage für den Datentransfer zu Forschungszwecken im Auftrag des Bundessozialministeriums eingeführt, eine institutionalisierte Evaluierungsfunktion war aber dennoch nicht gegeben. Dies änderte sich erst 2005 und nur für die Grundsicherung für Arbeitsuchende. In § 55 SGB II ist nun das IAB mit der regelmäßigen Wirkungsforschung beauftragt. Schwab (2017, S. 214) argumentiert, dass „eine institutionelle Verankerung [...] die Rolle von Evaluation [stärkt]“. Die von ihm angeführte „Gefahr einer Ritualisierung von Evaluierungsprozessen“ ließ sich in den hier aufgeführten Beispielen experimenteller Politik nicht er-

kennen. Inwiefern sie für die regelmäßige Wirkungsforschung des IAB besteht, das als Teil der BA von dieser nicht unabhängig ist, wäre gesondert zu untersuchen.

### 3.2 Vergleich der Evaluierungsarrangements

Hatten diese Unterschiede in den institutionellen Rahmenbedingungen Auswirkungen auf die Evaluierungsarrangements?

Im Vergleich zeigt sich, dass der Evaluierungsgegenstand im Sinne von „Eigenschaften der politischen Programme“ (Schwab, 2017, S. 215) die Rolle der Evaluierungen mitbestimmte. Das erste Beispiel griff Fragen der finanziellen und regulativen Ausgestaltung eines Instruments der Arbeitsförderung auf und kann als typisch für Programm-Evaluationen verstanden werden. Grundlegende Legitimationsfragen stellten sich dabei nicht, ein Scheitern der Kombilohn-Modelle hätte die Teilnehmenden nicht existenziell gefährdet. Das zweite Beispiel war mit materiellen Fragen der Bemessung der Existenzsicherung beschäftigt und berührte das grundgesetzlich geschützte Recht auf menschenwürdige Existenz. Derart existenzielle Gegenstände erzeugen ein hohes Interesse der von der Regelung Betroffenen, das allerdings nur dann sichtbar wird, wenn diese sich kollektiv für ihre Interessen einsetzen (können). Im zweiten Beispiel konnten die Betroffenen ihre durchaus berechtigten Interessen punktuell individuell, aber nicht kollektiv organisieren. Anders im dritten Beispiel, in dem es für Kommunen und die BA darum ging, wer die Grundsicherung als Leistungsträger zugeschrieben bekommt und wer diese Aufgabe „verlieren“ würde. Hier waren die Kommunalen Spitzenverbände als kollektive Akteure und die BA als interessierte Partei und zugleich mit der Generierung der notwendigen Daten beauftragter Umsetzungsakteur in die Ausgestaltung und Begleitung der Evaluierung einbezogen. Darüber hinaus organisierten sich auch die Jobcenter beider Umsetzungsmodelle über eigene Vertretungen und beteiligten sich aktiv an den Diskursen über die Evaluation. Hier zeigt sich die in der Evaluationsforschung thematisierte Bedeutung der Machtposition der Stakeholder bei der Gestaltung der Evaluierungsarrangements. „„Meaning‘ in evaluation research is not negotiated within a simple plurality of competing stakeholder interests. Different stakeholder views and interests are expressed from positions of more or less power“, argumentieren beispielsweise David Taylor und Susan Balloch (2005, p. 2).

Die Evaluierungen, die auf Sozialhilfe-Daten angewiesen waren, hatten einen schwierigeren Zugang zu Primärdaten als die Evaluierung im Bereich der Arbeitslosenversicherung und des SGB II, weil die Kommunen die Sozialhilfedaten dezentral und mit verschiedenen Softwaresystemen generierten und die Evaluierungen wesentlich auf zusätzliche Befragungen angewiesen waren. Dies führte dazu, dass eher Vorher-Nachher-Vergleiche als Kontrollgruppen-Vergleiche angestellt wurden. Ein Teil der Evaluierungen zu Kombilohn-Modellen konnte aber auf Daten der BA zugreifen (Kaltenborn, Krug, Rudolph, Weinkopf & Wiedemann, 2005, S. 13, 45). Deshalb konnten hier multivariate Analysen zum Kontrollgruppen-Vergleich auf der individuellen Ebene der Teilnehmenden eingesetzt werden. Die Evaluierung zum SGB II konnte auch auf zentral gehaltene Daten der BA zum SGB II zugreifen, weshalb ambitionierte wissenschaftliche Analysemethoden – neben multivariaten Analysen zum Kontrollgruppen-Vergleich auf der individuellen Ebene auch solche auf der Ebene der Jobcenter – eingesetzt werden konnten.

## 4 Fazit

Mit der vorliegenden Analyse konnten verschiedene Thesen Oliver Schwabs zur Rolle von Evaluierung in politisch-administrativen Systemen (Schwab, 2017) getestet werden. In der Gesamtbetrachtung der drei hier diskutierten Evaluierungen zeigt sich ein deutlicher Zusammenhang zwischen institutionellen Rahmenbedingungen und der Gestaltung der Evaluierungsarrangements. Je dezentralisierter der Evaluationsauftrag und je weniger institutionalisiert die Evaluationsfunktion war, desto weniger ambitioniert waren die eingesetzten Evaluierungsmethoden, desto weniger waren die Evaluationsergebnisse miteinander vergleichbar und desto unwahrscheinlicher war ihre vollständige Veröffentlichung. Zugleich hatten auch die Evaluationsgegenstände und die Verfügbarkeit von Datenquellen einen Einfluss auf die Evaluierungsarrangements.

Diese „varieties of evaluation“ (Schwab, 2017) erscheinen zudem davon abhängig, wie stark Akteure der jeweiligen Advokaten-Koalitionen politisches Interesse an der Deutung und der Veröffentlichung von Ergebnissen hatten. Die drei untersuchten Fälle unterlagen im Evaluierungsprozess unterschiedlich starken politischen Deutungskämpfen, die abhängig waren von der institutionellen Ebene und Umstrittenheit des Gegenstandes sowie dem Organisationsgrad und der Machtposition der handelnden Akteure.

## Anmerkungen

- 1 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Selbstverständlich sind alle Geschlechter gleichermaßen gemeint.
- 2 Zur grundlegenden Kritik an der Debatte um das Lohnabstandsgebot siehe Gerhard Bosch (1999).
- 3 Die Landesregierungen versprachen sich dabei vor allem finanzielle Unterstützung vom Bund, die dieser im Rahmen des Sonderprogramms CAST auch gewährte.
- 4 Damit passt das Beispiel zu den haushaltspolitisch motivierten Evaluierungsarrangements der zweiten Welle, wie Hellmut Wollmann (2004) herausgearbeitet hat.
- 5 Art. 87 Abs. 2 Grundgesetz sieht eigenständige Bundeseinrichtungen als Körperschaften öffentlichen Rechts für die Sozialversicherungen, aber nicht für Fürsorgesysteme, vor. Nach Art. 87 Abs. 3 Grundgesetz kann der Bund durch Bundesgesetz selbständige Bundesoberbehörden oder bundesunmittelbare Körperschaften errichten.

## Literatur

- AG Arbeitslosenhilfe/Sozialhilfe der Kommission zur Reform der Gemeindefinanzen (2003). *Bericht der Arbeitsgruppe „Arbeitslosenhilfe/Sozialhilfe“ der Kommission zur Reform der Gemeindefinanzen*.
- Andres, Gerd (2000). Über die Chancen, die Potenziale eines aktivierenden Sozialstaats in praktische Politik zu verwandeln. In Erika Mezger & Klaus W. West (Hrsg.) *Aktivierender Sozialstaat und politisches Handeln* (S. 41-49), Marburg: Schüren.
- Armbrorst, Christian (2003). § 101a Experimentierklausel. In Christian Armbrorst et al. (Hrsg.). *Bundessozialhilfegesetz. Lehr- und Praxiskommentar (LPK-BSHG) mit einer Kommentierung zum Asylbewerberleistungsgesetz* (S. 1149-1157). Baden-Baden: Nomos.
- Banner, Gerhard (2008). Logik des Scheiterns oder Scheitern an der Logik? Kommentar zu Lars Holtkamp ‚Das Scheitern des Neuen Steuerungsmodells‘. *dms – der moderne staat*, (1) 2, 447-455.
- Bender, Stefan & Rudolph, Helmut (1999). *Kosten eines gestaffelten Zuschusses zu den Sozialversicherungsbeiträgen. Simulation des Zuschußbedarfs auf der Basis des Jahreszeitraummaterials der Beschäftigtenstatistik*. IAB-Werkstattbericht 08/1999. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.

- Bertelsmann-Stiftung (2002). *Positionspapier: Eckpunkte einer Reform von Arbeitslosen- und Sozialhilfe*. Gütersloh: Bertelsmann-Stiftung.
- Bertelsmann-Stiftung (2007). *Jahresbericht 2006 für das SGB II-Benchmarking der Optionskommunen*. Gütersloh: Bertelsmann-Stiftung.
- Bertelsmann-Stiftung, Hans-Böckler-Stiftung & KGSt (Hrsg.) (2002). *Netzwerk Kommunen der Zukunft. Produkte der Netzwerkarbeit. Bd. 14. Kommunen und lokale Beschäftigungsförderung. Bausteine guter Praxis*. Frankfurt am Main: VAS.
- Blancke, Susanne (2004). *Politikinnovationen im Schatten des Bundes*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Boockmann, Bernhard, Brändle, Tobias, Fuchs, Philipp, Klee, Günther, Kugler, Philipp, Laub, Natalie, Sippli, Khira & Verbeek, Hans (2019). *Evaluation des ESF-Bundesprogramms zur Eingliederung langzeitarbeitsloser Leistungsberechtigter nach dem SGB II auf dem allgemeinen Arbeitsmarkt*. Zwischenbericht 31. März 2019.
- Bosch, Gerhard (1999). Niedriglöhne oder Innovation: Überlegungen zur Zukunft der Erwerbsarbeit. *WSI-Mitteilungen*, (52) 12, 861-869.
- Boss, Alfred (1994). *Explizite und implizite Besteuerung geringer Arbeitseinkommen - Aspekte der Armutsfalle in der Bundesrepublik Deutschland*. Kieler Arbeitspapier Nr. 643. Kiel: Institut für Weltwirtschaft.
- Bundesverfassungsgericht (2007). Urteil des Zweiten Senats vom 20. Dezember 2007 – 2 BvR 2433/04 - Abs. (1 - 228).
- Bundesverwaltungsgericht (2004). Urteil des 5. Senats vom 25. November 2004 – BVerwG 5 CN 1.03.
- Bundesnetzwerk Arge SGB II (2008). *Stellungnahme des Bundesnetzwerks Arge zur Geschäftsabweisung 13 / 2008 der Bundesagentur für Arbeit (BA)*. Verfügbar unter: [http://www.lag-arbeitsnrw.de/de/aktuelles/pdf-sgb/2008-04-25\\_Bundesnetzwerk\\_ARGE\\_SGB\\_II\\_Stellungnahme.pdf](http://www.lag-arbeitsnrw.de/de/aktuelles/pdf-sgb/2008-04-25_Bundesnetzwerk_ARGE_SGB_II_Stellungnahme.pdf) [12. Dezember 2019].
- Bundesrat (2003). *Gesetzesantrag der Länder Hessen, Baden-Württemberg, Bayern, Sachsen. Entwurf eines Gesetzes zur Sicherung der Existenzgrundlagen (Existenzgrundlagengesetz – EGG)*, Drucksache 654-03.
- CDU-Bundesgeschäftsstelle (1994). *Wir sichern Deutschlands Zukunft*. Regierungsprogramm von CDU und CSU.
- CDU-Bundesgeschäftsstelle (2001). *Neue Soziale Marktwirtschaft*. Diskussionspapier der CDU Deutschlands, 27. August 2001, Berlin.
- Deutscher Bundestag (1998). *Beschlußempfehlung und Bericht des Ausschusses für Gesundheit (14. Ausschuß) zu dem Gesetzentwurf der Bundesregierung – Drucksachen 13/10422, 13/10868 – Entwurf eines Ersten Gesetzes zur Änderung des Medizinproduktegesetzes (1. MPG-ÄndG)*, Drucksache 13/11021.
- Deutscher Bundestag (1999). *Beschlußempfehlung und Bericht des Ausschusses für Arbeit und Sozialordnung (11. Ausschuß) zu dem Gesetzentwurf der Bundesregierung – Drucksache 14/389 – Entwurf eines Siebten Gesetzes zur Änderung des Bundessozialhilfegesetzes*. Drucksache 14/820.
- Deutscher Bundestag (2001). *Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Pia Maier, Dr. Klaus Grehn, Dr. Ruth Fuchs, weiterer Abgeordneter und der Fraktion der PDS – Drucksache 14/7224 – Pauschalierung der Sozialhilfe*, Drucksache 14/7426.
- Deutscher Bundestag (2003). *Gesetzentwurf der Fraktionen SPD und BÜNDNIS 90/DIE GRÜNEN. Entwurf eines Vierten Gesetzes für moderne Dienstleistungen am Arbeitsmarkt*. Drucksache 15/1516.
- Deutscher Bundestag (2004). *Materialien für die öffentliche Anhörung von Sachverständigen am 26. April 2004 in Berlin zum Entwurf eines Gesetzes zur optionalen Trägerschaft nach dem Zweiten Buch Sozialgesetzbuch (Kommunales Optionsgesetz) – BT-Drs. 15/2816*. Zusammenstellung der schriftlichen Stellungnahmen. Ausschussdrucksache 15(9)1112.
- Deutscher Bundestag (2010). *Beschlußempfehlung und Bericht des Ausschusses für Arbeit und Soziales (11. Ausschuss) a) zu dem Gesetzentwurf der Bundesregierung – Drucksache 17/1940, 17/2057 – Entwurf eines Gesetzes zur Weiterentwicklung der Organisation der Grundsicherung für Arbeitsuchende b) zu dem Gesetzentwurf der Fraktionen der CDU/CSU, SPD und FDP –*

- Drucksache 17/1555 – Entwurf eines Gesetzes zur Weiterentwicklung der Organisation der Grundsicherung für Arbeitsuchende. Drucksache 17/2188.
- Empter, Stefan & Frick, Frank (Hrsg.) (1999). *Beschäftigungsorientierte Sozialpolitik in Kommunen. Strategien zur Integration von Sozialhilfeempfängern in das Erwerbsleben*. Gütersloh: Verlag Bertelsmann Stiftung.
- FDP-Bundesgeschäftsstelle (2002). *Bürgerprogramm 2002. Programm der FDP zur Bundestagswahl 2002 beschlossen auf dem 53. Ord. Bundesparteitag vom 10. bis 12. Mai 2002 in Mannheim*. Berlin: FDP-Bundesgeschäftsstelle.
- Felder, Bernd (2003). *Stellungnahme zur Anhörung „Entwurf eines Gesetzes zur Einordnung des Sozialhilferechts in das Sozialgesetzbuch“* – Bundestags-Drucksache 15/1514. (13) Ausschuss für Gesundheit und Soziale Sicherung. Ausschussdrucksache 0293. Berlin: Deutscher Bundestag.
- Fels, Gerhard, Heinze, Rolf G., Pfarr, Heide & Streeck, Wolfgang (1999). *Bericht der Wissenschaftlergruppe der Arbeitsgruppe Benchmarking über Möglichkeiten zur Verbesserung der Beschäftigungschancen gering qualifizierter Arbeitnehmer*. Verfügbar unter: [http://www.mpifg.de/people/ws/downloads/Bericht\\_Benchmarking-Gruppe.pdf](http://www.mpifg.de/people/ws/downloads/Bericht_Benchmarking-Gruppe.pdf) [6. April 2020].
- Fink, Ulf (1998). *Arbeit für Alle. Neue Initiativen zur Beschäftigungsförderung*. Bonn: Bouvier.
- Gerster, Florian & Deubel, Ingolf (2001). Das Mainzer Modell für Beschäftigung und Familienförderung. In Florian Gerster (Hrsg.). *Arbeit muss sich lohnen!* (S. 59-77). Baden-Baden: Nomos.
- Hartmann, Helmut (2002). *Organisationsmodell der Arbeitsförderung*. Revidierter Entwurf, erstellt für die Unterarbeitsgruppe „Reform der Arbeitslosen- und Sozialhilfe“ der Bertelsmann Stiftung, 25. Februar 2002. Hamburg: con\_sens GmbH.
- Hartmann, Helmut, Ahlrichs, Jan & Böttcher, Beate (2008). *Die 69 Optionskommunen der Bundesrepublik Deutschland. Benchmarking der Optionskommunen – Vergleichsring IV – Entwicklung und Tendenzen im SGB II. Berichtsjahr 2007*. Hamburg: con\_sens GmbH.
- Henneke, Hans-Günter (2004). Die Saga um „Hartz IV“: zur Reform der sozialen Sicherungssysteme und ihrer Umsetzung. *Zeitschrift für Staats- und Europawissenschaften*, (2) 4, 548-579.
- Hess, Doris, Schröder, Helmut, Smid, Meno & Reis, Claus (2004). MoZaT. *Neue Strukturen für Jobs. Abschlussbericht der wissenschaftlichen Begleitforschung*. Mit BMWa und den Ländern abgestimmte Endfassung. Bonn: infas Institut für angewandte Sozialwissenschaft.
- Hesse, Joachim Jens, Götz, Alexander, Rüdiger, Ronald & Schubert, Simon (2008). *Evaluation der Aufgabenträgerschaft nach dem SGB II 2005 – 2008. Abschlussbericht*. Untersuchung im Auftrag des Deutschen Landkreistages. Band 80 der Veröffentlichungen des Vereins für Geschichte der Deutschen Landkreise e.V., Berlin: Deutscher Landkreistag.
- Hoffmann, Susanne (2000). Hilfe durch Systemreform? *Bundesarbeitsblatt*, (51) 5, 5-10.
- Institut für Sozialforschung und Gesellschaftspolitik (2006). *Jahresbericht 2006 an das Bundesministerium für Arbeit und Soziales. Evaluation der Experimentierklausel nach § 6c SGB II – Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „zugelassene kommunale Träger“ und „Arbeitsgemeinschaften“*. Dienstleistungsauftrag: Administrative Unterstützung und wissenschaftliche Beratung, Köln: Institut für Sozialforschung und Gesellschaftspolitik GmbH.
- Kaltenborn, Bruno (2000). Reformkonzepte für die Sozialhilfe und ihre Konsequenzen für Fiskus und Arbeitsangebot. Mikroökonomische Analyse mit dem Sozio-ökonomischen Panel 1986 bis 1996 und dem Simulationsmodell SimTrans. *Mitteilungen aus der Arbeitsmarkt und Berufsforschung*, 33, 68-79.
- Kaltenborn, Bruno (2001). Kombilöhne in Deutschland – Eine systematische Übersicht. *IAB-Werkstattbericht Nr. 14*. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Kaltenborn, Bruno, Krug, Gerhard, Rudolph, Helmut, Weinkopf, Claudia & Wiedemann, Eberhard (2005). *Evaluierung der arbeitsmarktpolitischen Sonderprogramme CAST und Mainzer Modell*. BMAS Forschungsbericht 552. Berlin: Bundesministerium für Wirtschaft und Arbeit.
- Kaps, Petra (2006). *Arbeitsmarktintegration oder Haushaltskonsolidierung. Interessen, Zielkonflikte und Ergebnisse kommunaler Beschäftigungspolitik*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kaps, Petra (2008). Die Grundsicherung in der Politikverflechtungsfalle? *dms – der moderne staat*, (1) 2, 399-422.



- Kaps, Petra (2009). Konstitutionelles Experiment als Governanceform. In Sebastian Botzem, Jeanette Hofmann & Sigrid Quack (Hrsg.). *Governance als Prozess. Koordinationsformen im Wandel* (S. 607-633), Baden-Baden: Nomos.
- Klinger, Roland (2000). Pauschalierung der Sozialhilfe wird in den Stadt- und Landkreisen verwirklicht. Reformvorschlag des Landkreistags Baden-Württemberg setzt sich durch. *Landkreisnachrichten Baden-Württemberg*, (39) 3, 127-131.
- Konzendorf, Götz (2009). Institutionelle Einbettung der Evaluationsfunktion in Politik und Verwaltung in Deutschland. In Thomas Widmer, Wolfgang Beywl & Carlo Fabian (Hrsg.) *Evaluation. Ein systematisches Handbuch* (S. 27-39), Wiesbaden: VS Verlag.
- Kruse, Jürgen (2002). *Reform von Arbeitslosenhilfe und Sozialhilfe. Modell einer Zusammenführung beider Systeme*. Präsentation von Ergebnissen einer Arbeitsgruppe unter Moderation der Bertelsmann-Stiftung. Köln, 7. März 2002.
- Lessenich, Stephan (2009). *Das Grundeinkommen in der gesellschaftspolitischen Debatte*. WISO Diskurs. Bonn: Friedrich-Ebert-Stiftung.
- Maaß, Volker (2001). *Experimentierklauseln für die Verwaltung und ihre verfassungsrechtlichen Grenzen: zugleich ein Beitrag zu § 7a BerlHG*. Berlin: Duncker & Humblot.
- Martens, Rudolf (2006). Neue Regelsatzberechnung 2006: Zu den Vorschlägen des Paritätischen Wohlfahrtsverbandes und der Bundesregierung. *Soziale Sicherheit*, (55) 6, 182-194.
- Mummert Consulting (2003). Stand der bundesweiten Evaluation. Zwischenergebnisse von Mummert Consulting. *Informationen, Termine und Notizen zu Modellvorhaben der Kommunen zur Pauschalierung der Sozialhilfe. PASO Brief 7 (9-13)*, Köln: Univation e.V..
- Nothbaum, Norbert, Lübker, Susanne & Kämper, Andreas (2004). *Die Konsequenzen einer pauschalisierten Sozialhilfe für die HilfeempfängerInnen*. Zusammenfassung, Bielefeld: GOE Gesellschaft für Organisation und Entscheidung.
- Pautz, Hartwig (2008). Think-tanks in Germany: the Bertelsmann Foundation's role in labour market reform. *Zeitschrift für Politikberatung*, (1) 3/4, 437-456.
- Rothkegel, Ralf (2002). Der rechtliche Rahmen für die Pauschalierung von Sozialhilfeleistungen – insbesondere zur Experimentierklausel des 101 a BSHG. *ZFSH/SGB (41)* 10/11: 585-592, 657-668.
- Sabatier, Paul A. (1988). An Advocacy Coalition Framework of Policy Change and the Role of Policy Oriented Learning Therein. *Policy Sciences*, (21) 1, 129-168.
- Scharpf, Fritz W. (1997). Beschäftigungsfreundlich und sozial: ein Widerspruch? *Die Mitbestimmung*, (43) 4, 36-39.
- Schmachtenberg, Rolf (2003). *Zusammenlegung von Arbeitslosenhilfe und Sozialhilfe*. Vortrag auf der Fachtagung der nationalen Armutskonferenz (NAK) „Aktueller Stand der Arbeitsmarkt- und Sozialhilfereform“, Berlin, 14. Mai 2003.
- Schupp, Jürgen, Volz, Joachim, Wagner, Gert G. & Zwiener, Rudolf (1999). Zuschüsse zu den Sozialversicherungsbeiträgen im Niedriglohnbereich: wenig zielgerichtet und teuer. *DIW-Wochenbericht*, (66) 27, 499-509.
- Schwab, Oliver (2017). Varieties of Evaluation? Zur Rolle von Evaluierung in politisch-administrativen Systemen. Literaturbasierte Standortbestimmung und Ausblick. In Sabine Kuhlmann & Oliver Schwab (Hrsg.). *Starke Kommunen – wirksame Verwaltung. Fortschritte und Fallstricke der internationalen Verwaltungs- und Kommunalforschung* (S. 199-225). Wiesbaden: VS Verlag.
- Sell, Stefan (1999). Zur Schnittstellenproblematik zwischen Arbeitslosenversicherung und Sozialhilfe. Bestandsaufnahme der Risikoverlagerungen und Plädoyer für eine funktionsgerechte institutionelle und instrumentelle Modernisierung. *Sozialer Fortschritt*, (48) 1-2, 24-33.
- Siebert, Horst (1996). *Hundert Punkte für mehr Beschäftigung*. Kieler Diskussionsbeiträge 264. Kiel: Institut für Weltwirtschaft.
- Sinn, Hans-Werner, Holzner, Christian, Meister, Wolfgang, Ochel, Wolfgang & Werding, Martin (2002). Aktivierende Sozialhilfe. Ein Weg zu mehr Beschäftigung und Wachstum. *Ifo-Schnelldienst*, (55) 9, 7-14.
- SPD-Parteivorstand (1994). *Reformen für Deutschland*. Das Regierungsprogramm der SPD. Bonn.

- SPD & Bündnis 90/Die Grünen (1998). *Aufbruch und Erneuerung – Deutschland Weg ins 21. Jahrhundert*. Koalitionsvertrag zwischen der Sozialdemokratischen Partei Deutschlands und BÜNDNIS 90/DIE GRÜNEN.
- Spermann, Alexander (1996). Das „Einstiegsgehalt“ für Langzeitarbeitslose. *Wirtschaftsdienst*, (76) 5, 240-246.
- Spindler, Helga (2004). Die neue Regelsatzverordnung – Das Existenzminimum stirbt in Prozentschritten. *Info also*, (22) 4, 147-166.
- Taylor, David & Balloch, Susan (2005). *The Politics of Evaluation. Participation and Policy Implementation*. Bristol: The Policy Press.
- Univation. Institut für Evaluation und wissenschaftliche Weiterbildung e.V. & Dr. Beywl & Associates GmbH (2005). *Pauschalisierte Leistungsgewährung in sozialen Sicherungssystemen – Erfahrungen und Konsequenzen*. Abschlussbericht der wissenschaftlichen Begleitung zu den Modellvorhaben zur Pauschalierung von Sozialhilfe in NRW. Köln.
- Wienand, Manfred (2003). Verzahnung von Arbeitslosen- und Sozialhilfe: Das Kooperationsmodell des Deutschen Städtetages. *Der Städtetag*, (56) 12, 9-10.
- Wollmann, Helmut (2004). Zur (Nicht-)Verwendung von Evaluationsergebnissen in Politik und Verwaltung. Eine vernachlässigte Frage der Evaluationsforschung. In Sabine Kropp & Sabine Kuhlmann (Hrsg.). *Wissen und Expertise in Politik und Verwaltung*, dms Sonderheft 1 (S. 87-102). Opladen: Barbara Budrich.
- ZEW, IAQ & TNS emnid (2008). *Evaluation der Experimentierklausel nach § 6c SGB II - Vergleichende Evaluation des arbeitsmarktpolitischen Erfolgs der Modelle der Aufgabenwahrnehmung „Zugelassener kommunaler Träger“ und „Arbeitsgemeinschaft“*. Untersuchungsfeld 3: „Wirkungs- und Effizienzanalyse“. Abschlussbericht. Mannheim, Gelsenkirchen, Bielefeld.

#### *Anschrift der Autorin:*

Petra Kaps, ZEP – Zentrum für Evaluation und Politikberatung, Mansteinstraße 8, 10783 Berlin, E-Mail: kaps@zep-partner.de.

Cornelia Schendzielorz, Martin Reinhart

## Die Regierung der Wissenschaft im Peer Review

### Zusammenfassung

Dieser Beitrag widmet sich der fallvergleichenden Analyse des unübersichtlichen Evaluierungsphänomens *Peer Review*. Ausgehend von der Vielfalt des Peer Review als Gegenstand der Soziologie des Bewertens wird eine Untersuchungsperspektive entwickelt, die Peer Review als Entscheidungsmechanismus über knappe Ressourcen in seiner Verfahrensförmigkeit und Steuerungsfunktion einer empirisch vergleichenden Verfahrensanalyse zugänglich macht. Der Forschungsansatz kombiniert die praxeologischen (Stefan Hirschauer) und funktionalistischen Verfahrensanalysen (Niklas Luhmann) mit konzeptionellen Ansätzen aus der Soziologie des Bewertens und ermöglicht es, relevante Parameter zur Differenzierung und Typologisierung von Peer Review in der Wissenschaft zu bestimmen. Anhand dessen werden spezifische Tätigkeiten herausgearbeitet, die als elementare Bestandteile von Peer-Review-Verfahren identifizierbar sind. Unterschiedliche Peer-Review-Prozesse lassen sich damit anhand ihres Komplexitätsgrades bestimmen und im Spannungsverhältnis von Selbst- und Fremdsteuerung als Evaluationspolitik verorten.

**Schlagwörter:** Peer Review, Legitimation durch Verfahren, Soziologie der Bewertung, Evaluationspolitik, Wissenschaft

### Abstract

#### *Governing Science Through Peer Review*

This article deals with the comparative analysis of peer review as a complex evaluation phenomenon. Starting from the variety of peer review as an object for the sociology of evaluation, we develop a research perspective that makes peer review in its procedural form and control function accessible to an empirical comparative procedural analysis comprehending it as a decision-making mechanism for scarce resources. Our research approach combines the praxeological (Stefan Hirschauer) and functionalist analyses of processes (Niklas Luhmann) with the conceptual approaches from the sociology of evaluation and enables to determine relevant parameters for differentiating and typologising peer review in science. This allows the identification of specific activities that can be identified as elementary components of peer review procedures. Differing peer review processes can thus be determined according to their degree of complexity and positioned as evaluation policy between internal and external forms of control.

**Keywords:** Peer review, legitimation by procedures, sociology of evaluation, evaluation policy, science

## 1 Einleitung

Dieser Beitrag richtet das Augenmerk auf die „varieties of evaluation“, die aktuell in der Wissenschaft vorzufinden sind. Gegenstand der Analyse sind die Evaluierungsarrangements des wissenschaftlichen Peer Reviews unter Berücksichtigung der wissen-

schaftspolitischen Dimension dieser Evaluationsverfahren, die sich u. a. darin zeigt, dass sie über die Verteilung knapper Güter eine Steuerungsfunktion ausüben. Das Peer Review ist nicht nur das zentrale Verfahren wissenschaftlicher Wertbestimmung und damit „one of the fundamental conditions of possibility of academic knowledge and the construction of its value“ (Biagioli, 2002, p. 11). Es ist zudem gleichermaßen bemerkenswert stabil, wie es in seinen konkreten Ausprägungen in spezifischen Prozeduren erstaunlich divers erscheint. Es ist einerseits vielseitig einsetzbar und lässt sich andererseits an den jeweiligen Bedarf anpassen. So sind sowohl Entscheidungsprozesse über Publikationsplatz als auch über Fördermittel oder Stellen mittels Peer Review organisierbar, weil sich dieses an die jeweiligen Erfordernisse anpassen lässt, bspw. durch verschiedene Formen der Verblindung, Arbeitsteilung und Rollentrennung oder durch mehrstufige Entscheidungsprozesse (Biagioli, 2002; Baldwin, 2018). Ähnlich der Evaluierungslandschaft anderer Felder ist es einerseits von Homogenisierungs- andererseits von Heterogenisierungstendenzen geprägt, die nicht zuletzt mit den diversen organisationalen Rahmenbedingungen – Verlag, Zeitschrift, Förderorganisation, Hochschule – zusammenhängen dürften. Somit konstatieren wir angesichts seiner Erfolgsgeschichte als Evaluationsform in der Wissenschaft eine „Metastabilität des Peer Reviews“ (Reinhart, 2012, S. 193), die empirisch mit einer Vielfalt von Peer-Review-Formaten einher geht.

Diese Vielfalt spiegelt sich auch in der Peer-Review-Forschung wider, die sich in großen Teilen dem Zeitschriften-Peer-Review widmet und zur quantitativen Analyse und Prüfung der Reliabilität, Fairness und Validität der Verfahren neigt (Weller, 2001; Hirschauer, 2005; Reinhart, 2012; Neidhardt, 2016). Diese Ausrichtung ist maßgeblich durch die nur selektiv und selten im synchronen oder diachronen Vergleich zugänglichen Daten bedingt. Der mitunter von den Organisationen, die Peer Review nutzen und mitorganisieren – konkret der Wissenschaftsverlage, Förderorganisationen und Hochschulen – begrenzte oder gar verweigerte Zugang zu Daten, die im Laufe jener Evaluationsarrangements generiert werden, stellt seinerseits einen evaluationspolitisch relevanten Befund dar. Dennoch ist angesichts der Bedeutsamkeit des Peer Reviews in und für die Wissenschaft ausgehend von den vielfältigen Fallstudien mit entsprechend fallspezifischen Ergebnissen eine Systematisierung mit Blick auf eine mögliche Theoretisierung wünschenswert. Vor diesem Hintergrund folgt dieser Beitrag dem Desiderat einer fallvergleichenden Analyse des unübersichtlichen Evaluierungsphänomens Peer Review. Dabei wird auf der Basis von Sekundäranalysen umfangreicher empirischer Studien zu Forschungsförderprogrammen ein konzeptionelles Argument zur Differenzierung und Typologisierung von Peer-Review-Prozessen entwickelt.

Dabei ist zu berücksichtigen, dass Peer Review neben der Qualitätssicherung durch wissenschaftliche Expertenurteile auch ein wesentlicher Bestandteil der Evaluierungsarrangements ist, mit denen Wissenschaft beansprucht, sich selber zu steuern. Die politische Dimension jener Selbststeuerung manifestiert sich darin, dass zur Rechtfertigung einer Autonomie und Selbstverwaltung der Wissenschaft auf Peer-Review-Verfahren zurückgegriffen wird (Wilholt, 2012). Gleichwohl wird Peer Review beispielweise im Zuge themenbezogener, kompetitiver Vergabe von Förder- und Forschungsgeldern ebenso zur Fremdsteuerung genutzt. Die Funktionalität und Stabilität des Peer Review gründet offenbar gerade darin, dass es ermöglicht zwischen Selbst- und Fremdsteuerung zu vermitteln und hier eine Art Scharnier bildet. Dies hat in der Forschungsliteratur überwiegend dazu geführt, diese Steuerungskonstellation unter dem Begriff der

Governance zu fassen (Schimank, 2016). Im Gegensatz dazu laufen unsere Theoretisierungsbestrebungen im vorliegenden Beitrag auf den Begriff der Regierung in einem allgemein politischen Sinne (Rosanvallon, 2018) hinaus. Im Folgenden werden wir verschiedene Fallbeispiele und Anwendungskontexte von Peer Review in einer vergleichenden Verfahrensanalyse zusammenführen. Dabei gilt es, die Anpassungsfähigkeit des Peer Review an vielfältige Evaluierungszwecke mit Blick auf jene Scharnierfunktion aufzuschlüsseln. Ziel der Analyse ist es, relevante Parameter zur Differenzierung und Typologisierung von Peer Review in der Wissenschaft herauszuarbeiten, ohne dabei die Einheit des Phänomens Peer Review aus den Augen zu verlieren. Der Beitrag nimmt Peer Review somit als einen Aspekt der Evaluationspolitik von Wissenschaft in den Blick, indem er spezifische Peer-Review-Formate als Bestandteile eines umfassenderen Systems der Selbst- und Fremdsteuerung und damit der Regierung von Wissenschaft untersucht.

## 2 Die Vielfalt des Peer Review als Gegenstand der Soziologie des Bewertens: Bestandsaufnahme und konzeptionelle Überlegungen

Peer Review findet in der Wissenschaft an zentralen Stellen Anwendung und leistet dort Qualitätssicherung über Bewertungsprozesse, insbesondere in Prozeduren zur Manuskriptbegutachtung bei Zeitschriften und Verlagen, zur Antragbegutachtung bei der Einwerbung von Drittmitteln und in der Begutachtung von Personen mit Blick auf ihre Eignung für langfristige Anstellungen, bspw. bei der Berufung von Professor\*innen. In allen drei genannten Fällen, deren Peer Review auch innerhalb eines Anwendungskontextes und desselben organisationalen Rahmens sehr heterogen sein kann, geht es erstens um die Frage der Verteilung knapper Güter, zweitens sind die Peer-Review-Prozesse verfahrensförmig organisiert und drittens erfüllen sie eine Steuerungsfunktion. Angesichts der bereits viel beforschten Frage, wie die knappen Ressourcen verteilt werden, bzw. ob sie fair, wissenschaftlich oder gesellschaftlich sinnvoll verteilt werden sollten<sup>1</sup>, konzentrieren wir uns zunächst auf die Verfahrensförmigkeit des Evaluationsarrangements Peer Review und erörtern anschließend seine Steuerungsfunktion als ein Instrument der Regierung von Wissenschaft. Dabei gilt es, die Verfahren als durch spezifische Praktiken gekennzeichnete in den Blick zu nehmen und zu fragen, in welchem Maße eine Legitimation durch Verfahren erfolgt und inwieweit diese für die Steuerungsfunktion maßgeblich ist. Im Folgenden schließen wir daher an Stefan Hirschauers Verständnis an, demnach Peer Review als ein sozialer (und primär kommunikativer) Prozess begriffen und als solcher in seiner Praxeologie analysiert werden kann (Hirschauer, 2005). Auf dieser Basis entwickeln wir unseren Analyseansatz in kritischer Auseinandersetzung mit der Soziologie der Bewertung (Lamont, 2012; Heintz, 2018) und in Anlehnung an Niklas Luhmanns Begriff des Verfahrens als einer Prozedur, durch die Legitimation erzeugt wird, weiter (Luhmann, 1983).

Eine praxeologische Betrachtung von Peer-Review-Prozessen kann auf Stefan Hirschauers Analyse des Zeitschriften-Peer-Reviews zurückgreifen: Peer Review ist darauf ausgerichtet, die Qualität der publizierten Fachbeiträge zu sichern, die Kommunikationsangebote zu optimieren und dadurch die Lesezeit einer Disziplin zu kalibrieren

(Hirschauer, 2019, S. 227). Die Bewertung erfolgt im Prozess der Urteilsfindung, im Zuge derer diverse Lesemodi von Reviewer\*innen und Editor\*innen – tendenziell inhaltsfokussiert bei Fachgutachter\*innen, tendenziell form- bzw. stilfokussierte bei Herausgeber\*innen – diskutiert werden. Diese Erörterungen ereignen sich beispielsweise in Herausgebersitzungen oder verschriftlicht in Form von Gutachten und der auf diese eingehenden Überarbeitungen der Texte durch die Autor\*innen (Hirschauer, 2005, S. 59-80). In diesem inhaltlichen Austausch zeigt sich bereits die sequenzialisierte Struktur der Prozesse, die sich auch im jeweiligen Begutachtungsverfahren manifestiert, in das sowohl schriftliche als auch mündliche Kommunikationsakte eingehen (Hirschauer, 2019, S. 278-281; 2015, S. 38).<sup>2</sup> Die Charakteristiken der Prozessualität, der Sequenzialität, der Verfahrensförmigkeit und die Mischung von Schriftlichkeit und Mündlichkeit gelten ebenso für Peer-Review-Formate zur Vergabe von Forschungsgeldern und Stellenbesetzungen. Auch in diesen Fällen handelt es sich um komplexe Bewertungsprozesse. Sie sind in mehrstufigen Verfahren organisiert, die zumindest eine grobe sequenzielle Anordnung der jeweiligen Bewertungspraktiken vorbahren. Dabei werden ebenfalls meist schriftliche Gutachten eingeholt und anschließend Einschätzungen und Lesarten in Gutachtersitzungen bei Förderorganisationen (Olbrecht, 2014) oder in Berufungskommissionen (Kleimann & Klawitter, 2017a) kontrovers diskutiert und erörtert.

Wie ist nun der Bewertungsprozess zu konzeptionalisieren? Im Rahmen der Soziologie der Bewertung findet sich der Vorschlag, Bewertungskonstellationen als Ausgangspunkt zu verwenden, die als Triade von Bewertetem, Bewertendem und Publikum konstituiert sind (Meier, Peetz & Waibel, 2016). Den Autoren zufolge lassen sich diese Konstellationen darüber hinaus vor dem Hintergrund transituativ gültiger Regeln der Bewertung und technologischer Infrastrukturen fassen. Dieser Ansatz zielt darauf ab, sich von einem akteurszentrierten Zugriff der Analyse von Bewertungen zu lösen. Allerdings führt diese Triade die Unterscheidung zwischen Subjekt und Objekt der Bewertung, und somit der Differenzierung zwischen einem aktiv agierenden und passiv bezeichneten Element fort. Zudem behält sie mit dem Begriff des Publikums auch in der Nomenklatur Spuren einer akteurszentrierten Perspektive bei. Als möglicher Ausgangspunkt oder Fixpunkt verbliebe dann einzig der fallspezifisch zu beschreibende Gegenstand der Bewertung. Einen solchen als (ex ante) klar bestimmbar anzunehmen, kollidiert allerdings mit einem praxeologischen resp. pragmatischen Verständnis von (Be-)Wertung. Demnach ist die Ungewissheit Ausgangspunkt und Anlass der evaluierenden Untersuchung, die der Bewertung zu Grunde liegt (Dewey, 1939; Potthast, 2017; Bogusz, 2018). Die Bewertungskonstellation lässt sich folglich mit der vorgeschlagenen Triade nicht hinreichend trennscharf erfassen und begrifflich bestimmen, um sie als heuristisches Analyseinstrument für eine vergleichende Verfahrensanalyse nutzbar zu machen.

Demgegenüber identifizieren Anne Krüger und Martin Reinhart (2016; 2017) fünf Bausteine der Bewertung: Bewertungspraktiken, Bewertungsstrukturen, Bewertungsinfrastrukturen, Bewertungssituationen und die Reflexivität von Bewertung. Außerdem plädieren sie innerhalb der Bewertungspraktiken für eine Differenzierung von Wertzuschreibung und Wertabwägung. Diese Unterscheidung aufgreifend ließe sich zur Charakterisierung der Verfahren und ihrer Sequenzialität fragen: Wo, d. h. an welchem Punkt im Bewertungsprozess, erfolgt auf welcher Ebene von wem und wie, d. h. in welcher Form, die *Wertzuschreibung* und die *Wertabwägung*? Das Potential der Unter-

scheidung für eine vergleichende Verfahrensanalyse ist jedoch auch damit begrenzt, dass die Separierung beider Praktiken fall- und verfahrensspezifisch stark variiert und somit schwer systematisierbar ist. Der konkrete Vollzug der Wertzuschreibung und ihrer Hervorbringung ist zudem bei vielen Verfahren in der empirischen Untersuchung schwer zugänglich; denn diese wird häufig als Werteinschätzung individuell erarbeitet und erst anschließend als bereits gefasste Wertzuschreibung erläutert und expliziert. Somit können im Zuge der Peer-Review-Verfahren häufig nur Bestätigungen, Infragestellungen, kontroverse Diskussionen, Modifikationen, Plausibilisierungen und Rechtfertigungen von Wertzuschreibungen beobachtet werden. Damit geht einher, dass die Thematisierung der erfolgten Wertzuschreibungen in Peer-Review-Verfahren häufig unter Legitimationsdruck erfolgen und dementsprechend rhetorisch rationalisiert werden. Dabei werden die affektbehafteten Elemente der Wertzuschreibung häufig euphemisiert, um ihre Akzeptanz als scheinbar rein in der Sache liegende Argumente nicht zu gefährden. Damit weist die Differenzierung von Wertzuschreibung und Wertabwägung in eine ähnliche Richtung wie Hirschauers Verständnis eines relationierten Fachurteils. Überdies zeichnen sich Konvergenzen ab, zwischen dem als „Reflexivität der Bewertung“ gefassten Hinterfragen der eigenen Wertzuschreibung und Wertabwägung und dem, was Hirschauer als „Urteilen unter Beobachtung“ beschreibt (Hirschauer, 2019). Letzteres impliziert, dass Aufmerksamkeit und Sachlichkeit derart intensiviert werden, dass sie zur Steigerung von Meinungsverschiedenheit führen, die dann – zugespitzt formuliert – im „Beurteilungspanoptismus“ (Hirschauer, 2019, S. 287, 290) mündet und reflexiv relationiert werden muss.

Inwiefern trägt nun die Verfahrensförmigkeit, also die prozesshafte, sequenziell formalisierte Organisation der Praktiken, die im Peer-Review-Prozess zusammenkommen zur Legitimation jener Qualitätsbewertung und (Selbst-)Steuerung bei? Wir gehen davon aus, dass eine gelingende Legitimation von Verfahren zur Qualitätssicherung über den Aspekt der personalen Teilnahme oder der Repräsentation entscheidend vermittelt ist (Luhmann, 1983). In dieser Perspektive hat ein Verfahren kein Wahrheitskriterium, fördert die Richtigkeit der Entscheidung, ermöglicht und kanalisiert Kommunikation und garantiert, dass Entscheidungen getroffen werden (Luhmann, 1983, S. 12-13). Demzufolge lässt sich die Verfahrensgestaltung von Prozessen der Qualitätsentwicklung und Bestimmung von Qualität anhand der folgenden fünf Aspekte bestimmen (Luhmann, 1983, S. 40-49; Reinhart, 2012, S. 134-145):

- Offenheit des Verfahrens,
- Verlässlichkeit, mit der es eine Entscheidung herbeiführt,
- Stufenförmigkeit, d. h. die Unterteilung in Sequenzen und Verfahrensschritte,
- Grad der Arbeitsteilung, d. h. die tätigkeitsbezogene, inhaltliche oder funktionale Aufsplittung der zu verrichtenden Aufgaben und schließlich
- Rollendifferenzierung, also die (personenbezogene) Differenzierung von funktionalen Zuständigkeiten, ggf. nach Professionalisierungsgrad, beruflicher Position, organisationalem Status und Repräsentationszwecken.

Um diese konzeptionellen Überlegungen im Sinne einer Heuristik zur vergleichenden Verfahrensanalyse von Bewertungsprozessen in der Wissenschaft fruchtbar zu machen, bilanzieren wir die Divergenzen und Konvergenzen der herangezogenen Literatur. Demzufolge lassen sich Peer-Review-Prozesse in der Wissenschaft als Bewertungsverfahren fassen, in welchen dem bewerteten Objekt ein spezifischer Wert zugeschrieben

und dieser dann zu anderen Bewertungen ins Verhältnis gesetzt und abgewogen wird (Krüger & Reinhart, 2016). Auf diese Weise wird in allen Fällen eine relationale Bewertung hervorgebracht, die im Zuge der Entscheidungsfindung darüber, ob das Manuskript publikationswürdig, das Forschungsvorhaben förderungsfähig oder der Stellenanwärter gewinnungswürdig ist, konsolidiert wird. Peer Review in der Wissenschaft ist somit ein durch das Verfahren und das Fachurteil legitimer Bewertungsprozess, der Steuerungsfunktionen erfüllt.

Wir plädieren für einen Ansatz, der vom Verfahren aus denkt und somit ermöglicht, im Verfahren zwischen Bewertungspraktiken, Infrastrukturierung, Regeln und Kriterien zu differenzieren. Das Verfahren bietet dabei eine transssituative Analyseeinheit, die es erlaubt, diverse Elemente des Verfahrens zu erfassen, in der empirischen Analyse deren Spezifika aufzuzeigen und in exemplarischen Untersuchungen der Tatsache Rechnung zu tragen, dass distinkte Schritte des Verfahrens in konkreten Situationen vollzogen werden und verankert sind. Somit können Infrastrukturen (Star, 1999; Star & Ruhleder, 1996; Star & Bowker, 2002; Niewöhner, 2014) als Manifestationen spezifischer Verfahrensschritte oder manchmal des gesamten Verfahrens gelesen werden. Derart können sie ggf. auch als „Framing“ im Goffmanschen Sinne (Goffman, 1980) fungieren. Zudem ermöglicht der Fokus auf das Verfahren als eine transssituative Analyseeinheit auch mit Blick auf das Ziel der Bewertungsverfahren deutlicher zu differenzieren. Daraus folgt die Unterscheidung ziel- und zweckspezifischer Regeln, welche die Bewertungsprozesse im Sinne von teils expliziten, teils impliziten Spielregeln kennzeichnen. Dementsprechend sind Bewertungspraktiken als multivalent zu denken, insofern in ihnen Regeln sowohl reproduziert, aktualisiert, als auch modifiziert und teilweise konstituiert und infrastrukturisiert werden. Infrastrukturierung kommt deshalb eine gesonderte Bedeutung zu, weil sie als „materiell-semiotische Praxis“ Beziehungen herstellt und unterbindet, wodurch soziale Praxis als Verflechtung von Akteuren, Technologien und Wertordnungen artikuliert wird (Niewöhner, 2015, p. 123).

Angesicht der dargelegten konzeptionellen Unschärfe insbesondere des Publikumsbegriffs favorisieren wir, die diesem Begriff inhärenten Fragen der Involvierung, Partizipation und Repräsentation in Bewertungsprozessen alternativ mit Erving Goffmans Konzept des Teilnahmestatus zu fassen. Dieses dient dazu, genauer zu unterscheiden, wie jemand einer Situation beiwohnt: Aktiv, akzeptiert, passiv, ausgeschlossen, geduldet, etc. Derart können auch in komplexen Peer-Review-Verfahren die Modi der situativen Beteiligung differenzierter gefasst werden und beispielweise die recht groben Begriffe des Publikums, Zuschauers oder Zuhörers dekonstruiert werden (Goffman, 1986, S. 40-42; 2005, S. 42 ff.; Goffman, 1986, S. 41, Fn. 24; 2005, S. 51, 54, 58, 61, 68). Goffman weiterdenkend, der den Begriff zur Analyse sprachlicher Äußerungen in physischer Kopräsenz nutzt, schlagen wir vor, den Teilnahmestatus weniger in Bezug auf Situationen, sondern mehr in Bezug auf einzelne Verfahrensschritte zu denken, indem man die Aufgaben und Praktiken der Beteiligten herausarbeitet. In diesem Kontext wird auch Goffmans Rahmenbegriff nicht so sehr hinsichtlich der Situationen relevant gemacht, sondern ist als „Framing“ bedeutsam, das sich in der Verfahrensgestaltung vermutlich eher in infrastrukturierenden Regeln und Abläufen manifestiert, welche die Handlungspraktiken der Beteiligten vorbahnen.



### 3 Methodologie

Ausgehend von diesen konzeptuellen Überlegungen zielen wir auf eine vergleichende Betrachtung von Peer-Review-Verfahren. Um relevante Parameter zur Differenzierung und Typisierung von Bewertungspraktiken zu identifizieren, schlagen wir folgende Frageheuristik vor:

- Erstens: Zu welchem *Ziel und Zweck* wird die Bewertung vorgenommen? Die Frage richtet sich auf das verfahrensspezifische Ziel, das bspw. in einer funktionalen Zweckbestimmung des jeweiligen Peer Review im Wissenschaftssystem bestehen kann, und das in jedem Fall über den generellen Anspruch eines reliablen, validen und fairen Verfahrens hinaus geht.
- Zweitens: Welche expliziten und impliziten *Regeln* prägen das Verfahren? Damit sind nicht nur Regeln im Sinne von Bewertungskriterien gemeint, sondern dezidiert auch Regeln zur Personenkonstellation, zu Aufgaben-/Zuständigkeitsprofilen, zur Rollentrennung sowie zu Bewertungsordnungen und Begründungslogiken mitgedacht. Deskriptiv kann beispielweise unterschieden werden, ob es sich um formale Regeln, um rechtlich kodifizierte Regeln oder um informell eingeeübte Konventionen handelt und wie diese begründet werden. Zudem ist herauszuarbeiten, in welchem Maße es sich bei all dem um Regeln handelt, die Standards für wissenschaftliche Qualität setzen - bspw. Originalität oder Robustheit? In diesem Kontext ist auch zu eruieren, ob spezifische Praxislogiken vorliegen - bspw. in Form von hochschul- und wissenschaftspolitischen Maßgaben, Zielen und Interessen, davon unabhängigen oder damit verknüpften reputationslogischen Ansinnen, Bestrebungen oder Konventionen bspw. zwecks Sichtbarkeit und Aufmerksamkeit, oder auch von ökonomischen Erwägungen, die mit Blick auf Mitteleinwerbung und Profit sowie einer günstigen Platzierung auf dem Forschungs- und Wissenschaftsmarkt relevant sein können.
- Drittens: In welchen *Infrastrukturierungen* manifestieren sich die mehr oder weniger regulierten Verfahren? In Anschluss an Susan Star und Karen Ruhleder sowie an Jörg Niewöhner lassen sich mittels der differenzierten Erfassung von Infrastrukturierungen im Sinne einer prozessualen Analyseperspektive weitere Einblicke in Bewertungspraktiken gewinnen; denn die Brücke zwischen Praktiken und ihrer verfahrensförmigen Organisation wird anhand von Infrastrukturierungen geschlagen. Zudem geben die Infrastrukturierungen Auskunft über den variablen Formalisierungsgrad der Verfahren, welcher u. a. mit Blick auf die spezifische Steuerungsfunktion relevant sein kann.
- Viertens: Was wird bewertet? Die Frage nach dem *Bewertungsgegenstand* ist nur vermeintlich einfach und taucht deshalb hier als letzte auf. Der Bewertungsgegenstand ist in viel geringerem Maße ex ante gegeben, als die Akteure im Feld dies wahrnehmen oder eingestehen. Es gilt deshalb danach zu fragen, *wie* der Gegenstand der Bewertung *durch das* und *in dem Verfahren* bestimmt wird; aber auch wie lange er unterbestimmt bleibt? Die Frage derart zu stellen, trägt der Tatsache Rechnung, dass die Dimensionen der Bewertung, welche den Gegenstand der Bewertung bestimmen, nicht deckungsgleich sein müssen, mit den Parametern, entlang derer die gültige Entscheidung getroffen wird. Für viele Verfahren ist es typisch, dass auf eine Einzelbewertung (bspw. von Anträgen auf Förderung) eine

vergleichende Bewertung und Entscheidung folgt, die die Bewertungsgegenstände erweitert (bspw. wenn Anträge in einer abschliessenden Gutachtersitzung unter Berücksichtigung regionaler Kriterien rangiert werden).

Auf diese Weise trägt die Frageheuristik dem Deweyschen Verständnis von Valuationsprozessen (Dewey, 1939) Rechnung, die ihren Ausgangspunkt an konkreten Problemen bzw. der Ungewissheit einer Bestimmung haben. Dementsprechend wird auch der spezifische Gegenstand der Bewertung erst im Zuge des Bewertungsverfahrens konstituiert. Gleichwohl ist diese Konzeption in der Empirie mit sehr heterogenen Fallbeispielen und dementsprechend diversen Selbstbeschreibungen von Gegenständen der Bewertung im jeweiligen Kontext konfrontiert. Heuristisch lassen sich grob folgende drei Varianten skizzieren:

- a. Etwas wird als Gegenstand der Bewertung bezeichnet, dass a priori bereits als eine physisch-materielle Einheit gegeben ist, die wiederum in einem Kontext bewertet wird, aber als Träger und Konvergenzpunkt des Bewertungsgeschehens unumgänglich ist (bspw. die Bewerber\*in im Berufungsverfahren).
- b. Etwas wird zu Beginn des Verfahrens als Gegenstand definiert bzw. eingegrenzt, z. B. im Zuge einer Selektion oder indem festgelegt wird, was als Gegenstand im Verfahren zugelassen wird (bspw. Artefakte, wie das formgerecht eingereichte Manuskript, der formal korrekt gestellte Antrag auf Forschungsförderung und die formgerechte Projektbeschreibung oder Bewerbung inklusive der personenbezogenen Informationen)<sup>3</sup>.
- c. Erst im Zuge des Verfahrens bildet sich sukzessive und jeweils einzelfallspezifisch heraus, was Gegenstand der Bewertung ist, so dass der Gegenstand erst im Verlauf des Bewertungsprozesses konturiert wird, beispielsweise bei neuartigen Bewertungsverfahren, wie jenen im Rahmen der damals neuen Exzellenzinitiative oder aktuell der Verfahren zur Bewertung von Hochschulverbünden im Rahmen der Exzellenzstrategie (Möller, Antony, Hinze & Hornbostel, 2012; GWK, 2019).

Selbstverständlich sind in komplexen mehrstufigen Verfahren auch Mischformen denkbar, deren spezifische Ausgestaltung der Empirie zu entnehmen sein wird.

## 4 Verfahrensanalyse: Fragenspezifischer Vergleich

Diese Frageheuristik wird im Folgenden an konkreten Peer Review Prozessen zur Vergabe von Forschungsgeldern durchgespielt. Den Ausgangspunkt bildet das Fallbeispiel der Förderung von Sonderforschungsbereichen (SFB) durch die Deutsche Forschungsgemeinschaft (DFG). Dieser seit 1968 bestehende, langjährig erprobte zweistufige Peer-Review-Prozess beinhaltet verschiedenste Bewertungspraktiken (Expertenbegutachtung, Panel-Begutachtung, Vor-Ort-Begehung, Berichterstattung) und eignet sich daher als exemplarischer Fall, der umfassenden Einblick in die Komplexität der Peer-Review-Verfahren bietet. Eine bemerkenswerte Besonderheit dieses Peer-Review-Prozesses zur Entscheidung über die Einrichtung von Sonderforschungsbereichen besteht darin, dass diese als Verbundvorhaben konzipiert sind. Allerdings hat der Anteil der Drittmittel, die zur Finanzierung größerer Forschungsverbünde mit längerer Laufzeit<sup>4</sup> vergeben werden im Zuge steigender Drittmittelquoten seit den 1990er Jah-

ren deutlich zugenommen (Klein, Kraatz & Hornbostel 2012, S. 164 f.; Deutsche Forschungsgemeinschaft, 2018, S. 188, 194), so dass die Analyse der Verbundförderung aussagekräftige Einblicke in die Entwicklung und den gegenwärtigen Stand von Förderformaten verspricht. In der vergleichenden Analyse werden außerdem zur gezielten Ergänzung weitere Fallbeispiele herangezogen. Die Auswahl der Förderprogramme zielt darauf ab, auch die im deutschsprachigen Raum etablierten Formate der Einzelprojekt- und der Personenförderung fragenspezifisch zu berücksichtigen. Derart bezweckt die komparative Überprüfung zu eruieren, inwiefern die Parameter der Analyseheuristik fallvergleichend zur Differenzierung und Typologisierung von Begutachtungsprozessen geeignet sind.<sup>5</sup>

## 5 Ziel und Zweck der Peer-Review-Verfahren

Um das Ziel zu bestimmen, gilt es, sich zu vergegenwärtigen, worüber mittels der Peer-Review-Verfahren entschieden wird. Selbstverständlich variieren je nach Förderlinie die Dauer, der finanzielle Umfang und die Adressaten der Förderung auf institutioneller sowie personeller Ebene: Beispielsweise wird beim Begutachtungsverfahren zur Einrichtung von Sonderforschungsbereichen über eine sehr langfristige (bis zu 12 Jahre) und im Gesamtvolumen vergleichsweise üppige und interdisziplinär angelegte Forschungsförderung befunden. Angesichts dessen ist das Peer-Review-Verfahren des SFB darauf ausgerichtet, fachlich kompetente und plausibel begründete Bewertungen hervorzubringen, welche diese langfristigen Förderentscheidungen hinreichend legitimieren.<sup>6</sup> Im Vergleich dazu wird bei der Einzelprojektförderung des Schweizer Nationalfonds (SNF) lediglich über eine drei- bis fünfjährige disziplinspezifische Förderung befunden, welche die Durchführung der geplanten Forschung ermöglichen soll.<sup>7</sup> Beim Emmy Noether-Programm der DFG wird über die mittelfristige Finanzierung einer Nachwuchsgruppe und ihrer Leitung (früher bis zu sieben Jahren, aktuell bis zu 6 Jahren) entschieden.<sup>8</sup> Das Emmy Noether-Programm bezweckt damit eine Weiterqualifizierung in Form des Erwerbs von Leitungserfahrung und der Möglichkeit zur fachlich eigenständigen Profilierung als Forscher\*in mit Blick auf das Karriereziel Professur.

Trotz unterschiedlicher Programme sticht der von allen geteilte Zweck ins Auge, fachwissenschaftlich kompetente und intersubjektiv abgesicherte Bewertungen vorzunehmen, die auch forschungs- und wissenschaftspolitisch legitimierbar sind, u. a. weil das Fachurteil eine realistische Abschätzung der jeweiligen Gelingensbedingungen, Zielerreichung und erfolgreichen Durchführung gewähren soll. Auffällig ist hier, dass die Verfahren immer auch dem Fortbestand und dem Nachweis der Funktionsfähigkeit der sie ausrichtenden Organisationen (DFG/SNF) selbst dienen; denn deren Aufgabe besteht darin, an der Schnittstelle zur Politik Forschungsmittel zu verteilen und zuverlässig Entscheidungen zu fällen. (Guston, 2000; Reinhart, 2012)

## 6 Regeln der Peer-Review-Verfahren

Mit Blick auf die zweite Frage, *welche expliziten und impliziten Regeln* die Peer-Review-Verfahren zur Vergabe von Drittmitteln prägen, lassen sich folgende Regeln entlang ihrer spezifischen Funktionen differenzieren:

- Regeln zur Legitimation des Urteils,
- Regeln zur Personenkonstellation,
- Bewertungsregeln im Sinne von Kriterien der Beurteilung,
- Verfahrensregeln und Regeln der Entscheidungsfindung (*decision making rules*).

Wiederum variiert je nach Förderprogramm, wie genau diese Regeln mit Blick auf die jeweilige Funktion ausgestaltet werden.

Bei der Bewertung der längerfristig einzurichtenden SFB werden zur *Legitimation der fachlichen Beurteilung* folgende Aspekte in Anschlag gebracht: Auffällig ist der hohe und auf die Bewilligung fokussierte Zeitaufwand von zwei vollen Arbeitstagen. Neben der Antragslektüre durch die Panelmitglieder wird hier in Form der Vor-Ort-Begehungen sowie der Klausursitzungen der Panelmitglieder und DFG-Berichtersteller\*innen beträchtlich in die systematische diskursive Aushandlung der Bewertung investiert. Dabei wird im Rahmen der Panelbegutachtung eine intersubjektive Nachvollziehbarkeit der individuellen Einschätzungen der Teilprojektgutachter\*innen angestrebt (Klein, Kraatz & Hornbostel, 2012, S. 165-166). Zudem steht am Ende der fachlichen Bewertung eine kollektiv geteilte Notenvergabe<sup>9</sup>. Derart zielt die Panelbegutachtung von SFBs auf eine konsensuale Bewertung ab. Die *Personenkonstellation* im SFB-Peer-Review-Verfahren ist durch eine starke Rollentrennung gekennzeichnet. Sie erfolgt zwischen vier Rollen: Erstens den Gutachter\*innen, welche die fachliche Expertise beisteuern sollen; zweitens den zwei Vertreter\*innen der DFG-Geschäftsstelle, die als Verfahrenshüter\*innen und Moderator\*innen fungieren; drittens den von der DFG berufenen Fachwissenschaftler\*innen, die als Berichtersteller\*innen das Verfahren begleiten und die Aushandlung der Bewertung überwachen; und viertens den zwei stimmberechtigten Politiker\*innen, die erst im letzten Verfahrensschritt – der Entscheidung im Bewilligungsausschuss – hinzugezogen werden. Die fachliche *Bewertung* erfolgt nach den *Kriterien* der Originalität der Forschungsidee und der Fragestellung. Darüber hinaus sind die Vorarbeiten und die Qualifikationen der Antragstellenden ein entscheidendes Kriterium (Olbrecht, 2014, S. 272 ff.). Wie Meike Olbrecht herausarbeitet, bieten diese Kriterien erheblichen Deutungsspielraum mit Blick auf ebenso schwer bestimmbare Maßstäbe wie Relevanz, Innovationspotentiale und zu erwartende Erkenntnisgewinne der Forschungsvorhaben. Dieser Befund trifft ebenso auf die Kriterien anderer Peer-Review-Verfahren zur Vergabe von Drittmitteln zu, bspw. auf die Bewertungskriterien beim Peer Review im Rahmen der Exzellenzinitiative (Möller, Antony, Hinze & Hornbostel, 2012; Hornbostel & Möller, 2015) oder des Emmy Noether-Programms (Böhmer, Hornbostel & Meuser, 2008)<sup>10</sup>. Bereits an der verwendeten Nomenklatur lässt sich ablesen, dass im Verfahren anhand der Bewertungskriterien auch Bewertungsmaßstäbe zu ihrer eigenen Bekräftigung inszeniert werden (Olbrecht, 2014, S. 272 ff.). Für eine vergleichende Verfahrensanalyse lassen sich die Terminologien der expliziten Bewertungskriterien somit selbst als Element der Regularien zur Legitimation des Qualitätsurteils lesen.

In den Regelungen zur Legitimation des Qualitätsurteils, der Personenkonstellation und den Kriterien, nach denen die Bewertung erfolgt, manifestieren sich folgende *Ver-*

*fahrensregeln* des SFB-Peer-Review-Prozesses: Es handelt sich um ein mehrstufiges Verfahren, das entsprechend der Rollentrennung arbeitsteilig organisiert ist. Die inhaltliche Bewertung der SFB-Anträge in der Vor-Ort-Begutachtung ist von der Entscheidung im Bewilligungsausschuss formal und auch personell weitgehend entkoppelt. Vermittels der im Verlauf der Verfahrensstufen und Sequenzen wechselnden Personenkonstellationen präformiert das Verfahren rollen- und aufgabenspezifische *Teilnahmestatus*. Die Gutachter\*innen sind an der Bewertung aktiv beteiligt und in ihrer fachlichen Urteilskompetenz voll akzeptiert. Von der abschliessenden Entscheidung sind sie jedoch ausgeschlossen. Die Mitarbeiter\*innen der DFG-Geschäftsstelle sind in der Bewertung zugegen, allerdings bleiben sie fachlich passiv, da sie nur begleitende, dokumentierende und teils moderative Funktionen erfüllen. Von der Entscheidung sind sie ebenfalls ausgeschlossen. Die Berichterstattenden sind als einzige sowohl in der Bewertung als auch in der Entscheidung zugegen. Sie übernehmen sowohl eine verfahrenshütende Funktion im Zuge der gemeinsamen Hervorbringung der Bewertung als auch eine synthetisch kommunizierende Funktion bei der Einigung auf eine konsensuale Bewertung im Panel; konkret in Form der kollektiv vertretenen Notengebung und deren Formulierung für das Entscheidungsgremium. Dort sind sie als Präsentierende aktiv und als berichterstattende Informant\*innen akzeptiert, allerdings nicht stimmberechtigt. Die Politiker\*innen sind von dem gesamten Bewertungsprozess ausgeschlossen und lediglich im Entscheidungsgremium zugegen. Auf Basis der ihnen vorliegenden Informationen sind sie dort aktiv stimmberechtigt und damit auch ohne fachliche Expertise gewissermaßen als funktionale Vertreter\*innen derjenigen, die Infrastrukturen und Gelder bereitstellen, an der Entscheidung beteiligt. Mit Blick auf den Teilnahmestatus kommt hinzu, dass die Verfahrensmäßigkeit durch wechselseitige Beobachtung der Verfahrensbeteiligten untereinander in Panel-Sitzungen, Diskussionen und Begehungen und mittels der Verfahrensbegleitung durch die Berichtersteller\*innen abgesichert wird. Die wechselseitige Kontrolle ist somit relativ stark formalisiert ausgeprägt. Bemerkenswert ist zudem, dass durch die direkte persönliche Begegnung von Gutachtenden, Verfahrensverantwortlichen und Antragstellenden die Anonymität der beteiligten Personen aufgehoben ist. Gleichwohl wird durch das kollektive Urteil in Form einer Note, die von der Gruppe als undifferenzierte Einheit vertreten wird und als Konsens kommuniziert wird, der Rückschluss auf Individualbewertungen einzelner Panel-Mitglieder verhindert.

Im Unterschied dazu erfolgt die *Legitimation des fachlichen Urteils* sowohl bei der Projektförderung des SNF als auch bei der Personalförderung der Nachwuchsgruppenleitung im Emmy Noether-Programm durch zwei bis drei jeweils unabhängig voneinander, schriftlich verfasste Gutachten von Fachkolleg\*innen, die von der Förderorganisation bestellt werden. Derart soll in Zusammenschau der Gutachten pro Antrag ein gegen einseitige Vorlieben oder Vorbehalte abgesichertes Fachurteil erzeugt werden. Beim Emmy Noether-Programm kommt ebenso wie bspw. bei der Förderung von Exzellenzclustern eine mündliche Präsentation der Antragstellenden hinzu. Diese sind somit, anders als beim SNF-Verfahren, nicht anonym, so dass auch die Informationen zum beruflichen Werdegang der Antragsteller\*innen explizit in die Begutachtung mit eingehen. Die *Personenkonstellation* dieser Verfahren umfasst neben den Antragsteller\*innen und den Fachgutachter\*innen auch die Mitglieder jener Förderorganisationen, bspw. Mitglieder des Hauptausschusses der DFG oder des Forschungsrates bzw. Referent\*innen des SNF. Somit ist die fachliche Begutachtung unterteilt in individuell

unabhängig voneinander verfasste Fachgutachten und eine gemeinsame Diskussion der Gutachten sowie der mündlichen Präsentation und des Diskussionsverlaufs im Entscheidungsgremium.

In diesen im Detail zwar unterschiedlichen Regeln zur Legitimation des Urteils, zur Personenkonstellation und zu den Bewertungskriterien manifestieren sich fallübergreifende *Verfahrensregeln und -merkmale*: Eine ausgeprägte Stufenförmigkeit, Rollendifferenzierung und Arbeitsteilung. Diese zeigt sich insbesondere darin, dass die beteiligten Akteure mitunter mit exklusiven Zuständigkeiten für bestimmte Dimensionen der Bewertung betraut sind: Beispielsweise kommt den Gutachter\*innen die Aufgabe zu, die fachliche Qualität und Relevanz einzuschätzen. Das Personal der Förderorganisationen übernimmt meist administrative und kontextualisierende Aufgaben. Den Mitgliedern der Entscheidungsgremien fällt auf Basis der vorangegangenen Bewertungsprozesse die relationierende Bewertung zu, in der meist expliziter als zuvor auch finanzielle und wissenschaftspolitische Aspekte mit abgewogen werden. Zudem sind in allen Fallbeispielen die fachwissenschaftliche Begutachtung und die Entscheidung über die Finanzierung somit formal ebenso wie personell entkoppelt und manifestieren sich für die jeweilige Verfahrensstufe und Sequenz als rollen- und aufgabenspezifisch differenzierte Teilnahmestatus. Auffällig ist auch die sequenzielle Organisation der verschiedenen Verfahrensschritte. Dabei wird im Zuge der Sammlung diverser fachnaher Einschätzungen im Sinne einer Perspektivenvielfalt, eine breite und komplexe Informationsbasis erzeugt, die dann im Zuge des Verfahrens, in der Relationierung zu anderen Bewertungsdimensionen sukzessive wieder reduziert und mit Blick auf die Entscheidung enggeführt wird.

## 7 Infrastrukturierungen der Peer Review Verfahren

Hinsichtlich der dritten Frage nach den *Infrastrukturierungen*, in denen sich die Regularien der Peer-Review-Verfahren niederschlagen, lassen sich folgende Differenzierungen vornehmen: Zunächst ist zwischen Schriftlichkeit und Mündlichkeit zu unterscheiden. Dabei sind Subunterscheidungen zu berücksichtigen: Einerseits zwischen Verschriftlichung in Form von Fließtext (bspw. bei Gutachten oder der Vergabe semantischer Noten) und numerischer Verschriftlichung in Form von Zahlen, Zeichen, Kennzahlen u. a. (bspw. numerische Noten oder Voten); andererseits zwischen mündlichen Diskussionen, Beratungen, Erörterungen in Begutachtungs-Panels und mündlichem Austausch mit Antragsteller\*innen in der unmittelbaren persönlichen Begegnung mit den an der Bewertung Beteiligten (Diskussionen, Frage-Antwort Dialoge etc.). Darüber hinaus kann der Formalisierungsgrad auf einer Skala von informell zu formell, ggf. sogar bis hin zu rechtlich kodifizierten Regeln differenziert werden. Nicht zuletzt ist festzuhalten, welche prozeduralen Abläufe in physischer Kopräsenz oder zeitlich parallel und welche auch raum-zeitlich versetzt erfolgen und in welchem Maße die Prozesse analog bzw. digital erfolgen.

Alle erwähnten Peer-Review-Verfahren zur Vergabe von Drittmitteln basieren auf schriftlichen Anträgen zu Forschungsvorhaben, die inhaltlich teilstrukturiert sowie teilformalisiert sind. Sie alle enthalten neben freien Fließtextteilen auch skalierte Noten- oder Zeichensysteme, numerische und tabellarische Informationen zu Antragsteller\*innen sowie zur Finanzierung der geplanten Projekte. Die Verfahren zur Vergabe

von Drittmitteln sind zudem weitestgehend analog verankert. Auch wenn Dokumente und Informationen digital versendet werden, bedarf es für die aktenauglichen Dokumente meist zusätzlich einer postalischen Zustellung. Bisher erfolgt das Peer Review meist zumindest nicht ausschließlich über eine digitale Plattform oder Vergleichbares.<sup>11</sup> Unterschiede zeigen sich insbesondere mit Blick auf das Verhältnis und die Abfolge von Mündlichkeit und Schriftlichkeit im Begutachtungsverfahren. Im SFB-Peer-Review-Verfahren ist die Erarbeitung und Aushandlung der Bewertung im Zuge der Diskussionen im Panel und der Vor-Ort-Begehung von Mündlichkeit geprägt. Die Verschriftlichung erfolgt in diesem Verfahren konsequent im Nachgang der mündlichen Behandlung. Dabei nimmt der Formalisierungsgrad im Laufe des Verfahrens im Wechsel von mündlicher Diskussion anschließenden Verschriftlichungen in Beratungsvermerken, Protokollen, Notenvergabe und Entscheidungsvorlagen sukzessive zu. Diese Schrift Dokumente fungieren somit als Instrument der sukzessiven Schließung eines mündlichen Gruppendiskussionsprozesses. Demgegenüber steht die weit verbreitete hybride Infrastrukturierung, die mit Blick auf Mündlichkeit und Schriftlichkeit gewissermaßen den umgekehrten Weg geht. Ausgehend von ausführlichen Fließtextverschriftlichungen hin zu einer mündlichen Diskussion, mündet sie in einer numerisch kondensierten Bewertung. Auch hier nimmt der Formalisierungsgrad im Zuge der Verknappung der schriftlichen Dokumente im Verfahrensverlauf zu.

## 8 Gegenstandsbestimmung der Peer Review Verfahren

Mit Blick auf die vierte Frage, *was im Zuge der beschriebenen Verfahrensweisen zum Gegenstand der Bewertung gemacht wird*, zeigt die Analyse, dass es sich keinesfalls um einen von Anfang bis Ende des Verfahrens fix zu identifizierenden Gegenstand handelt, sondern vielmehr um Gegenstandskonglomerate. Diese Konglomerate lassen sich in drei Komponenten untergliedern:

1. Die jeweiligen Textartefakte, d. h. Projektskizzen und Anträge, Gutachten, Sitzungsprotokolle, Vermerke, Berichte, Entscheidungsvorlagen;
2. Mündliche Einschätzungen, Diskussionsbeiträge und deren Relationierungen der Bewertenden und Entscheidenden (Gutachter\*innen in Panels, DFG Fachkollegiat\*innen, Berichterstatter\*innen);
3. Die schriftlich rhetorisch ausgearbeitete argumentative Überzeugungskraft der Antragsteller\*innen, die je nach Verfahren auch mündlich in Präsentationen und Diskussionen mit den an der Bewertung Beteiligten vorgeführt wird. Sie bezieht sich neben den inhaltlichen fachwissenschaftlichen Darlegungen insbesondere auch auf die strukturelle Passung zum Förderprogramm und ggf. auch auf die Darstellung der Persönlichkeit in der direkten Begegnung.

Mit Blick auf die Fallbeispiele ist deutlich, dass neben der qualitativen fachwissenschaftlichen Bewertung des Forschungsvorhabens anhand der Textartefakte und mündlichen Darlegungen sowohl bei der Verbundprojektförderung als auch bei der Personenförderung die geplante institutionelle Verortung des Forschungsvorhabens eine wichtige Rolle spielt. In diesen Förderlinien, beim Emmy Noether-Programm noch expliziter als beim SFB-Verfahren oder den Exzellenzclustern, werden zudem auch die Antragstellenden Personen (ob als Nachwuchsrgruppenleiter\*innen oder SFB- bzw. Cluster-Sprecher\*in-

nen) als Personen und in ihren bisherigen wissenschaftlichen Leistungen und Erfahrungen der Bewertung unterzogen. Hinzu kommen Form- und Stilaspekte der schriftlichen und mündlichen Darlegungen, die sich nicht nur in formaler Korrektheit, sondern darüber hinaus in situativer Angemessenheit, der Wortwahl, des Tonfalls, des Eingehens auf Fragen, der Kontextualisierung etc. niederschlagen. Auch diese können sich darauf auswirken, ob die an der Bewertung Beteiligten von der Relevanz und der Sinnhaftigkeit des beantragten Forschungsvorhabens überzeugt werden können.

Darüber hinaus werden im Zuge jener mehrstufigen Peer-Review-Verfahren zunehmend Meta-Bewertungen Teil des Gegenstands, beispielsweise wenn in den Diskussionen zu den Gutachten Stellung genommen wird und diese mit Blick auf ihre Trefflichkeit und Tauglichkeit als Bewertungen bewertet werden oder aber wenn Empfehlungen von Referent\*innen, Panelmitgliedern, Vermerke von Geschäftsstellenmitarbeiter\*innen oder Entscheidungsgrundlagen ihrerseits bewertet werden. Diese Aneinanderreihung von Bewertungen und Meta-Bewertungen, geht damit einher, dass zwar nicht unbedingt in physischer Kopräsenz (wie im Panel), sondern auch raum-zeitlich versetzt (Gutachten, Vermerke, Empfehlungen) die Bewertenden selber wiederum in ihrer bewertenden Praxis einer Bewertung unterzogen werden. In der Zuspitzung läuft diese Anhäufung von Metabewertungen auf das hinaus, was Hirschauer als „Beurteilungsphanoptismus“ bezeichnet (Hirschauer, 2019, S. 286 ff.).

Dabei fällt auf, dass in Peer-Review-Verfahren zur Drittmittelvergabe der Gegenstand der Entscheidung meist nicht identisch mit dem Gegenstand der Bewertung ist, weil die Begutachtungsprozesse häufig von den Entscheidungsprozeduren entkoppelt sind. Nicht zuletzt bedingt die sequenzielle Organisation der Verfahren, dass sich die Dokumentensammlung, über die bewertet und entschieden wird, im Laufe des Peer-Review-Verfahrens wandelt. Die ausgeprägte Rollendifferenzierung und Arbeitsteilung der Verfahrens beteiligten bedingt überdies, dass in den von der inhaltlichen Fachbegutachtung separierten Entscheidungsgremien, auf der Grundlage von Entscheidungsvorlagen mitunter auch wissenschaftspolitische und mit Blick auf den gesamten Zuständigkeitsbereich der Förderorganisation verteilungsstrategische Gesichtspunkte eine Rolle spielen.

## 9 Erweiterte Prüfung der Heuristik

Angesichts des Befunds, dass sich bei den erörterten Fallbeispielen aus der Forschungsförderung die Gegenstände der Bewertung erst im Zuge des Verfahrens konstituieren und es sich somit um im Verfahren herausgebildete Gegenstandskonglomerate handelt, wird im Folgenden ein kursorischer Blick auf die Verfahren des Zeitschriften-Peer-Review und des Peer Reviews im Rahmen von Berufungsverfahren gerichtet. Dies dient dazu, die Beständigkeit und Tauglichkeit der entwickelten Heuristik zu eruieren und das Potential der fallvergleichenden Analyse für Peer-Review-Prozesse jenseits der Drittmittelvergabe auszuloten. Aufgrund des begrenzten Platzes begnügen wir uns damit, diese Prüfung lediglich anhand der Eckpunkte der Heuristik, d. h. für die *Ziele und Zwecke* der Verfahren sowie für die *Gegenstandsbestimmung* im Verfahren vorzunehmen.

Ziele des Zeitschriften-Peer-Reviews sind die Kalibrierung der Lesezeit einer Disziplin sowie allgemein die wissenschaftliche Qualitätssicherung der publizierten Fachartikel (Hirschauer, 2005, S. 53 ff.; 2019, S. 227). Ziel eines Berufungsverfahrens ist es, die bestmögliche Passung von Entwicklungszielen der Hochschule und Fakultät mit der



langfristig angestellten Person zu erzielen, wie die Studien von Bernd Kleimann und Maren Klawitter zeigen (Klawitter, 2015; Kleimann, In der Smitten & Klawitter, 2015; Kleimann & Klawitter, 2017a; 2017b). Die Berufung dient damit dem Zweck der Profilbildung der Hochschule sowie der Aufgabenerfüllung ihres Personals in Forschung und Lehre.

Mit Blick auf den Gegenstand wird im *Zeitschriften-Peer-Review* anhand der Gutachten die inhaltliche und formale Qualität des Manuskripts, seine Passung zur Zeitschrift, zur Ausgabe, die Angemessenheit und Gültigkeit des im Manuskript erhobenen Geltungsanspruches und der im Gutachten vorgebrachten Beurteilung und Kritik bewertet. Somit handelt es sich auch hier um ein Gegenstandskonglomerat, das der Bewertung unterzogen wird. Diese richtet sich auf das eingereichte Manuskript und den darin erhobenen Geltungsanspruch der Autor\*innen, etwas im Rahmen dieser Zeitschrift und ihrer Leserschaft Passfähiges und Publikationswürdiges mitzuteilen. Im *Berufungsverfahren* werden die Werdegänge und Merkmale der Bewerber\*innen auf die ausgeschriebene Professur, d. h. die Eigenschaften bewertet, an denen ihre Eignung, Befähigung und ihr Leistungspotential abgeschätzt werden. Somit ist hier die Person in Hinblick auf ihre Passung zu den hochschul- und ggf. fakultäts- und institutsspezifischen Entwicklungen Gegenstand der Bewertung. Mit Blick auf die im Verfahren zusammengetragenen und eingeschätzten Informationen, entpuppt sich allerdings auch die Person als ein Konglomerat aus nachgewiesenen, beschriebenen, dargestellten und zugeschriebenen Merkmalen. Sie wird als ein Bündel aus a) manifesten, d. h. in Publikationen, eingeworbenen Mitteln, dokumentierten Forschungsleistungen etc.; b) einem biographischen Verlauf, der Aufenthalts- und Arbeitsorte listet, die sich mit Blick auf qualitätsdifferenzierende Zuschreibungen wie Reputation der Einrichtung, Internationalität, Wechselhäufigkeit etc. lesen und deuten lassen und c) der Persönlichkeit der Bewerber\*innen, die grob als soziales Verhaltensprofil mit Blick auf kommunikative, kooperative, argumentative, lehrdidaktische Fähigkeiten und eigene Interessen und nicht zuletzt der Gewinnungswahrscheinlichkeit der Bewerber\*in gefasst und bewertet. In vergleichender Perspektive hat dabei das Zeitschriften-Peer-Review mit manchen Verfahren zur Drittmittelvergabe, bspw. mit der Projektförderung des SNF die Anonymisierung der Einreichenden gemein. Hingegen teilen die Berufungsverfahren mit manchen Verfahren zur Vergabe von Fördergeldern die persönliche Präsentation der Bewerber\*innen.

## 10 Zwischenbilanz – Elemente des Peer Review

Trotz der Unterschiedlichkeit der dargelegten Fallbeispiele werden im Vergleich somit Überschneidungspunkte und Gemeinsamkeiten sichtbar: Zum einen der formale Befund, dass trotz divergierender Ziele, die Peer Review Verfahren immer selbst ein konstitutiver Bestandteil für die Bestimmung der Gegenstandskonglomerate sind, die hier einer Bewertung unterzogen werden. Zum anderen kristallisieren sich im Zuge der vergleichenden Verfahrensanalyse der Bewertungspraktiken spezifische Tätigkeiten heraus, die sich als elementare Bestandteile von Peer-Review-Verfahren identifizieren lassen:

### *Postulierende Tätigkeit*

Am Beginn des Prozesses steht ein Postulat. Die Einreichung eines Manuskripts, eines Förderantrags sowie der Bewerbung verstehen wir als postulierende Tätigkeit im Sinne

eines Antrags auf Genehmigung der damit verbundenen Ansprüche. Genauer gesagt betrachten wir den Anspruch auf Publikationsfähigkeit, den Anspruch auf Förderfähigkeit für ein Projekt oder eine Person und den Anspruch, dem ausgeschriebenen Stellenprofil zu entsprechen und es kompetent ausfüllen zu können, als Forderung nach Akzeptanz des Gesuchs. Diese Postulate schaffen den Anlass für die Bewertung, ihre Durchführung und ihre Organisation in Verfahren, um später eine evaluationsbasierte Entscheidung zu ermöglichen.

### *Konsultative Tätigkeit*

Zudem sind konsultative Aktivitäten wesentlich für Peer Review. Die Prozesse, in denen das Manuskript, der Vorschlag oder die Bewerbung und gegebenenfalls die persönliche mündliche Präsentation überprüft werden, verstehen wir als Beratungspraxis. Die schriftlichen Überprüfungen, Notizen und mündlich geäußerten Schätzungen zu den Postulaten werden als konsultative Gutachten eingeholt, um eine kompetente Bewertung zu gewährleisten.

### *Dezisive Tätigkeit*

Ebenso konstitutiv für den Peer-Review-Prozess ist, dass er eine Entscheidung zustande bringt. Der Verfahrensschritt, in dem die berechtigten Teilnehmer\*innen des jeweiligen Verfahrens eine verbindliche Entscheidung über die Annahme oder Ablehnung des Postulats treffen – sei es auf der Grundlage von Noten, Rankings, Volltextmitteilungen wie Berichten oder mündlichen Präsentationen – bezeichnen wir als entscheidungstreffende Tätigkeit.

### *Administrative Tätigkeit*

Unerlässlich sind zudem verwaltende Aktivitäten. Die Prozesse, in denen Postulate empfangen und ihre Behandlung eingeleitet und koordiniert werden, begreifen wir als administrative Tätigkeit. Diese Aktivitäten übernehmen in der Regel jene Organisationen, die Peer-Review-Prozesse einsetzen. Dementsprechend werden sie von deren Vertreter\*innen wie Herausgeber\*innen, editorialen Manager\*innen oder den Mitarbeiter\*innen der Fördereinrichtungen, Verlagen oder universitären Personalabteilungen durchgeführt. Die Verwaltungstätigkeit als eigenständige Tätigkeit zu definieren, trägt der Tatsache Rechnung, dass Peer-Review-Verfahren gewöhnlich in einem bestehenden organisatorischen Rahmen erfolgen.

### *Diskutierende Tätigkeiten*

Hinzu kommen diskursive und erörternde Aktivitäten. Wir definieren die Prozesse, in denen die Gutachten, Notizen, Einschätzungen, die zu den Bewertungen beitragen, ausgetauscht, kommentiert und diskutiert werden als Diskusstätigkeiten. Diese Diskussionspraktiken finden z. B. in Redaktionen, Herausgeber\*innengremien, Gutachter\*innensitzungen und Panels bzw. Entscheidungsgremien oder Berufungskommissionen statt und können schriftliche Gutachten im Fließtext sowie Memos, Sprechnotizen oder numerische bzw. zeichenbasierte Bewertungen und mündliche Beiträge beinhalten.

### *Präsentierende Tätigkeiten*

Ergänzend sind darbietende und vorzeigende Aktivitäten zu differenzieren. Die Prozesse, in denen die Postulierenden ihren Projektvorschlag, ihr Profil und ihre Anfrage persönlich vorstellen und erläutern, einschließlich der Beantwortung der damit verbundenen Fragen, fassen wir als präsentierende Tätigkeiten. Bei diesen Aktivitäten sind die

Postulierenden in den Bewertungsprozess mit eingebunden. Diese Aktivitäten sind in meist mehrstufigen Verfahren für Bewerbungen und personalisierte Zuschüsse sowie für Projekt- und Verbundprojektfinanzierungen zu finden, z. B. wenn im Rahmen der Begutachtung oder Evaluation Vor-Ort-Begehungen oder Besuche erfolgen. Meistens sind diese Tätigkeiten im Nachgang einer mehr oder weniger intensiven Einschätzung oder Bewertung schriftlicher Postulate veranschlagt.

### *Beobachtende Tätigkeiten*

Zu unterscheiden sind überdies observative Aktivitäten, die auch der Verfahrenskontrolle dienen. Die Prozesse, in denen das Peer-Review-Verfahren mit Blick auf die Einhaltung der jeweiligen Aufgaben, Rechte und Pflichten jeder Rolle, des Verfahrensablaufs und des Gesamtziels der Prozedur begleitet und gehütet wird, werden von uns als beobachtende Aktivitäten eingestuft. Diese Aktivitäten sind oft mit der Verpflichtung verbunden für die Entscheidungsgremien zu dokumentieren oder Bericht zu erstatten.

### *Moderierende Tätigkeiten*

Darüber hinaus bedingt die Ausdifferenzierung der Tätigkeiten einen Bedarf an Moderation, welche die situative Rahmung gewährleistet, in dem die Einzelaktivitäten zueinander in Beziehung gesetzt werden. Das geregelte Verfahren als solches erfordert moderierende Tätigkeiten, welche die Prozesse begleiten und bei Bedarf erläutern und erklären. Je komplexer und mehrstufiger sie sind, desto stärker muss das Zusammenspiel der verschiedenen Aktivitäten im laufenden Verfahren explizit abgestimmt und koordiniert werden, z. B. indem die Beteiligten in das Verfahren eingeführt und durch dieses begleitet werden. Dies kann auch die Übernahme von Sitzungsleitungen bei Diskussionen oder Präsentationen von Antragsteller\*innen oder Bewerber\*innen beinhalten.

Auf Grundlage der identifizierten elementaren Bestandteile von Peer-Review-Verfahren lässt sich wiederum der Blick für eine typologisierende Differenzierung der Verfahren öffnen. Angesichts der acht unterschiedenen Elemente des Peer Review lassen sich nun einerseits jene davon identifizieren, die als notwendige Bestandteile ein Minimalverfahren kennzeichnen: Postulierende, konsultative, dezisive und administrative Tätigkeiten. Andererseits kommen weitere Elemente modulhaft hinzu, die eine Anpassung an weitere Verfahrensziele und Kontexte leisten können: Diskutierende, präsentierende, observative und moderierende Tätigkeiten konturieren alle zusammengekommen ein maximales Verfahren (siehe *Tabelle 1*).

*Tabelle 1: Verfahrenskonstellationen des Peer Review*

| Minimalverfahren | Zusatzmodule in Richtung Maximalverfahren |
|------------------|-------------------------------------------|
| Postulat         | Diskussion                                |
| Administration   | Präsentation                              |
| Konsultation     | Observation                               |
| Dezision         | Moderation                                |

*Quelle:* Eigene Darstellung.

Derart können die Tätigkeiten auch als verschiedene Elemente gesehen werden, die ähnlich einem Baukastensystem je nach Ziel und Zweck des Verfahrens, Steuerungs- und Legitimationsbedarf unterschiedlich kombiniert werden können. Dabei sind diverse Anteilsverhältnisse vorzufinden, die bedingen, welche Tätigkeiten das Verfahren in

welchem Maße prägen oder gar dominieren. Beispielsweise fällt das Fachurteil von Expert\*innen in den Rahmen der konsultativen Tätigkeiten, so dass bei minimalistischen Peer-Review-Verfahren das Expert\*innenurteil einen größeren Stellenwert im Evaluierungsarrangement einnimmt. Dessen relatives Gewicht nimmt bei aufwändigeren und dadurch stärker modularisierten Verfahren tendenziell ab, da dort auch moderierende, präsentierende, beobachtende und diskutierende Aktivitäten im Verfahren integriert sind. Diese Binnendifferenzierung ist häufig zu beobachten, wenn zusätzlich zur fachlichen Qualitätssicherung die Scharnierfunktion zur Vermittlung unterschiedlicher Steuerungsinteressen bspw. von Wissenschaftsorganisationen und politischen Akteuren gewichtig ist. Zudem steigt mit zunehmender Komplexität der Verfahren auch der Bedarf, die unterschiedlichen Aktivitätsmodule miteinander zu artikulieren, was eine stärkere Formalisierung der Maximalverfahren u. a. in der Prozessgestaltung bedingt. Diese geht dann bevorzugt auch mit zunehmend formaler Infrastrukturierung mitunter gar mit Standardisierung einher.

## 11 Fazit: Evaluationspolitik der Wissenschaft - Regierung durch Peer Review

Die entwickelte Heuristik zielte darauf ab, Peer-Review-Formate als Bewertungsprozesse zu erschließen, um sie einer vergleichenden Verfahrensanalyse zugänglich zu machen. Die Heuristik fokussiert dementsprechend verfahrensrelevante Aspekte und ermöglicht dadurch viererlei: Erstens können Peer-Review-Verfahren nach ihrem *spezifischen Zweck* differenziert werden; zweitens ist der *Bewertungsgegenstand* nicht zwingend von Beginn an als scheinbar gegebener gesetzt; drittens erlaubt der weit gefasste Begriff von *Regeln*, die verfahrensförmige *Infrastrukturierung* der Praxis zu analysieren; und viertens ermöglicht dies, die häufig im Laufe des Verfahrens wechselnden Teilnahmestadien, bspw. anhand der Personenkonstellationen, Arbeitsteilung und Rollentrennung zu beschreiben und zu erfassen, ohne sie dabei statisch Einzelakteuren dauerhaft zuschlagen zu müssen. Derart werden der Prozessualität und Sequenzialität jener Verfahren konsequent Rechnung getragen.

Mittels der Analyseheuristik können somit Gemeinsamkeiten und Unterschiede der vielfältigen Peer-Review-Verfahren identifiziert werden, so dass greifbar wird, wie die Vielfalt, innerhalb der allen gemeinsamen Konzeption als Peer Review möglich ist. Erkennbar wird dabei auch, dass eine solche Diversität in der Einheit gewissermaßen Bedingung der Möglichkeit jener auch historisch zu konstatierenden „Metastabilität des Peer Review“ sein muss. Es scheint nachgerade das kennzeichnende Merkmal wissenschaftlicher Evaluierungsarrangements zu sein, dass sie auf den ersten Blick den Anschein erwecken müssen, einheitlich und wissenschaftsspezifisch zu sein, damit sie als Peer Review gelten können. Die Diversität innerhalb dieser Einheit der Peer-Review-Form offenbart hingegen, dass von Einheitlichkeit allenfalls in Bezug auf einzelne modulare Tätigkeiten gesprochen werden kann.

Angesichts der diversen Aktivitätsmodule und deren vielfältiger Fremdreferenzen ist zudem evident, dass es im Peer Review keinesfalls nur darum geht, zu bewerten. Die Kalibrierung der Lesezeit einer Disziplin, die fachkompetente intersubjektiv abgesicherte Qualitätsbewertung, die in Themensetzung, personeller Ausstattung und finanzieller Investition wissenschaftspolitisch legitimierbar ist, die Personalpassung mit

Blick auf fachliche und strukturelle Faktoren im Verhältnis zu den Zielen der Hochschulentwicklung und weitergedacht für Promotions- oder Habilitationsverfahren die Vergabe qualifikationsverbürgender Titel; all das sind exemplarische Leistungen, die dem Peer Review überantwortet werden. Neben der evidenten Steuerungsfunktion ist damit die Legitimationsfunktion augenfällig zentral und zwar nicht nur nach innen gegenüber der wissenschaftlichen Gemeinschaft, sondern auch nach außen, nicht zuletzt gegenüber der Politik. In diesem Spannungsfeld ist eine einheitliche Adressierbarkeit nötig, da es um die Verwendung öffentlicher Gelder für Forschung geht und zudem die Zugänglichkeit von daraus resultierendem als gesichert geltendem Wissen zur Debatte steht. Insofern liegt von vielfältigen Seiten – Wissenschaft, Politik, im Fall von Wissenschaftsverlagen auch Wirtschaft – ein Steuerungsinteresse vor, das in einer geteilten Terminologie ausgehandelt werden muss, um es handhabbar im Sinne von regierbar zu machen. Das vereinheitlichende *Label* „Peer Review“ gewährleistet für alle Akteure eine Erkennbarkeit, die nicht nur Steuerungsinteressen vermittelt, sondern in erheblichem Maße Legitimation für eine Regierung durch Peer Review verbürgt.

Dieses zeigt sich einerseits an der Geschichte des Begriffs „Peer Review“. Als verbreitete Bezeichnung für wissenschaftliche Evaluierungsarrangements taucht dieser Terminus nämlich erst in den 1970er Jahren auf und vermittelt erstmals zwischen neuen politischen Steuerungsinteressen und wissenschaftlichen Autonomieforderungen (resp. Selbststeuerungsinteressen) (Baldwin, 2018). Hier findet sich der historische Moment, in dem vielfältige wissenschaftliche Evaluierungsarrangements durch eine vereinheitlichende Terminologie legitimatorisch gewendet werden. Darüber hinaus sind die in der Folge dessen komplexer werdenden Verfahren eine Reaktion darauf, dass das Peer Review selbst wegen seiner Steuerungsfunktion erhöhtem Legitimationsdruck ausgesetzt ist. Reflexive Modernisierung und Wissensgesellschaft sind die Schlagworte, mit denen der gesellschaftliche Kontext seit den 1970er Jahren bezeichnet werden kann, der diesen Druck erzeugt (Weingart, 2001). Derart ist Peer Review nicht nur selbst Gegenstand von Politik und insbesondere von Evaluationspolitiken geworden, sondern hat dadurch auch eine immense Verbreitung seiner Anwendung und Nutzungsformen sowie Ausdehnung seiner Wirkmächtigkeit erfahren. Folgerichtig eröffnen sich im Ausblick für die Wissenschaftsforschung Möglichkeiten, den Themenkomplex um Qualitätssicherung, Ressourcenzuweisung und Expertise aus der leitenden Unterscheidung von Selbst- und Fremdsteuerung herauszulösen.

Angesichts des zunehmenden Legitimationsdrucks ließe sich die Frage weiterverfolgen, inwiefern eine zunehmende Ausdifferenzierung der Verfahren durch zusätzliche Aktivitäten deren Komplexität übermäßig steigert. Einerseits erfordern die Verfahren dadurch immer mehr Vorkenntnisse und Einarbeitung, bis hin zu Professionalisierung, die bspw. am Element der konsultativen Tätigkeiten anhand der Notwendigkeit von Gutachter\*innen-Trainings diskutiert werden. Extremes Beispiel hierfür ist die Einführung von sog. „impact case studies“ im Research Excellence Framework Grossbritanniens, wo sowohl die Evaluierten als auch die Evaluierenden in vorgängigen Testläufen regelrecht auf das Verfahren hin „kalibriert“ wurden (Manville et al., 2015). Andererseits droht die zunehmende Verfahrenskomplexität auch die Einheit dessen zu sprengen, was unter Peer Review gefasst wird und was seine Meta-Stabilität bedingt. Alternative Verfahren zum Peer Review erscheinen dadurch zumindest auf den ersten Blick effektiver und effizienter, so wie das gegenwärtig bspw. in der Diskussion um Losverfahren in der Forschungsförderung sichtbar wird. Zudem tritt auch das spezifisch wissenschaftliche am Peer Re-

view dann in den Hintergrund zugunsten der vielfältigen Fremdreferenzen, die Peer-Review-Formate ermöglichen und die insbesondere die Scharnierfunktion zum Politischen im Sinne eines Adaptationsvermögens in den Vordergrund treten lassen.

Als Schlussfolgerung unserer fallvergleichenden Analyse ist jedoch die Identifikation konstitutiver Aktivitäten und deren zielgerichtete modularisierte Zusammenstellung u. a. im Zuge von Maximalverfahren von vorrangiger Bedeutung: Die praxeologische Analyse befördert aus dem empirischen Material zu Tage, in welchem Maße Peer Review nicht nur in Evaluationspolitiken eingebettet ist, sondern auch mit politischen Gestaltungsfragen verwoben ist. In dieser Perspektive lassen sich die vielfältigen Evaluierungsarrangements des Peer Review mit Pierre Rosanvallon (2018) auch als verschiedene Regierungsvarianten in der Wissenschaft begreifen. Diese unterscheiden sich danach, wie in den divergierenden Verfahren die verschiedenen Steuerungsinteressen prozessualisiert und in sequenzialisierte und formalisierte Abläufe überführt werden, welche Tätigkeiten dabei wem zufallen, kurzum welche Verfahrensförmigkeit welche Teilnahmestatus, d. h. Repräsentations- und Beteiligungsformen zur Entscheidung über die Verteilung knapper Ressourcen vorsieht. Derart lassen sich die Evaluierungsarrangements des Peer Review als Regierungsarrangements der Wissenschaft begreifen. Im Zuge dessen werden wissenschaftliche Anforderungen an die Qualitätsprüfung innerhalb des Peer Review bspw. anhand des (langjährig erprobten) Verfahrenselements der Konsultation von Fachleuten an vorrangig politisch geprägte Aushandlungs-, Bewertungs- und Entscheidungsprozeduren wie bspw. die diskursiven und observativen Verfahrenselemente anschlussfähig gemacht. Die Kombination von auf das Spezialist\*innenurteil verweisenden und auf Einbindung, Aushandlung sowie auf die Kontrolle und Gewährleistung einer akuraten Rollentrennung und kompetenzgeleiteten Arbeitsteilung ausgerichteten Verfahrensmodule steht sinnbildlich dafür, wie Peer Review als ein Adapter fungiert, der wissenschaftliche und politische Qualitäts- und Legitimationserfordernisse ineinander überführt und füreinander kompatibel macht. Eine gouvernementale Nomenklatur (Foucault, 2006) erlaubt hier das verwobene Verhältnis von Regierenden und Regierten differenzierter bspw. als Momente von Repräsentation, Partizipation, Rollentrennung, aber auch von Veridiktion, Lesbarkeit und Responsivität zu untersuchen. Nicht zuletzt würden dadurch wissenschaftliche Evaluierungsarrangements noch deutlicher als Evaluationspolitik erkennbar.

## Anmerkungen

- 1 Siehe dazu die Rezeption der Literatur in Weller (2001), Hirschauer (2005), Reinhart (2012), Neidhardt (2016).
- 2 Darüber hinaus diagnostiziert Hirschauer dem Zeitschriften-Peer-Review Strategien der Kommunikationsintensivierung, die auf die Steigerung der Aufmerksamkeit, der Sachlichkeit, der Meinungsverschiedenheit und der Disziplinierung der Autorschaft zielen (Hirschauer, 2019, S. 294).
- 3 Mit Blick auf Fallanalysen ließe sich auch hier wieder die Frage produktiv machen, in welchem Maße diese Festlegung als eine Wertzuschreibung erfolgt und/oder in welchem Maße sie im Modus der Wertabwägung getroffen wird. Aufgrund des erläuterten mangelnden Generalisierungspotentials wird diese Frage allerdings hier zurückgestellt.
- 4 Darunter fallen in Deutschland sowohl die von der DFG im Rahmen sogenannter „koordinierter Programme“ – neben Sonderforschungsbereichen auch Forschungsgruppen, Forschungszentren, Schwerpunktprogramme und Graduiertenkollegs – als auch die seit Anfang der 2000er Jahre im Zuge der Exzellenzinitiative bzw. seit 2019 der Exzellenzstrategie von der öffentlichen Hand verausgabten Forschungsgelder.

- 5 Die Informationen zu den Fallbeispielen speisen sich aus verschiedenen Einzelfallstudien und der dazugehörigen Literatur. Insbesondere bei den Drittmittelverfahren wird auf umfangreiche empirische Studien im Umfeld des Instituts für Forschungsinformation und Qualitätssicherung (iFQ), heute Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), zu Sonderforschungsbereichen, zum Emmy Noether-Programm und zur Projektförderung des Schweizerischen Nationalfonds (SNF) zurückgegriffen, die ergänzt durch Webseiten- und Dokumentenanalyse sekundäranalytisch ausgewertet werden.
- 6 Auffällig ist, dass auch strukturelle Faktoren und wissenschaftspolitische Erwägungen eine gewichtige Rolle spielen. Aufgrund der langfristig angelegten Förderung (4+4+4) wird die Förderungswürdigkeit auch mit Blick auf inner- und interdisziplinäre Entwicklungen, die Relevanz im internationalen Vergleich und die Komplementarität zu bereits bestehenden Schwerpunkten eingeschätzt und abgewogen (Olbrecht, 2014).
- 7 Das Verfahren des Schweizerischen Nationalfonds zielt auf eine möglichst gut abgesicherte und somit fachwissenschaftlich und forschungspolitisch legitimierbare, da angemessene und sachgerechte Entscheidung über die Vergabe von finanziellen Mitteln zur Projektförderung (Reinhart, 2012).
- 8 Das Begutachtungsverfahren des Emmy Noether-Programms ist als Personenförderung für promovierte Wissenschaftler\*innen angelegt. Die Antragsstellenden beantragen die Einrichtung und Leitung einer Nachwuchsgruppe zu einer spezifischen Problemstellung, die an eine etablierte Forschungseinrichtung angegliedert ist (Deutsche Forschungsgemeinschaft, 2019)
- 9 Zu den Effekten gruppendynamischer Prozesse in der Panelbegutachtung bspw. auf die kollektive Urteilsfindung in konkurrenzialen Verfahren und die Notenvergabe siehe auch Klein, Kraatz & Hornbostel, 2012, S. 170-181.
- 10 Die fachliche Bewertung des Forschungsvorhabens für eine Nachwuchsgruppe wird anhand von Kriterien der Originalität und Konzeption des Forschungsprojektes eingeschätzt. Darüber hinaus wird auch die Passung des Forschungsvorhabens und der/des sich bewerbenden Nachwuchsleiter\*in/s zum Programm bewertet, die u. a. trotz der mit Blick auf das wissenschaftliche Alter relativ jungen Zielgruppe bereits Anforderungen wie internationale Forschungserfahrung und hochrangige Publikationen miteinbezieht. Die Bedeutung dieser Kriterien zeigt sich insbesondere in dem nach Meinung der Antragssteller\*innen sehr bedeutsamen Merkmal der persönlichen Publikationsleistung (inklusive der Reputation der Zeitschriften) und bisheriger Forschungsaktivitäten (Böhmer, Hornbostel & Meuser, 2008, S. 49).
- 11 Sowohl bei den genannten Verfahren als auch generell ist allerdings eine Verschiebung der Kommunikation auf digitale Plattformen deutlich erkennbar.

## Literatur

- Baldwin, Melinda (2018). Scientific Autonomy, Public Accountability, and the Rise of “Peer Review” in the Cold War United States. *Isis*, 109 (3), 538-558. <https://doi.org/10.1086/700070>.
- Biagioli, Mario (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures*, 12 (1), 11-45. <https://doi.org/10.1080/1045722022000003435>.
- Bogusz, Tanja (2018). *Experimentalismus und Soziologie: Von der Krisen- zur Erfahrungswissenschaft*. Frankfurt: Campus Verlag.
- Böhmer, Susan, Hornbostel, Stefan, & Meuser, Meuser (2008). Postdocs in Deutschland: Evaluation des Emmy Noether-Programms. *iFQ-Working Paper No.3*. Bonn.
- Deutsche Forschungsgemeinschaft (2018). *Jahresbericht 2018*. Aufgaben und Ergebnisse. Bonn. Verfügbar unter: [https://www.dfg.de/download/pdf/dfg\\_im\\_profil/geschaeftsstelle/publikationen/dfg\\_jb2018.pdf](https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/dfg_jb2018.pdf) / [17. März 2020].
- Deutsche Forschungsgemeinschaft (2019). *Emmy Noether-Programm*. Bonn. Verfügbar unter: [https://www.dfg.de/foerderung/programme/einzelfoerderung/emmy\\_noether/](https://www.dfg.de/foerderung/programme/einzelfoerderung/emmy_noether/) [29. November 2019].
- Dewey, John (1939). *Theory of Valuation*. Chicago: University of Chicago Press.
- Foucault, Michel (2006). *Sicherheit, Territorium, Bevölkerung*. (C. Brede-Konersmann & J. Schröder, Übers.). Frankfurt am Main: Suhrkamp.
- Goffman, Erving (1980). *Rahmen-Analyse: Ein Versuch über die Organisation von Alltagserfahrungen* (10. Aufl.; H. Vetter, Übers.). Frankfurt am Main: Suhrkamp Verlag.

- Goffman, Erving (1986). *Interaktionsrituale: Über Verhalten in direkter Kommunikation* (1. Aufl.; R. Bergsträsser & S. Bosse, Übers.). Frankfurt am Main: Suhrkamp Verlag.
- Goffman, Erving (2005). *Rede-Weisen: Formen der Kommunikation in sozialen Situationen* (H. Knoblauch, C. Leuenberger, & B. Schnettler, Hrsg.). Konstanz: UVK Verlagsgesellschaft.
- Guston, David H. (2000). *Between Politics and Science: Assuring the Integrity and Productivity of Reseach*. (2 Aufl.) Cambridge; New York: Cambridge University Press.
- GWK-Bonn (2019). *Die Exzellenzstrategie und ihre Vorgängerin, die Exzellenzinitiative*. Bonn. Verfügbar unter: <https://www.gwk-bonn.de/themen/foerderung-von-hochschulen/exzellenzstrategie-exzellenzinitiative/> [29. November 2019].
- Heintz, Bettina (2018). Von der Allmacht der Zahlen und der Allgegenwart des Bewertens. *Soziologische Revue*, 41 (4), 629-642.
- Hirschauer, Stefan (2005). Publierte Fachurteile. Lektüre und Bewertungspraxis im Peer Review. *Soziale Systeme*, 11(1), 52-82.
- Hirschauer, Stefan (2015). Oral Communication in Journal Peer Review. *Human Studies* 38: 37-55.
- Hirschauer, Stefan (2019). Urteilen unter Beobachtung. In Stefan Nicolae, Martin Endreß, Oliver Berli, & Daniel Bischur (Hrsg.), *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit* (S. 275-298). [https://doi.org/10.1007/978-3-658-21763-1\\_12](https://doi.org/10.1007/978-3-658-21763-1_12).
- Hornbostel, Stefan & Möller, Torger (2015). Die Exzellenzinitiative und das deutsche Wissenschaftssystem: eine bibliometrische Wirkungsanalyse. *Wissenschaftspolitik im Dialog* 12. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften.
- Klawitter, Maren (2015). Effects of Institutional Changes on Requirements for Vacant Professorships in Germany: A Quantitative Content Analysis of Employment Advertisements. *Working Papers in Higher Education Studies*, 1 (2). <http://www.wphes-journal.eu/index.php/wphes/article/view/15>.
- Kleimann, Bernd & Klawitter, Maren (2017a). Berufungsverfahren an deutschen Universitäten aus Sicht organisationaler Akteure. *Beiträge zur Hochschulforschung*, 39 (3/4), 52-73. Verfügbar unter: <http://www.bzh.bayern.de/uploads/media/3-4-2017-Kleimann-Klawitter.pdf>
- Kleimann, Bernd & Klawitter, Maren (2017b). An Analytical Framework for Evaluation-Based Decision-Making Procedures in Universities. In Jeroen Huisman & Malcolm Tight (Eds.), *Theory and method in higher education research* (First edition). Bingley, UK: Emerald Publishing Limited.
- Kleimann, Bernd, In der Smitten, Susanne & Klawitter, Maren (2015). Forschungserfahrung als Berufungskriterium. Anforderungen an zukünftige Professorinnen und Professoren. *Forschung und Lehre*, 2015 (8), 644-645.
- Klein, Tamar, Kraatz, Alexandra & Hornbostel, Stefan (2012). Begutachtungsprozesse im Wettbewerb um Drittmittel. Das Beispiel der Sonderforschungsbereiche. *Die Hochschule*, 2012 (2), 164-182.
- Krüger, Anne K. & Reinhart, Martin (2017). Theories of Valuation-Building Blocks for Conceptualizing Valuation Between Practice and Structure. *Historical Social Research/Historische Sozialforschung*, 263-285. <https://doi.org/10.12759/hsr.42.2017.1.263-285>.
- Krüger, Anne K. & Reinhart, Martin (2016). Wert, Werte und (Be)Wertungen – Ansätze für eine Soziologie des Wertens und Bewertens. *Berliner Journal für Soziologie*. <https://doi.org/10.1007/s11609-017-0330-x>.
- Lamont, Michèle (2012). Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology*, 38, 201-221. <https://doi.org/10.1146/annurev-soc-070308-120022>.
- Luhmann, Niklas (1983). *Legitimation durch Verfahren*. Frankfurt am Main: Suhrkamp.
- Manville, Catriona, Guthrie, Susan, Henham, Marie-Louise, Garrod, Bryn, Sousa, Sonia, Kirtley, Anne, Castle-Clarke, Sophie & Ling, Tom (2015). *Assessing Impact Submissions for REF 2014: An Evaluation*. RAND Europe: Cambridge, UK.
- Meier, Frank, Peetz, Thorsten & Waibel, Désirée (2016). Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung. *Berliner Journal für Soziologie*, 26(3), 307-328. <https://doi.org/10.1007/s11609-017-0325-7>.
- Möller, Torger, Antony, Philipp, Hinze, Sybille & Hornbostel, Stefan (2012). Exzellenz begutachtet. Befragung der Gutachter in der Exzellenzinitiative. *iFQ-Working Paper No.11*. Berlin.



- Neidhardt, Friedhelm (2016). Selbststeuerung der Wissenschaft durch Peer-Review-Verfahren. In Dagmar Simon, Andreas Knie, Stefan Hornbostel, & Karin Zimmermann (Hrsg.), *Handbuch Wissenschaftspolitik* (2. Aufl., S. 261-277). <https://doi.org/10.1007/978-3-658-05455-7>.
- Niewöhner, Jörg (2015). Infrastructures of Society, Anthropology of. *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition (12), 119-125. <https://doi.org/10.18452/19371>.
- Niewöhner, Jörg (2014). Perspektiven der Infrastrukturforschung: Care-ful, relational, ko-laborativ. In Diana Lengersdorf & Matthias Wieser (Hrsg.), *Schlüsselwerke der Science & Technology Studies* (2014. Aufl., S. 341-352). Wiesbaden: VS Verlag.
- Olbrecht, Meike (2014). *Entscheidungsfindungsprozesse von Gutachtergruppen*. Humboldt-Universität zu Berlin, Philosophische Fakultät III. <https://doi.org/10.18452/16946>.
- Potthast, Jörg (2017). The Sociology of Conventions and Testing. In Claudio E. Benzecry, Monika Krause, & I. Ariail Reed (Eds.), *Social Theory Now*. The University of Chicago Press.
- Reinhart, Martin (2012). *Soziologie und Epistemologie des Peer Review*. (1. Aufl.) Baden-Baden: Nomos.
- Rosanvallon, Pierre (2018). *Die gute Regierung* (1. Aufl.; M. Halfbrodt, Übers.). Berlin: Suhrkamp Verlag.
- Schimank, Uwe (2016). Governance der Wissenschaft. In Dagmar Simon, Andreas Knie, Stefan Hornbostel & Karin Zimmermann (Hrsg.), *Handbuch Wissenschaftspolitik* (2. Aufl., S. 39-57).
- Star, Susan Leight (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, 43 (3), 377-391. <https://doi.org/10.1177/00027649921955326>.
- Star, Susan Leight & Bowker, Geoffrey C. (2002). How to Infrastructure. In Leah A. Lievrouw & Sonia Livingstone (Eds.), *Handbook of New Media: Social Shaping and Consequences of ICTs* (pp. 151-162).
- Star, Susan Leight & Ruhleder, Karen (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, 7 (1), 111-134. <https://doi.org/10.1287/isre.7.1.111>.
- Weingart, Peter (2001). *Die Stunde der Wahrheit?: Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft*. Weilerswist: Velbrück Wissenschaft.
- Weller, Ann C. (2001). *Editorial Peer Review: Its Strengths and Weaknesses*. Information Today, Inc.
- Wilholt, Torsten (2012). *Die Freiheit der Forschung: Begründungen und Begrenzungen*. Originalausgabe. Berlin: Suhrkamp Verlag.

### *Anschriften der Autor\*innen:*

Dr. Cornelia Schendzielorz, Humboldt-Universität zu Berlin, Institut für Sozialwissenschaften, Unter den Linden 6, 10099 Berlin, E-Mail: [schendzc@hu-berlin.de](mailto:schendzc@hu-berlin.de) und Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, Abteilung Forschungssystem und Wissenschaftsdynamik, E-Mail: [schendzielorz@dzhw.eu](mailto:schendzielorz@dzhw.eu).

Prof. Dr. Martin Reinhart (Korrespondenzautor), Humboldt-Universität zu Berlin, Institut für Sozialwissenschaften, Unter den Linden 6, 10099 Berlin, E-Mail: [martin.reinhart@hu-berlin.de](mailto:martin.reinhart@hu-berlin.de).

Regine Paul

# Analyse and rule? A conceptual framework for explaining the variable appeals of ex-ante evaluation in policymaking

## Abstract

This article integrates disparate explanations for increasing (but variable) turns to ex-ante policy evaluation, such as risk analysis, across public administrations. So far unconnected silos of literature – on policy tools, policy instrumentation, the politics of evaluation and the political sociology of quantification – inconsistently portray ex-ante evaluation as rational problem-solving, symbolic actions of institutional self-defence, or (less often) political power-seeking. I synthesise these explanations in an interpretivist and institutionalist reading of ex-ante evaluation as contextually filtered process of selective meaning-making. From this methodological umbrella emerges my unified typology of ex-ante evaluation as instrumental problem-solving (I), legitimacy-seeking (L) and power-seeking (P). I argue that a) these ideal-types co-exist in policymakers' reasoning about the expected merits of ex-ante evaluation, whilst b) diverse institutional contexts will favour variable weightings of I, L and P in policymaking. By means of systematisation the typology seeks to inspire an interdisciplinary research agenda on varieties of ex-ante evaluation.

**Keywords:** policy tools, policy evaluation, politics of evaluation, institutionalism, interpretive policy analysis

## Zusammenfassung

*Ein integrierter Erklärungsansatz für die variable Hinwendung zu ex-ante Evaluierungstools in der Politikgestaltung*

Der Beitrag entwirft einen integrierten Analyserahmen für das Phänomen der zunehmenden und zugleich variablen Hinwendung zu Instrumenten der ex-ante Politikevaluation, beispielsweise der risikobasierten Entscheidungsvorbereitung, in vielen politisch-administrativen Systemen. Bisher unverbundene und oft kontrastierende Literaturen – zu *policy tools*, *policy instrumentation*, Evaluationspolitik, aber auch der politischen Soziologie der Quantifizierung – konzeptualisieren ex-ante Evaluation als rationale Problemlösung, symbolische Handlung institutioneller Selbstverteidigung oder (seltener) politisches Machtspiel. Der Beitrag synthetisiert solche scheinbar konträren Erklärungsansätze in einer interpretivistischen und institutionalistischen Lesart, nach welcher Akteure des politisch-administrativen Systems den *Toolkits* der ex-ante Evaluation drei verschiedene Leistungsversprechen – instrumentelle Problemlösung, Legitimitätsbeschaffung und Machtgewinn (*ILP-Typologie*) – zuessen. Der Beitrag argumentiert, dass a) diese drei Idealtypen in der Bedeutungszuweisung an ex-ante Evaluationsinstrumente immer ko-existieren, b) diverse institutionelle Kontextbedingungen allerdings eine unterschiedliche Gewichtung der respektiven I-, L- und P-Interpretationsrahmen in der Politikgestaltung begünstigen. Durch Systematisierung bisher nebeneinanderstehender Forschungsstränge soll die integrierte Typologie eine interdisziplinäre Forschungsagenda zu Variation von Politikevaluation und Evaluationspolitik inspirieren.

**Schlagworte:** Politikinstrumente, Politikevaluation, Evaluationspolitik, Institutionalismus, interpretative Policyanalyse

# 1 Introduction

For some decades now, governments and public administrations across the globe have invested heavily in analytical tool-sheds for assessing policy options *before* decision-making. Famous examples include President Reagan's Executive Order 12291 of 1981, requiring that all US legislation be pre-examined with cost-benefit analysis (CBA) (Posner, 2001), or the English demand that all regulatory enforcement activities be prioritised through systematic risk analysis (Hampton, 2005). There is also increased promotion of such tools by international organisations (for risk analysis see: Organisation for Economic Co-Operation and Development, 2010).

Scholars have recently started examining ex-ante evaluation tools, too (e.g., Atkinson, 2015; Black, 2005; Dunlop & Radaelli, 2016; Radaelli, 2005; 2010; Rothstein, Borraz & Huber, 2013) and thus revive their foundational role in the birth of the policy sciences in the tradition of Harold D. Lasswell<sup>1</sup>. And yet: we so far find rather contradictory and scattered conceptual answers to the question why exactly such tools appeal to policymakers and how their appeal varies across contexts:

- On the one hand, standard definitions of so-called “policy-analytical methods” (I adopt the term from Turnpenny, Jordan, Benson & Rayner, 2015, p. 4) in policy studies suggest they fulfil a predominantly instrumental role in policymaking far off political attention or infighting<sup>2</sup>. CBA and its cousins count as precursors for policy formulation and adoption. By analysing policy problems and proposed solutions *before* decision-making they “[generate] options about what to do about a public problem” (Howlett, 2011, p. 29).
- On the other hand, such a technocratic and instrumental conceptualisation of the appeal of ex-ante evaluation is fiercely contradicted in the literature on policy evaluation and instrumentation (Weiss, 1973, 1993; Linder & Peters, 1989; Peters, 2002; Lascombes & Le Galès, 2007; Kassim & Le Galès, 2010; Raimondo, 2018). Scholars in these fields have amply highlighted the political nature of the design, choice and use of evaluation and policy tools. Read jointly, as this article will, these literatures suggest that ex-ante evaluation involves what I call *positional struggles* over organisational legitimacy, power and political influence in a policy domain.
- In addition, both the policy studies and the policy tools literature disregard insights from the political sociology of quantification, though it offers great analytical mileage for specifying the argumentative underpinnings of legitimacy and power-oriented turns to ex-ante evaluation among public administration actors.

This article argues that we need a systematic integration of insights from across these so far unconnected silos of literature to explain the increasing but contextually variable rise of ex-ante evaluation. Building on the issue's reflections on ‘varieties of evaluation’, I contribute a unified analytical framework for explaining the adoption of policy-analytical methods as stemming from administrators'/policymakers'<sup>3</sup> contextualised perceptions of their likely ability to pursue different sets of goals. I propose a unified typology<sup>4</sup> of instrumental problem-solving (I), legitimacy-seeking (L) and power-seeking (P) to capture the duality of instrumental and more political adoptions of ex-ante evaluation frameworks. My interpretivist premise is that to adopt ex-ante evaluative tools is a powerful act of interpreting and pre-structuring the socio-political world as it is to be known and governed against existing structural conditions.

To develop the conceptual-analytical argument, the article first reviews existing typologies of policy tools and evaluation uses. On that basis, it conceptualises the adoption of ex-ante evaluation tools as *multidimensional* acts of *institutionally embedded* but *selective meaning-making* on the part of policymakers (section 2). On that terminological and conceptual footing, the article then carves out the three ideal types (section 3-5) – my ILP typology – and develops expectations about the institutional contexts in which meaning-making actors will vest instrumental, legitimacy or power-seeking goals in the adoption of ex-ante evaluation. I illustrate the logic of each ideal type through the example of risk-analytical frameworks, a case well-charted in scholarly work. The article last maps out how the unified typology can stimulate a future research agenda on varieties of ex-ante evaluation by examining interactions between and the dynamic evolution of types over time (section 6).

## 2 Ex-ante evaluation as *institutionally embedded* but *selective* act of meaning-making

To anchor my conceptual approach to policymakers' motivations for adopting ex-ante evaluation I start with a brief review of existing typologies of policy tools and evaluation. This also highlights this article's contribution in bridging disparate literatures with a distinct interpretivist and institutionalist framework.

### 2.1 Typologies of policy tools and evaluation uses

In the policy tools literature, typologisation has long focused on different kinds of tools for achieving a nominal policy goal. This is exemplified in the discussion of “carrots, sticks, and sermons” (Bemelmans-Videc, Rist & Vedung, 2011) but also Christopher Hood's (1983) NATO typology of substantial policy tools with four different vantage points for behaviour change (nodality, authority, treasure and organisation). These typologies assume that civil servants within the public administration have a rational view on how a specific tool will achieve a policy goal at the lowest cost, least visibility and most effective targeting of regulatees (for example, incentives for self-regulation vs. sanctioned prohibitions). While such theorisation implies that context will shape perceptions of the appropriateness of tools (both among regulatees and public administration actors), it does not theorise context as an independent variable.

Calls to contextualise policy tool choice in institutionalist perspectives include Peters's argument that tools do not pre-exist ‘conceptualised and defined’ and ready to be picked for problem-solving. Instead, they constitute ‘their own politics’ in interaction with empirical settings (Peters, 2002, p. 553). A more recent discussion of tool choice (Capano & Lippi, 2017) confirms that “neither legitimacy nor instrumentality is objectively given, but is perceived by decisionmakers in a subjective manner, depending on the context in which they act and make decisions” (p. 271). Their typology systematises decision-makers' perception of tools as an attempt to balance between the expected instrumentality of a tool and its perceived legitimacy. Missing, however, is a systematic integration of power-political drivers of tools choice. While the authors acknowledge that a tool “allocates opportunities and duties, creates values or symbols, as well as

stakes, and also distributes costs and benefits” (ibid., p. 279), they offer no explanation of how power-related perceptions might variably interact with policymakers’ instrumental and legitimacy motivations.

Claudio Radaelli’s (2010) discussion of ‘four images’ of regulatory impact assessment (RIA) (2010) distinguishes rational policymaking (=instrumental), symbolic action (=legitimatory), control of the bureaucracy and public management reform. Control of the bureaucracy type arguments, however, suggest a power-political dimension with actors in multi-level politico-administrative regimes negotiating the room for manoeuvring between policy decisions and implementation (Radaelli, 2010, pp. 169, 172 f.). It is less clear how public management reform forms a distinct ideal type, as arguably the goals of such reform have been well shown to connect to all of the other ideas (rational problem-solving, control of the bureaucracy and legitimacy symbolic action) (Pollitt & Bouckaert, 2011). Convincingly, Radaelli develops specific hypotheses about how variable institutional contexts – such as dominant ‘administrative models’, the degree of decentralisation, or the power of the central executive – shape uses of regulatory impact assessment (see sections 3-5).

Critical evaluation studies produced context-sensitive conceptualisations of policymakers’ evaluation use (Weiss, 1973; 1993). Political contexts shape the extent to which evaluations are used “(a) *instrumentally*, to give direction to policy and practice; (b) *politically or symbolically*, to justify pre-existing preferences and actions, and (c) *conceptually*, to provide new generalizations, ideas, or concepts that are useful for making sense of the policy scene.” (Weiss, Murphy-Graham & Birkeland, 2005, p. 13, highlight added) However, this taxonomy, based on extensive empirical research, remains disconnected from systematic theorisation of how institutional contexts shape the charted variety. Political science readings of evaluation provide such comparative systematisation. Christina Boswell (2008, p. 471), for example, argues that the prevalence of a) legitimacy orientations and/or b) policymakers’ attempts to substantiate their pre-existing policy preferences over c) rational-instrumental perspectives on expert knowledge depends on the perceived stability and source of an organisations’ legitimacy (claim a) as well as the degree of contention and mode of settlement with regard to policy goals (claim b). Read jointly, existing typologies highlight the *multidimensionality* of policymakers’ perceptions of policy tools/evaluative frameworks and (to a lesser extent) the relevance of institutional contexts for shaping these perceptions.

## 2.2 An interpretivist framework for building a unified typology

I propose an interpretivist policy analysis framework (e.g. Stone, 2012; Yanow & Schwartz-Shea, 2006) to explain policymakers’ variable (sets of) motivations for adopting tools of ex-ante evaluation. This means 1) to treat the diversity of motivations described in existing typologies not as mutually exclusive but as assemblages of meaning in policymakers’ sense-making processes, and 2) to systematise explanations of variety via a focus on the institutional embeddedness of sense-making processes. Rather than assuming instrumental problem-solving to be a dominant motivator of ex-ante evaluation, analysts must deconstruct its empirical relevance in a *multidimensional semiotic space* through careful interpretive analysis. To systematise the relevant semiotic dimensions as idealtypical yardsticks in a unified typology is the first key contribution of this article.

In so doing my interpretive perspective contributes, second, a nuancing of expected variety. Interpretivists have not only stressed the ambiguity and multiplicity of meanings vested in policies (and tools) but also the *co-existence of substantial interpretations* of policies with *perceptions about policies*' implications for *positional struggles* in a domain (Béland, 2009; Carmel, 2019). As discussed above, ex-ante evaluation's perceived appeal as a tool in power struggles has been rather underdeveloped in existing typologies. My typology therefore expands existing work – usually involving an instrumental and legitimacy-related (or symbolic) dimension – with a distinct focus on power-seeking.

Third, I contribute a set of expectations on how policymakers' perceptions of ex-ante evaluation *interact with their specific institutional contexts*. Institutions are understood here as both formal (as in the constitutional set-up of a polity and formal interactions between key actors) and informal (specific norms about state-market interactions etc.) rules of interactions in a given context which shape, according to various readings of neo-institutionalism, actors' preferences, senses of appropriateness as well as fueling path dependencies over time (Hall & Taylor, 1996). These include, for the purposes of our debate, key markers of 'politico-administrative regimes' (Pollitt & Bouckaert, 2011), including constitutional relationships between actors in unitary or federalised polities, degrees of centralisation or decentralisation, and the formal relationships of accountability and competition that emerge from such institutional diversity. They also concern the influence of dominant policy types and policy styles in diverse settings. As part of my conceptualisation around the ILP typology (sections 3-5), I theorise how policymakers' perceptions of ex-ante evaluation will vary alongside diverse institutional contexts and their attempts of negotiating (and potentially also re-make) their own position within these contexts.

### 3 Ideal type I: instrumental problem-solving

Studies on the policymaking process and policy change but also rational choice inspired accounts of policy(-analytical) tools highlight the instrumental appeal of CBA and its cousins in fulfilling pre-set policy goals. They also specify how exactly ex-ante evaluation triggers policymakers' problem-solving orientations through their distinct ability of screening possible solutions to a policy problem before decision-making.

#### 3.1 Policy tools as instrumental problem-solvers

Influential theories of the policy process and policy change situate policy tools – and policy-analytical tools – near the bottom within a hierarchy of scales of change. This is true for both the concept of policy paradigm change (Hall, 1993) and the advocacy coalitions framework (ACF) (Sabatier & Jenkins-Smith, 1993) which design causally complex but yet *hierarchically ordered* heuristics of the policy process with a rather marginal and unpolitical role for tools (let alone analytical tools). Tool choice, design and calibration appear as *functionally subordinated* to politically inaugurated policy goals and wider ideational frameworks. Indeed, in Hall's conceptual world of policy paradigms, tools belong in the realm of rather apolitical 'normal policymaking' occur-

ring far off the (assumably analytically more interesting) scenes of (rare) sweeping paradigmatic change. It is in that sense that a hierarchal view of policy change “remains the essence of *positivistic* notions of *rationality* in policy analytical theory” even in ideational scholarly work otherwise well able to capture causal complexity beyond linear policy models (Béland & Howlett, 2016, p. 393, highlight added).

Scholars of policy tools echo the instrumental view on policy tools when assuming a neat alignment of tools with pre-set goals. This view dominates Hood’s seminal *The Tools of Government* where he writes that “it is by applying its tools that government makes the link between wish and fulfilment”<sup>5</sup> (Hood, 1983, p. 8). Those adopting methods of ex-ante evaluation can hence be expected to seek enforcement of this link by founding each decision on analytical evidence *before* the fact. This assumed instrumentality of policy-analytical methods in the policymaking process also resonates with the idea of an “anticipatory policy style” where (analytical) expertise rationalises bureaucratic decision-making and prevents knee-jerk ‘reactive’ responses to public opinion (Richardson, Gustafsson & Jordan, 1982; Howlett & Tosun, 2018).

Focusing on the instrumental relationship between policy purposes (ends or goals) – the normative decisions about what a policy should aim to achieve – and the means selected to achieve an end (tools) is still in vogue. A recent study on policy tool change in the European Union (EU) captures policy instruments with Hood’s (then) thirty-year-old definition: as “means and techniques in order to achieve policy goals” (Bähr, 2010, p. 15). Once decided upon in political struggles, it seems in such accounts, dominant ideational “prisms through which policymakers [see]” (Hall, 1993, p. 279) policy problems trickle down to determine their perception of the likely performance of (analytical) tools.

### 3.2 Ex-ante evaluation as solution screening

The framing of evaluative tools in other parts of the literature – the studies discussed so far do not include evaluation tools – joins the instrumental proposition. CBA, risk analysis and RIA, for example, count – according to scholarly observers (Viscusi, 1983; Baldwin & Veljanovski, 1984; Sunstein, 2000; Posner, 2001; Majone, 2010; Radaelli, 2010) – as evaluation mechanisms for establishing, *before* decision-making, whether proposed (subsets of) policies are likely to achieve policy goals (effectiveness), whether they can do so in a cost-effective manner (efficiency) and without undesired adverse consequences (impact). In the instrumental reading, analytical tools thus serve as screening devices to identify rational policy solutions.

Giandomenico Majone, for example, argues that to apply the same analytical formula across regulatory issues<sup>6</sup> will make decision-making more consistent, avoid over- and under-regulation and equalise the treatment of regulatees (also see Viscusi, 1983, Sunstein, 2002). Policymaking and enforcement thus “become technical matters of calculation, policy-oriented scientific discourses that transcend political contests” (Tombs & Whyte, 2013, p. 67). Whilst related de-politicisation agendas have been rightly critiqued, for example in the case of risk analysis (e.g., Shrader-Frechette, 1990; Stone, 2012), the ability of ex-ante evaluation to ‘rationalise’ decision-making may well be instrumental from a problem-solving viewpoint.

Exemplary is the use of risk analysis in the United Kingdom (UK). Adhering to the cost-effectiveness goals of ex-ante evaluation, a parliamentary review suggests regulators

across all policy domains use “risk assessment to concentrate resources on the areas that need them most” (Hampton, 2005, p. 7). Since the approach gained statutory force in 2008 through the “Regulators’ Compliance Code” (Rothstein et al., 2019), UK policymakers must escalate their interventions proportionally with risk level. An enthusiastic promoter of this approach, the Health and Safety Executive turned to ex-ante evaluation already in the late 1980s to ensure that the cost, time and effort required to achieve pre-set regulatory goals (work risk reduction in the case of the HSE) be not grossly disproportionate to the benefit gained in its regulations and enforcement policies (*ibid.*). This example provides, at first sight, a textbook-like adoption of instrumental ex-ante evaluation, but its re-consideration throughout the article equally illustrates the multiplicity of rationales policymakers and policy operators vest in ex-ante evaluation tools.

### 3.3 Institutional contexts to favour instrumental appeals

An instrumental perspective on ex-ante evaluation is likely to emerge in contexts where policymakers perceive problem-solving as their genuine goal. This claim may seem mundane, but its relevance will become obvious in distinction from non-instrumental goals of evaluation (legitimacy-seeking and power-seeking) in the subsequent sections. Based on existing typologies of policymaking/administrative decision-making I theorise the institutional conditions to favour policymakers’ instrumental interpretations of ex-ante evaluation.

The classic policy styles distinction (Richardson, Gustafsson & Jordan, 1982, p. 13) suggests analytical tools’ currency to be rising in settings with an anticipatory policy style whereas those with reactive policy styles will not use ex-ante evaluation for instrumental reasons (they may well use analysis to legitimate decisions ex-post, of course). While the original typology has been theoretically and empirically extended (Howlett & Tosun, 2018), comparative public administration reform research indicates persisting variation as to the relative openness of different politico-administrative systems to expertise and sciences-led bureaucratic decision-making in line with policy style type arguments (Pollitt & Bouckaert, 2011). For example, in his comparative analysis of RIA, Radaelli suggests that this tool will emerge as a form of rational problem-solving in contexts with a ‘strong evaluation culture’ (2010, p. 174).

In a similar vein, Theodore Lowi’s policy types (1972) associate regulatory policies with the strongest role for expertise (this could include pre-decision analysis) while distributive and redistributive policies follow interest group and partisan politics logics, respectively. Even though very much focused on US-American policymaking, Lowi’s typology implies that ex-ante evaluation as instrumental problem-solving might feature strongly in cases of regulatory policymaking, either where a domain leans towards regulatory interventions as a whole and/or because the political setting overall is prone to use regulation (Majone, 1994; Moran, 2003). Exemplary is, again, the British administration which, since the 1990s (Black, 2005; Dodds, 2006), has sought to respond to “alleged over-regulation, legalism, inflexibility and an alleged absence of attention being paid to the costs of regulation” (Hutter, 2005, p. 1) with the development of analytical frameworks for decision-making.

As resource pressure, notions of regulatory crisis and related modernisation agendas have been wide-spread phenomena (Pollitt & Bouckaert, 2011) the instrumental



promises of ex-ante evaluation are likely to gain purchase elsewhere, too. Indeed, the discussion of reform trajectories in different politico-administrative regimes suggests the role of evaluation to be particularly relevant in Neo-Weberian state modernizers (ibid., pp. 168 f.). Accordingly, New Public Management reformer regimes, associated with strict principal-agent-accountability relationships, are likely to use analysis as one of the tools to aid politicians set strategic policy goals and increase the political control over the bureaucracy (as for RIA in the UK, Radaelli, 2010). In the absence of the control requirements associated with agencification (Döhler, 2002) policymakers within a hierarchically-integrated administration will rather adopt ex-ante evaluation as an expertise-driven technical implementation aid.

## 4 Ideal type L: legitimacy-seeking

This section draws together insights from various branches of research which adopt a sociological institutionalist perspective on policy tools, evaluation and specific methods of quantification (such as risk analysis or CBA). Inspired by neo-institutionalist sociology, such research interprets the adoption of ex-ante evaluation as policymakers' endogenous sense-making about the 'appropriateness' of different paths of action (March & Olsen, 1989, p. 23). Though few studies (Black, 2005; Rothstein, Huber & Gaskell, 2006; Radaelli, 2010) address the question of ex-ante forms of evaluation and its drivers explicitly, the joint review enables me to carve out a 'legitimacy-seeking' ideal type of ex-ante evaluation.

### 4.1 Policy tools and evaluations as legitimacy boosters

Institutionalist scholars of policy tools have, for some time now, claimed that considerations of "institutional self-interest" and the stability and legitimacy of the respective administrative organisation (Peters, 2002, pp. 557 f.) co-motivates policymakers' tool adoptions. The literature on the politics of instrumentation specifies this general proposition by arguing that any policy tool implies specific regulatory concepts which can help (re-)structure (strategically or unintendedly so) both the accountability relationships between administrators and regulatees and the ascription of authority in a policy domain (Lascoumes & Le Galès, 2007). In other words: tools come with distinct ideas about who is to be held accountable, affecting the perception of *legitimacy* of an administration, as well as ideas about who should legitimately be in charge.

Correspondingly, the instrument constituencies literature assumes actors from across most diverse organisations (government departments, international organisations, NGOs, interest groups, scientific associations, the media etc.) jointly elaborate policy tools, analytical models and technical solutions (e.g. specific risk algorithms, emission trading certificates or funded pension schemes) and promote – in a distinctly *endogenous* process of knowledge, identity and 'preference' generation – a policy tool as suitable and appropriate problem solution (Voß & Simons, 2014; Béland & Howlett, 2016; Simons & Voß, 2017).

Legitimacy-seeking claims have long been central for scholars of the politics of evaluation, too. They argue that 'political pressures' taint any purely instrumental con-

ceptualisation or use of analytical frameworks (Weiss, 1973). Policymakers can adopt evaluations selectively and strategically, they argue, to “provide support for policies decided on the basis of intuition, professional experience, self-interest, organizational interest, a search for prestige... [or] to buttress an existing point of view” (Weiss, Murphy-Graham & Birkeland, 2005, pp. 13 f.; similar points are made in the newer discussion on the politics of evaluation by Bjørnholt & Larsen, 2014, p. 407; Raimondo, 2018, p. 34). Whilst Carol Weiss, Erin Murphy-Graham and Sarah Birkeland (2005) examine mainly ex-post evaluation, comparative scrutiny of RIA confirms similar “symbolic” – or in my typological language legitimacy – uses of ex-ante evaluation tools in EU member states (Radaelli, 2005; 2010).

## 4.2 ‘Trust in numbers’ and institutional risk management

Key works on the (political) sociology of quantification help us go beyond a general assumption of legitimacy efforts and symbolic policy-making. This so far often ignored literature (from the viewpoint of policy and policy tool studies) specifies how and why ex-ante evaluation appeals as legitimacy booster: ‘trust in numbers’ rallies external support for an organisation and enables it to risk-manage its own mandate vis-à-vis accountability demands.

Theodore Porter (1995), for example, describes quantification in administrative decision-making as a strategy for organisations to counter ascribed legitimacy deficits. Where one’s authority to decide over specific policy issues meets doubt quantification becomes an important “technology of trust” (ibid., p. 15):

“A decision made by the numbers (or by explicit rules of some other sort) has at least the appearance of being fair and impersonal. Scientific objectivity ... provides an answer to a moral demand for impartiality and fairness. Quantification is a way of making decisions without seeming to decide. Objectivity lends authority to officials who have very little of their own.” (ibid, p. 8)

Porter’s vivid example is the introduction of the first ex-ante evaluation tool there was: CBA in the early 19<sup>th</sup> century by the U.S. Army Corps of Engineers (a body responsible for planning and executing drainage, levee and canal projects). Somewhat paradoxically perhaps, the expert regulator turned to ex-ante evaluation precisely because its bureaucratic expert regulation did not enjoy enough legitimacy within the political system. CBA became a strategy of convincing Congress, the Courts, and potential opponents in competing departments that the Corps’ funding decisions on water projects were reasonable, fair, and politically unbiased. Such isomorphic dynamics also feature for our illustrative example of risk analysis: in the UK case, its “apparent objectivity and transparency could be used to explain the allocation of resources, in a way which was well tested and trusted by the business community” (Hutter, 2005, pp. 2 f.).

Importantly, however, the political sociology of quantification goes beyond a conceptualisation of “symbolic” uses of ex-ante evaluation. When Porter argues that the wide-spread trust in numerical evidence enabled the Corps to position itself as legitimate regulator, I suggest, he reaches beyond mere symbolism and window-dressing. Like the politics of evaluation literature he highlights the *strategic institutionalising* role of analytical frameworks for organisations in complex multi-lateral policy domains (Raimondo, 2018; Jankauskas & Eckhard, 2019). Scholars of risk-based decision-making perceive risk analysis as chief organisational strategy for such institution-

alising processes. Ex-ante analysis helps specifying one's mandate, activities, and decision-making processes vis-à-vis the (perceived) expectations of those who can effectively blame an organisation<sup>7</sup>. In addition to abstract legitimacy gains, ex-ante evaluation hence helps policymakers define "the parameters of blame" (Black, 2005, p. 520) or their own "limits of governance" (Rothstein, Borraz & Huber, 2013). Organisations pro-actively manage their own "institutional risks" by stating the limits of their goals and activities upfront, based on forms of analysis that count as rigorous and objective (ibid., pp. 99-103).

An example for such legitimacy-seeking is once more the British Health and Safety Executive. When under pressure to cut costs during recent austerity programs, the agency shifted the focus of its long-established risk-analytical framework from cost-effective problem-solving to one of managing the parameters for being blameable for non-detected health and safety risks at workplaces (Demeritt, Rothstein, Beaussier, & Howard, 2015, on a similar argument for German work safety inspection see Paul, 2016).

#### 4.3 Institutional contexts to favour legitimacy-seeking appeals

Two strands of institutional conditions will shape the extent to which legitimacy-seeking matters in the adoption of ex-ante evaluation. The first is a strong perception among public administration bodies as being confronted with severe institutional risks (Rothstein, Huber & Gaskell, 2006), for example due to increased budgetary pressures, a heightened problem load (for example after a focusing event) or perceived regulatory failure and scandals (Hutter, 2005). Comparative risk regulation research confirms that pressure from business and public opinion, as well as the allocation suggests of accountability for dealing with residual risk, shapes the extent of institutional risk perceptions in a public administration body (Hood, Rothstein & Baldwin, 2001).

Second, a legitimacy-seeking perspective is more likely to emerge where public administration bodies experience structural pressure on their mandate. Institutional competition between actors over a regulatory mandate is likely, for example, to create continuous legitimisation pressure for a respective mandate. Such competition is heightened in multi-level settings where the involvement of different levels of decision-making (subnational, national, EU) but also the inclusion of non-state actors in decision-making (corporatist and self-regulatory regimes) reduce the scope for administrations to govern unilaterally (Loughlin, 2017). Radaelli's RIA studies suggest that minority and coalition governments (Sweden and the Netherlands in his case) are more likely to use ex-ante evaluation for 'symbolic' reasons (2010, p. 175). He sees a similar defensive dynamic for the EU Commission which seeks to "defend its proposals vis-à-vis a plethora of aggressive stakeholders" by the help of seemingly rigorous ex-ante evaluation (ibid.).

## 5 Ideal type P: power-seeking

The third ideal type forms around the assumption that actors might consider ex-ante evaluation as a weapon in power-positional struggles. Based on a joint venture of the literatures on policy instrumentation, the politics of evaluation and the political sociology of quantification, I show that analytical frameworks can appeal as tools for restructuring power relations both between different actors and administrative levels *and* as to the vision of how state, market and society ought to interact.

### 5.1 Policy instrumentation as struggle over influence

To be clear, I do not argue that mainstream policy analysis disregards political power in policymaking. Rather, as section 3 highlighted, many analyses treat policy tools more generally (Hood, 1983; Hall, 1993; Sabatier & Jenkins-Smith, 1993), and ex-ante evaluation in particular (Viscusi, 1983; Posner, 2001; Majone, 2010), as rather apolitical and instrumentally subordinated to the outcomes of political struggles over policy goals.

Some scholars of policy tools have, however, acknowledged that the design, choice and use of tools is embedded in positional struggles over “whose ox will be gored or whose nest will be feathered” (Peters, 2002, p. 554). The instrumentation literature, in particular, argues that policy tools always also (re-)define roles and relationships in a domain. This is because they imply, at their very core, specific theories about the “relationship between the governing and the governed” as well as a “condensed form of knowledge about social control and ways of exercising it” (Lascoumes & Le Galès, 2007, p. 1). Not only do policy tools have unintended structuration effects but they can be promoted for how they promise to (re-)structure a policy domain and the relationship of actors within it. It is less relevant, for the analytical focus of this article that such promises may often be disappointed in practice. The instrument constituencies literature shares the notion of power-related positional struggles, arguing that endogenous constituency-building around instruments inherently entails an envisioning of “shift[s] in responsibilities between governmental agencies or organisational departments” (Voß & Simons, 2014, p. 739). In both literatures, policy tool adoptions thus (also) become politically motivated struggles over influence in policymaking in ways their conventional treatment as technocratic and instrumental fulfillers of policy goals overlooks (and in ways which may not easily be discerned empirically).

### 5.2 Evaluation tools and strategic influence-seeking

From a different angle, the politics of evaluation has amply highlighted evaluation frameworks’ appeal for actors who are “manoeuvring for advantage and influence” and seeking to enhance “the influence of a particular division or department” (Weiss, 1973, pp. 180 f.). In a later contribution, Carol Weiss (1993) specified that such “interagency and intraagency jockeying for advantage and influence” (p. 95) concerns both

- abstract questions of relative power within and between public administration bodies
- *and* substantial questions about whose values and normative premises get inscribed into an evaluation framework (p. 99).

Weiss's first point is detailed in actor-centered analysis of specific ex-ante evaluation projects, often situated in institutionalist conceptualisations of international bureaucracies such as the EU Commission. For example, a recent study on the Commission's 2015 Better Regulation Reform concludes that, even if sold as a technical device for instrumental problem-solving and increased accountability, "the reform has contributed to the strategic actorness of the Commission as an international bureaucracy" (Jankauskas & Eckhard, 2019, p. 683). Studies on RIA confirm that ex-ante evaluation tools may tilt the balance of power towards seemingly technocratic actors within the EU, such as the EU Commission's Secretariat (Radaelli & Meuwese, 2010, see similar arguments in the wider literature on the open method of coordination, e. g. Tholoniati, 2010). This also holds true for our example of risk analytical frameworks: the EU Commission has embraced common risk analysis strategically in several domains to expand its influence over member states' administrative decisions *de facto* in *de jure* weakly integrated domains (for flood risk management see: Paul, Boudier & Wesseling, 2016; for border control see Paul, 2017).

### 5.3 Quantification as a governmental technology of power

Weiss's second point – the question of whose norms receive promotion via the specification of an evaluation tool – attracts more attention in poststructuralist critiques of diverse forms of analysis and related scholarly claims of rationality (e. g. for risk analysis: Viscusi, 1983; Majone, 2010). Without space to review this wide-ranging literature here, let me selectively refer to the attempt to deconstruct risk analysis as a political project tied to neo-liberal governmentality and its beneficiaries (Ewald, 1986; O'Malley, 2004; Stone, 2012).

The political sociology of quantification discusses how analytical templates help constitute processes of (neo-liberal) commodification and simultaneous shifts of authority towards central state bodies. Alain Desrosières (2002), in his *Politics of Large Numbers*, stresses the mutually constitutive relationship between the growth of mathematical analytical models, common market formation and governmental authority over populations and subnational state entities. For example, *Statistik* in Germany did not coincidentally emerge in the 19<sup>th</sup> century when the territory was split up into over 300 individual states with highly diverse local identities: comparative analyses were perceived as a chief requirement for the desired creation of a common market, a national identity and unified nation-state. James Scott's discussion of the Napoleonic meter<sup>8</sup> – one of his examples of statutory attempts to increase 'legibility' for enhanced governability – further supports this power-related conceptualisation of analytical and quantification techniques. He argues that, as the meter replaced the locally highly fragmented measurement practices with a common standard, it induced not only a more effective common market but also enabled far-reaching administrative centralisation: "no effective central monitoring ... [was] possible without standard, fixed units of measurement" in post-revolutionary France (Scott, 1998, p. 30).

The example of risk-analytical evaluation frameworks suggests a similar governmentality logic to be at play today. In flooding policies, for example, the introduction of a common model for risk analysis in 2007 by the EU has induced a paradigm shift from security-oriented technical flood defence towards the much less risk-averse ideal

of flood risk management in France and Germany (Paul, Boudier & Wesseling, 2016). Such studies highlight the Commission's attempt to present the evaluative framework as neutral and a-political all while articulating a desire to change member states' own concepts and normative parameters for enforcement (without, however, interfering with their sovereign decision-making).

## 5.4 Institutional contexts to favour power-seeking appeals

Power-seeking around evaluation tools will be relevant for policymakers in any institutional setting that fragments and/or decentralises authorities policymaking and/or implementation. I suggest that it will be *pronounced in constitutional multi-level systems* where the competences for policymaking are constitutionally shared across different actors and where decentral execution is not simply granted (and withdrawn) at the goodwill of a *de jure* dominant centre (for the difference between unitary/federalised statehood vs. centralisation/decentralisation see Loughlin, 2017).

Firstly, as for the institutional condition of decentralisation more generally, well-known arguments about politicians' wish to increase political 'control in a bureaucracy' and overcome implementation gaps (see Dunsire, 1978; for RIA: Radaelli, 2010) apply in any setting where policy formulation/decision and implementation are delegated to actors other than those determining policy goal-setting. Control of decentral administrative actors is usually wielded either via ministerial hierarchies in regimes following the New-Weberian state model (such as Germany or France) or, in the New Public Management reform countries (such as the UK or the US), via strict accountability patterns between principal and agent (see Pollitt & Bouckaert, 2011; Paul, 2016 for German risk-based policy-making). Wherever there is a structural need for central level policymakers to ensure vertical policy control in the administrative setting, central level actors will likely value ex-ante evaluation frameworks as tools to align decentral autonomous activities with central-level priorities.

Procedural standards and policy-analytical tools in particular are widely conceived of as tools of softly-induced and non-coercive behaviour change which enable central governments to "pre-dispose" (Balla, 1998) subcentral actors' use of discretion (see the discussion of behavioural policymaking and „soft paternalism“: Strassheim & Beck, 2019). The soft manipulation of autonomous subcentral administrative activities through ex-ante evaluation is more likely in policy domains where central level actors – be it an EU Directorate General or a Federal Ministry – possess the actual power to coerce policy tools onto lower levels. The adoption of ex-ante evaluation tools within multi-level administrations will hence vary with the respective supra-national and national level government's power *and* appetite to shape subcentral policy execution.

This segues into my second claim of institutional variation. With a view on variable territorial state organisation, we can assume that power-seeking rationales vested in ex-ante evaluation in decentralised but unitary states will focus on confirming the status quo of power relations (central level authority as constitutionally granted *ultimo ratio*) rather than seeking to re-arrange formal decision-making powers *per se*. Instead, from a power-seeking perspective, the *political* currency of analytical evaluation will increase in contexts of constitutionally limited central control where neither hierarchy nor strict accountability bind sub-central executive decision-making. This is typically

(but variably so) the case of federal polities and those integrated in stronger forms of regional integration of policymaking, such as member states in the EU setting. As scholars of federalism and multi-level governance explain, substantial policymaking in such settings increases coordination requirements between actors and therefore also carries with it the need to constantly negotiate one's relative authority and power vis-à-vis others (Benz, 2009; Benz, Corcaci & Doser, 2016). I suggest that, in such settings, policymakers will therefore favour uses of analytical tools which promote their own realm as the *best locus* of policymaking and enforcement in multi-level settings.

## 6 A research agenda on varieties of ex-ante evaluation

This article systematised existing conceptualisations of how and why ex-ante evaluation appeals to public administrations. It provided an integrated conceptual framework with analytical space for instrumental problem-solving perspectives *alongside* more political explanations of legitimacy- or power-seeking. In line with those highlighting that any 'rift' between rationalist and constructivist views on the role of evidence in policymaking is 'obsolete' (Strassheim, 2017, p. 238), my typology includes instrumental *alongside* non-instrumental rationalities and suggests that their specific weighting is shaped against policymakers' perceptions of how to best navigate (and even restructure) their specific institutional contexts (summary in table 1). My unified analytical framework has at least four implications for how the research agenda on ex-ante evaluation in public administrations might move forward, especially in comparative endeavours.

*Table 1:* Three idealtypical appeals of ex-ante policy evaluation

|                                                   | Instrumental<br>problem-solving                                                                                                                                                                                     | Legitimacy-seeking                                                                                                                                                                                                                        | Power-seeking                                                                                                                                                                                                                          |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Logic of</i>                                   | consequentiality                                                                                                                                                                                                    | appropriateness                                                                                                                                                                                                                           | influence                                                                                                                                                                                                                              |
| <i>Motivation</i>                                 | wish to solve policy problem                                                                                                                                                                                        | wish to defend mandate                                                                                                                                                                                                                    | wish to expand influence                                                                                                                                                                                                               |
| <i>Expected performance of ex-ante evaluation</i> | analysis helps screen best solution before decision-making and thus increase effectiveness & efficiency of policymaking                                                                                             | 'trust in numbers' and appearance of rigorous analysis justifies one's decision-making & secures one's mandate against criticism/competition                                                                                              | analytical frameworks help predispose and control others' decision-making; analytical harmonisation can bypass political fragmentation or even induce harmonisation                                                                    |
| <i>Favourable Institutional contexts</i>          | dominance of anticipatory policy style; openness to scientific expertise, rationality drive and existence of strong evaluation culture in a policy domain; non-partisan and hierarchically structured civil service | strong perception of legitimacy pressure (e.g. through stakeholder pressure or public opinion, or competition with other actors over a mandate); templates for ex-ante evaluation available and activatable (e.g. via coordination for a) | coercive powers over/ ability to tilt incentive structures for other actors favour control of bureaucracy; multi-level governance settings produce desire to control policy substance and advance one's own influence vis-à-vis others |

*Source:* Own illustration.

First, if we understand the adoption of ex-ante evaluation as an institutionally embedded act of selective meaning-making, it follows that a nominally similar analytical method can display highly variable appeals across countries, over time, but also – and

this is often less attended to – across policy sectors *within the same jurisdiction*. This requires overcoming methodological nationalism in comparative research on ‘evaluation regimes’ (see Strassheim & Schwab in this special issue) in favour of cross-sectoral analysis and joins in the longer-standing call to treat policy domains as their own analytical category (Döhler, 2015).

Second, analytical attention in future research ought to be paid to *interaction patterns and dynamics* of I, L or P types of ex-ante evaluation. The unified typology would be grossly misconceived if it were taken to mean ‘everything matters’. Instead, it encourages ideal type-based analytical strategies for identifying patterns within complex variety. Rather than justifying a relaxation of comparative analytical rigor, my typology offers a steppingstone towards systematic cross-contextual analyses and further theorisation of ex-ante evaluation in policymaking: are there specific patterns of ILP interaction in specific domains or countries? Does ex-ante evaluation lean towards specific combinations of I, L, or P in any temporally staggered manner? Is one prerequisite or hindrance for another and, if so, in which specific institutional contexts? Indeed, my own work on risk analysis in the German public administration suggests that policymakers in multi-level governance settings turn to evaluation tools to clear the decks off legitimacy and power struggles in order to enable serious problem-solving in the first place (Paul, 2021/fc).

These are not only theoretically interesting queries. Rather, comparative work evidences that policymakers’ dominant orientations for tool choice durably affects the ways in which a tool is institutionalised in policymaking and unfolds implications for policy outcomes as well as the attribution of legitimacy and power in a domain (for RIA: Radaelli, 2010; for risk analysis Rothstein et al., 2019).

Third, comparative explorations with the ILP typology can help clarify the role of actors in the adoption and diffusion of ex-ante evaluation<sup>9</sup>. While my contribution did not provide a nuanced conceptualisation of actors (see endnote 3) idealtypical analysis with the ILP framework can well explore whether and why specific actors are inclined to develop a particular ideal-typical interpretation of an evaluative tool. Such actor-centered analysis might well find that civil servants, higher-rank political actors such as ministers or EU commissioners, external evaluators or technical experts in a domain adopt different interpretive frameworks on the merits of ex-ante evaluation and could further also examine their relative influence in decision-making.

Last, despite its focus on contextual variation and institutional patterning the ILP framework neither wants to nor can be a predictive model of ex-ante evaluation. The interpretivist paradigm treats complex contexts as dynamic infrastructures which can be mobilised or immobilised, amplified or muted in meaning-making processes (e.g., Stone, 2012; Yanow & Schwartz-Shea, 2006). While institutional environments shape the relative plausibility of instrumental, legitimacy and power-oriented perceptions of ex-ante evaluation, this context does not fully pre-determine actors’ interpretations of the benefits of such analytical methods. As the whole policy process, ex-ante evaluation methods, therefore, ought to be analysed as “context-in-use” (Wagenaar, 2011, pp. 110 f.); as the dynamic, always-preliminary and unpredictable results of contextually filtered but selective meaning-making processes.



## Notes

- 1 Scholarly elaborations on planning, programming and budgeting system models (PPBS) – an analytical approach to decide how to spend public money – stood at the cradle of a mushrooming policy-analytical industry in the US (prominently: Lasswell 1971; Turnpenny, Jordan, Benson & Rayner, 2015).
- 2 This judgement call before decision-making arguably also renders ex-ante evaluation less politically sensitive than ex-post evaluation of policy programmes.
- 3 For better legibility, this article uses the term policymakers as a shorthand for all actors in a politico-administrative regime involved in the adoption of ex-ante evaluation tools.
- 4 I discuss how this relates to existing typologies in section 2.
- 5 Hood (1983) is well prepared to accept that the large variety of potential tools fit to do the job opens tool choice up for political struggles. But he insists that the conception of rationality matters at least “up to a point” (p. 135).
- 6 Majone calls this “procedural rationality”.
- 7 The notion of “laying blame at someone’s door” features prominently already in the seminal study on “Risk and Culture” (Douglas & Wildavsky, 1983, p. 38).
- 8 While James Scott’s study describes striking instances of control failure, the historic episode on the meter can be interpreted as a case of successfully increases ‘legibility’.
- 9 I thank one anonymous reviewer for highlighting this point

## References

- Atkinson, Giles (2015). Cost-benefit analysis: a tool that is both useful and influential? In Andrew J. Jordan & Turnpenny, John R. (Eds.), *The tools of policy formulation: actors, capacities, venues and effects* (pp. 76-99). Cheltenham Spa: Edward Elgar.
- Bähr, Holger (2010). *The Politics of Means and Ends: Policy Instruments in the European Union*. London: Routledge.
- Baldwin, G. R. & Veljanovski, C. G., (1984). Regulation by cost-benefit analysis. *Public Administration*, 62 (1), 51-69.
- Balla, Steven J. (1998). Administrative procedures and political control of the bureaucracy. *The American Political Science Review*, 92 (3), 663-673.
- Béland, Daniel (2009). Ideas, institutions, and policy change. *Journal of European Public Policy*, 16 (5), 701-718.
- Béland, Daniel & Howlett, Michael (2016). How solutions chase problems: instrument constituencies in the policy process. *Governance*, 29 (3), 393-409.
- Bemelmans-Videc, Marie-Louise, Rist, Ray C. & Vedung, Evert Oskar (Eds.) (2011). *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation*. New Brunswick: Transaction Publishers.
- Benz, Arthur (2009). *Politik in Mehrebenensystemen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Benz, Arthur, Corcaci, Andreas & Doser, Jan Wolfgang (2016). Unravelling multilevel administration. Patterns and dynamics of administrative co-ordination in European governance. *Journal of European Public Policy*, 23 (7), 999-1018.
- Bjørnholt, Bente & Larsen, Flemming (2014). The politics of performance measurement: ‘Evaluation use as mediator for politics. *Evaluation*, 20 (4), 400-411.
- Black, Julia (2005). The emergence of risk-based regulation and the new public management in the UK. *Public Law*, 2005(Autumn), 512-549.
- Boswell, Christina (2008). The political functions of expert knowledge: knowledge and legitimization in European Union immigration policy. *Journal of European Public Policy*, 15 (4), 471-488.
- Capano, Giliberto & Lippi, Andrea (2017). How policy instruments are chosen: patterns of decision makers’ choices. *Policy Sciences*, 50 (2), 269-293.
- Carmel, Emma (Ed.) (2019). *Governance analysis: a new approach to politics, policy and practice*. Cheltenham Spa: Edward Elgar Publishing.

- Demeritt, David, Rothstein, Henry, Beaussier, Anne-Laure & Howard, Michael (2015). Mobilizing risk: explaining policy transfer in food and occupational safety regulation in the UK. *Environment and Planning A*, 47, 373-391.
- Desrosières, Alain (2002). *The politics of large numbers – A history of statistical reasoning*. Harvard University Press.
- Dodds, Anneliese (2006). The Core Executive's Approach to Regulation: From 'Better Regulation' to 'Risk-Tolerant Deregulation'. *Social Policy & Administration*, 40 (5), 526-542.
- Döhler, Marian (2002). Institutional Choice and Bureaucratic Autonomy in Germany. *West European Politics*, 25 (1), 101-124.
- Döhler, Marian (2015). Das Politikfeld als analytische Kategorie. *dms – der moderne staat*, 8 (1), 51-69.
- Douglas, Mary & Wildavsky, Aaron (1983). *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. University of California Press.
- Dunlop, Claire A. & Radaelli, Claudio M. (Eds.) (2016). *Handbook of Regulatory Impact Assessment*. Cheltenham Spa: Edward Elgar.
- Dunsire, Andrew (1978). *Control in a bureaucracy*. New York: St Martin's Press.
- Ewald, François (1986). *L'état providence*. Bernard Grasset.
- Hall, Peter A. (1993). Policy paradigms, social learning, and the state: the case of economic policy-making in Britain. *Comparative Politics*, 25 (3), 275-296.
- Hall, Peter A. & Taylor, Rosemary C. R. (1996). Political science and the three new institutionalisms. *Political Studies*, XLIV, 936-957.
- Hampton, Philip (2005). *Reducing administrative burdens: effective inspection and enforcement*. London: HM Treasury.
- Hood, Christopher (1983). *The Tools of Government*. London: Macmillan.
- Hood, Christopher, Rothstein, Henry & Baldwin, Robert (2001). *The Government of Risk: Understanding Risk Regulation Regimes*. Oxford University Press.
- Howlett, Michael (2011). *Designing Public Policies: Principles and Instruments*, 1st edition. Abingdon; New York: Routledge.
- Howlett, Michael & Tosun, Jale (2018). Policy styles: A new approach. In Michael Howlett & Jale Tosun (Eds.). *Policy Styles and Policymaking: Exploring the Linkages* (pp. 3-21). London; New York: Routledge.
- Hutter, Bridget M., (2005). The attractions of risk-based regulation: accounting for the emergence of risk ideas in regulation. *CARR Discussion Paper*, 33.
- Jankauskas, Vytautas & Eckhard, Steffen (2019). International bureaucracies as strategic actors: how the better regulation reform strengthens the European Commission. *Politische Vierteljahresschrift*, 60 (4), 681-699.
- Kassim, Hussein & Le Galès, Patrick (2010). Exploring governance in a multi-Level polity: a policy instruments approach. *West European Politics*, 33 (1), 1-21.
- Lascoumes, Pierre & Le Galès, Patrick (2007). Introduction: understanding public policy through its instruments – from the nature of instruments to the sociology of public policy instrumentation. *Governance*, 20 (1), 1-21.
- Lasswell, Harold D. (1971). *A pre-view of policy sciences*. New York: American Elsevier.
- Linder, Stephen H. & Peters, B. Guy (1989). Instruments of government: perceptions and contexts. *Journal of Public Policy*, 9 (1), 35-58.
- Loughlin, John (2017). Federal and local government institutions. In Daniele Caramani (Ed.). *Comparative politics* (pp. 188-204). Oxford University Press.
- Lowi, Theodore J. (1972). Four systems of policy, politics and choice. *Public Administration Review*, 32 (4), 298-310.
- Majone, Giandomenico (1994). The rise of the regulatory state in Europe. *West European Politics*, 17 (3), 77-101.
- Majone, Giandomenico (2010). Strategic issues in risk regulation and risk management. In OECD. *Reviews of Regulatory Reform* (pp. 93-128), Paris.

- March, James G. & Olsen, Johan P. (1989). *Rediscovering institutions: the organizational basis of politics*. New York: Free Press.
- Moran, Michael (2003). *The British Regulatory State: High Modernism and Hyper-innovation*. Oxford University Press.
- O'Malley, Pat (2004). *Risk, Uncertainty and Government*, London: Glasshouse Press.
- Organisation for Economic Co-Operation and Development (OECD) (2010). *Risk and Regulatory Policy: Improving the Governance of Risk*. Paris: OECD.
- Paul, Regine (2016). Entgegen aller Wahrscheinlichkeiten? Erklärungsansätze für die Diffusion risikobasierter Regulierung im föderalen System Deutschlands. *dms – der moderne staat*, 9 (2), 325-352.
- Paul, Regine (2017). Harmonization by risk analysis? Frontex and the risk-based governance of European border control. *Journal of European Integration*, 39 (6), 689-706.
- Paul, Regine (2021/fc). *Analyse and rule: The politics of analytical innovation in public administrations*. London: Routledge.
- Paul, Regine, Boudier, Frédéric & Wesseling, Mara (2016). Risk-based governance against national obstacles? Comparative dynamics of Europeanization in Dutch, French, and German flooding policies. *Journal of Risk Research*, 19 (8), 1043-1062.
- Peters, B. Guy (2002). The politics of tool choice. In Lester Salomon (Ed.). *Designing Government - From Instruments to Governance* (pp. 552-564). Oxford University Press.
- Pollitt, Christopher & Bouckaert, G. (2011). *Public Management Reform: A Comparative Analysis*. Oxford University Press.
- Porter, Theodore M. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.
- Posner, Eric A. (2001). Controlling Agencies with Cost-Benefit Analysis: A Positive Political Theory Perspective. *University of Chicago Law Review*, 68, 1137-1200.
- Radaelli, Claudio M. (2005). Diffusion without convergence: how political context shapes the adoption of regulatory impact assessment. *Journal of European Public Policy*, 12 (5), 924-943.
- Radaelli, Claudio M. (2010). Rationality, power, management and symbols: four images of regulatory impact assessment. *Scandinavian Political Studies*, 33 (2), 164-188.
- Radaelli, Claudio M. & Meuwese, Anne C. M. (2010). Hard questions, hard solutions: proceduralisation through impact assessment in the EU. *West European Politics*, 33 (1), 136-153.
- Raimondo, Estelle (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26-41.
- Richardson, Jeremy, Gustafsson, Gunnell & Jordan, Grant (1982). The concept of policy style. In Jeremy Richardson (Ed.). *Policy Styles in Western Europe* (pp. 1-16). London: Allen and Unwin.
- Rothstein, Henry et al. (2019). Varieties of risk regulation in Europe: Coordination, complementarity & occupational safety in capitalist welfare states. *Socio-Economic Review*, 17 (4), 993-1020.
- Rothstein, Henry, Borraz, Olivier & Huber, Michael (2013). Risk and the limits of governance: Exploring varied patterns of risk-based governance across Europe. *Regulation & Governance*, 7 (2), 215-235.
- Rothstein, Henry, Huber, Michael & Gaskell, George (2006). A theory of risk colonization: The spiraling regulatory logics of societal and institutional risk. *Economy and Society*, 35 (1), 91-112.
- Sabatier, Paul A. & Jenkins-Smith, Hank C. (1993). *Policy change and learning: an advocacy coalition approach*. Westview Press.
- Scott, James (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
- Shrader-Frechette, K.P. (1990). *Risk and Rationality. Philosophical Foundations for Populist Reforms*. University of California Press.
- Simons, Arno & Voß, Jan-Peter (2017). The concept of instrument constituencies: accounting for dynamics and practices of knowing governance. *Policy and Society*, 37 (1), 14-35.
- Stone, Deborah (2012). *Policy paradox: the art of political decision making*. New York: W.W. Norton & Co.
- Strassheim, Holger (2017). Bringing the political back in: reconstructing the debate over evidence-based policy. A response to Newman. *Critical Policy Studies*, 11 (2), 235-245.

- Strassheim, Holger & Beck, Silke (Hrsg). (2019). *Handbook of Behavioural Change and Public Policy*. Cheltenham Spa: Edward Elgar.
- Strassheim, Holger & Schwab, Oliver (2020). Politikevaluation und Evaluationspolitik. *dms – der moderne staat*, 13 (1), 3-23.
- Sunstein, Cass R. (2000). Cost-benefit default principles. *Michigan Law Review*, 99, 1651-1723.
- Sunstein, Cass R. (2002). *Risk and Reason: Safety, Law, and the Environment*. Cambridge University Press.
- Tholoniati, Luc (2010). The career of the Open Method of Coordination: lessons from a ‘soft’ EU instrument. *West European Politics*, 33 (1), 93-117.
- Tombs, Steve & Whyte, David (2013). Transcending the deregulation debate? Regulation, risk, and the enforcement of health and safety law in the UK. *Regulation & Governance*, 7 (1), 61-79.
- Turnpenny, John R., Jordan, Andrew, Benson, David & Rayner, Tim (2015). The tools of policy formulation: an introduction. In Andrew J. Jordan & John R. Turnpenny (Eds.). *The tools of policy formulation: actors, capacities, venues and effects, New horizons in public policy*, (pp. 3-29). Cheltenham Spa: Edward Elgar.
- Viscusi, W. Kip (1983). *Risk by Choice: Regulating Health and Safety in the Workplace*. Harvard University Press.
- Voß, Jan-Peter & Simons, Arno (2014). Instrument constituencies and the supply side of policy innovation: the social life of emissions trading. *Environmental Politics*, 23 (5), 735-754.
- Wagenaar, Henk (2011). *Meaning in action: interpretation and dialogue in policy analysis*, New York: M.E. Sharpe.
- Weiss, Carol (1973). The politics of impact measurement. *Policy Studies Journal*, 1 (3), 179-183.
- Weiss, Carol H. (1993). Where politics and evaluation research meet. *Evaluation Practice*, 14 (1), 93-106.
- Weiss, Carol Hirschon, Murphy-Graham, Erin & Birkeland, Sarah (2005). An alternate route to policy influence: how evaluations affect D.A.R.E. *American Journal of Evaluation*, 26 (1), 12-30.
- Yanow, Dvora & Schwartz-Shea, Peregrine (Eds.) (2006). *Interpretation and method. Empirical research methods and the interpretive turn*. New York: M.E. Sharpe.

### *Contact details of the author:*

Dr. Regine Paul, Universität Kassel, Fachbereich für Gesellschaftswissenschaften, Vertretungsprof. Vergleichende Politikwissenschaft, Nora-Platiel-Straße 1, 34127 Kassel, E-Mail: regine.paul@uni-kassel.de

Jens Weiß, Dino Schubert

## Doppelte Reform mit geteiltem Erfolg: Zur Wirkung der DOPPIK-Reform auf Haushaltsführung und Verwaltungssteuerung in den Kommunen

### Zusammenfassung

Die DOPPIK-Reform war für die kommunalen Verwaltungen in Deutschland eines der zentralen Modernisierungsprojekte der letzten 20 Jahre. Dieser Beitrag untersucht den aktuellen Stand der Implementierung des doppischen Rechnungswesens und managerialistischer Steuerungsinstrumente. Auf Basis einer Sichtung der internationalen Literatur zur Implementierung managerialistischer Formen der Steuerung wird vertiefend untersucht, in welchem Umfang Ansätze des Performance Managements und des Strategischen Managements in Kommunen implementiert wurden und welche Faktoren hierfür relevant sind. Empirische Grundlage ist eine Befragung der Kommunen in Niedersachsen, Nordrhein-Westfalen und Sachsen-Anhalt. Die Ergebnisse zeigen, dass die Umstellung auf eine doppische Rechnungsführung weitgehend vollzogen ist. Eine managerialistische Form der Steuerung wurde dagegen nur in sehr wenigen der untersuchten Kommunen eingeführt. Gleichzeitig ist festzustellen, dass die derzeit verfügbaren theoretischen Modelle zur Erklärung der Implementierung managerialistischer Instrumente in deutschen Kommunen nicht ausreichen.

**Schlagerworte:** Verwaltungssteuerung, Performance Management, Strategisches Management, DOPPIK

### Abstract

*Double the reform, half the success? The implementation of accrual accounting and managerial instruments through the DOPPIK reforms in German local governments*

The implementation of a double-entry full accrual accounting system combined with ideas of output- or impact-oriented management—known as DOPPIK—has been the most important reform project for German local governments within the last 20 years. Based on a review of international literature on the use of managerial instruments in the public sector, this article analyses the present state of the reform focusing on the implementation of Performance Management and Strategic Management. The dissemination of managerial instruments and the main factors influencing their adoption are examined using data from a survey of local governments in the federal states of Lower Saxony, North Rhine-Westphalia and Saxony-Anhalt. Whereas many local governments implemented accrual accounting, only very few have implemented managerial ideas and instruments. Moreover, public management concepts and models cannot explain the variations in the implementation of managerial practices.

**Keywords:** Performance Management, Strategic Management, Local Government Management, DOPPIK

## 1 Einleitung

Die Einführung der als DOPPIK bekannt gewordenen Variante einer produkt- und ressourcenorientierten, doppelten Buchführung war für die kommunalen Verwaltungen in Deutschland eines der zentralen Reformprojekte der letzten 20 Jahre. Der der DOPPIK-Reform zu Grunde liegende Beschluss der *Ständigen Konferenz der Innenminister und -senatoren der Länder* (IMK) (2003) verknüpfte zwei bis dahin weitgehend unabhängig voneinander geführte Reform-Vorhaben: Einerseits Ideen zur Erneuerung des kommunalen Rechnungswesens (Lüder, 1996) und andererseits das Konzept zur outputorientierten Steuerung über einen Produkthaushalt, das – basierend auf Ideen des *New Public Management* – von der Kommunalen Gemeinschaftsstelle für Verwaltungsmodernisierung (KGSt) (1993) mit dem *Neuen Steuerungsmodell* propagiert wurde. Entsprechend finden sich im IMK-Beschluss zwei Ziele der Reform: Erstens sollte mit der Reform „das kommunale Haushalts- und Rechnungswesen von der bislang zahlungsorientierten Darstellungsform auf eine ressourcenorientierte Darstellung umgestellt“ werden und zweitens sollte zukünftig „die Steuerung der Kommunalverwaltungen statt durch die herkömmliche Bereitstellung von Ausgabeermächtigungen (Inputsteuerung) durch die Vorgabe von Zielen für die kommunalen Dienstleistungen (Outputsteuerung)“ erfolgen (Ständige Konferenz der Innenminister und -senatoren der Länder, 2003, S. 2).

16 Jahre nach diesem Beschluss ist schon formal ein äußerst heterogener Entwicklungsstand zu konstatieren (Mühlenkamp & Glöckner, 2009; Böhme & Heindl, 2014; Böhme, 2017; Adam & Mahnke, 2018): In Bayern, Schleswig-Holstein und Thüringen können die Kommunen zwischen einer kameralistischen und einer doppelischen Haushaltsführung wählen; diese Regelung gilt derzeit unbefristet. In Baden-Württemberg, einst Pionier der doppelischen Haushaltsführung (Lüder, 1996; 2001), wurde die Frist zur Umstellung auf einen doppelischen Haushalt auf den 01.01.2020 verschoben. Für die Kommunen der anderen Bundesländer galten zur Umstellung auf eine doppelische Haushaltsführung Fristen zwischen dem 01.01.2009 (Nordrhein-Westfalen, Rheinland-Pfalz und Saarland) und dem 01.01.2015 (Hessen). Die rechtlichen Grundlagen zur Umsetzung der doppelischen Haushaltsführung weisen darüber hinaus nicht nur zu den Details des Rechnungswesens, sondern auch hinsichtlich der Regelungen zur Implementierung einer outputorientierten Steuerung über Ziele und Kennzahlen eine erhebliche Heterogenität auf.

Der vorliegende Beitrag untersucht den Stand der DOPPIK-Reform insbesondere mit Blick auf Aspekte der Verwaltungssteuerung und leistet einen Beitrag zu einer Bewertung der Reformfolge. Die zentralen Fragestellungen des Beitrags sind erstens, in welchem Umfang managerialistische Instrumente, in diesem Fall die Steuerung über strategische Ziele und Kennzahlen – Performance Management und Strategisches Management –, bisher implementiert wurden, und zweitens, welche Faktoren einen wesentlichen Einfluss auf deren Einführung hatten. Empirische Grundlage ist eine Befragung der 1.003 Kommunen in Niedersachsen, Nordrhein-Westfalen und Sachsen-Anhalt, die im Sommer 2019 durchgeführt wurde. Die Ergebnisse zeigen, dass die Umstellung auf eine doppelische Rechnungsführung weitgehend vollzogen ist. Eine managerialistische Form der Steuerung wurde dagegen nur in wenigen der untersuchten Kommunen eingeführt. Dort, wo Leistungsinformationen, Kennzahlen und Ziele verfügbar sind, werden diese kaum für operative oder politische Entscheidungen genutzt. Gezeigt

wird darüber hinaus, dass die verfügbaren theoretischen Modelle zur Erklärung der Implementierung managerialistischer Instrumente nicht ausreichen, um die Reformdynamik in deutschen Kommunen zu erklären.

Im folgenden zweiten Kapitel werden zunächst die bisher vorliegenden Ergebnisse zur Umsetzung der Reform diskutiert. Anschließend werden theoretische Grundlagen zur Erklärung der Implementierung managerialistischer Instrumente diskutiert und das für die empirische Untersuchung genutzte Analysemodell entwickelt. Im vierten Kapitel wird die Methodik der durchgeführten Befragung erläutert. Anschließend werden die Ergebnisse der empirischen Erhebung dargestellt. Der Beitrag endet mit einer Diskussion der empirischen Ergebnisse und einer kurzen Zusammenfassung.

## 2 Stand der Diskussion zur Nutzung managerialistischer Instrumente in deutschen Kommunen

Bereits Ende der 2000er Jahre wurde deutlich, dass in den betroffenen Kommunen die Umstellung auf eine doppischen Haushaltsführung überwiegend wie geplant realisiert werden konnte, dass aber andererseits kaum neue, ziel- und kennzahlenbasierte Steuerungsstrukturen und -prozesse implementiert wurden (Innenministerium des Landes Nordrhein-Westfalen & steria mummert consulting, 2008; KGSt, 2008; 2010; Rechnungshof Rheinland-Pfalz, 2011). Eine erste Befragung in kreisfreien Städten und größeren Gemeinden zeigte eine tendenziell positive Einstellung der Kämmerereien zur DOPPIK, während aus Sicht der politischen Vertretungen der Verlust an Übersichtlichkeit des Haushalts bemängelt wurde (Burth & Hilgers, 2012). Auch Andreas Schmid (2012; 2014) kommt zu dem Ergebnis, dass die Reformen zwar aus kaufmännischer Sicht richtig und in Anbetracht der Verschuldung des öffentlichen Sektors notwendig seien, die Einführung von Managementinstrumenten aber nicht zu Verhaltensänderungen von Politikerinnen und Politikern geführt hätten.

Die haushaltsrechtlichen Regelungen von Brandenburg, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen, und Sachsen-Anhalt enthalten Vorgaben zur Evaluation der DOPPIK-Einführung. Die seit 2012 hierzu durchgeführten Verfahren beruhen überwiegend auf wenig strukturierten Formen des Erfahrungsaustausches der für die gesetzlichen Grundlagen verantwortlichen Ministerien mit einzelnen Kommunen, den kommunalen Spitzenverbänden und Gemeindeprüfungsanstalten, entsprechenden Anhörungen oder dem Einholen von Stellungnahmen. Schwerpunkte sind die praktische Anwendbarkeit der gesetzlichen Regelungen für die Buchführung und mögliche Verbesserungen in diesem Bereich (Landtag Brandenburg, 2012; Landtag von Sachsen-Anhalt, 2014; Landtag Rheinland-Pfalz, 2015; Sächsisches Staatsministerium des Innern, 2016; Ministerium für Heimat, Kommunales, Bau und Gleichstellung des Landes Nordrhein-Westfalen, 2017). Insgesamt wird in den vorliegenden Untersuchungen deutlich, dass die Umstellung des Rechnungswesens in den meisten Kommunen zu den festgelegten Stichtagen realisiert werden konnte, während Aspekte der „neuen Steuerung“ bestenfalls partiell umgesetzt wurden. Für das Land Sachsen-Anhalt liegt eine Evaluation auf Basis einer Vollbefragung der dortigen Kommunen ergänzt durch Interviews und die Analyse von Haushaltsplänen vor (Weiß, 2014). Die Ergebnisse zeigen, dass ein größerer Teil der Kommunen ihre Haushaltsführung nicht zum Stichtag umstellen konnte. Die Akzeptanz der DOPPIK sei weder in den Verwaltungen noch in den

Gemeinderäten besonders hoch und Ansätze für eine verbesserte Steuerung kaum zu finden (ebd., S. 41 f.).

In den Jahren 2011 und 2012 war der Nutzen der DOPPIK in den Kommunen Gegenstand einer breiteren Debatte in der Zeitschrift *Verwaltung & Management*. Kritikerinnen und Kritiker argumentierten, die Implementierung des neuen Steuerungsmodells sei im Wesentlichen gescheitert und die doppelte Haushaltsführung bringe keine Impulse für eine „neue Steuerung“ in den Kommunen, sondern werde eher als intransparent empfunden, und letztlich sei das Verhältnis von Aufwand zu Nutzen schlechter als in der kameralistischen Haushaltsführung (Bogumil, Ebinger & Holtkamp, 2011; 2012, Bogumil & Holtkamp, 2012). Die Befürworterinnen und Befürworter der Reform forderten eine Differenzierung zwischen dem Rechnungswesen und Aspekten der Verwaltungssteuerung und unterstrichen die betriebswirtschaftliche Rationalität der doppelten Haushaltsführung (Reichard, 2011; 2012; Burth & Hilgers, 2012; Fischer, 2012).

Zur Verfügbarkeit und Nutzung von Leistungsinformationen liegen einige detailliertere Ergebnisse vor: In einer frühen Untersuchung in der Berliner Sozialhilfeverwaltung stellt Tobias Polzer (2010) teilweise erhebliche Widerstände der Führungskräfte gegen die Nutzung von Leistungsinformationen fest, insbesondere aufgrund ihrer politischen Relevanz und unterschiedlicher Informationsinteressen in Politik und Verwaltung. Für die kreisfreien Städte in Deutschland kommen Alexander Kroll und Isabella Proeller (2012) zu dem Ergebnis, dass eine Steuerung über Ziele und Kennzahlen auf der operativen Ebene nur in Ansätzen entwickelt sei und an der mangelnden Qualität der Messsysteme scheitern könnte. Gleichzeitig sei ein geringes Interesse der Politik an Performance-Informationen (PI) zu konstatieren, eine Einschätzung, die auch von Jörg Bogumil, Falk Ebinger und Lars Holtkamp (2011) geteilt wird. Eine empirische Untersuchung von kommunalen Haushaltsplänen zeigt, dass in vielen Verwaltungen keine oder kaum Informationen über die Leistungen der Verwaltung zur Verfügung stehen (Weiß, 2013b). In Verwaltungen, in denen mit PI gearbeitet wird, ist deren Verfügbarkeit und Qualität in der Regel heterogen, d. h. stark von bereichsspezifischen Aspekten geprägt (Kroll, 2012). Christoph Reichard (2014) kommt zu dem Ergebnis, dass nur ca. 25% der Kommunen PI in ihren Haushaltsplänen ausweisen, wobei sich diese Informationen kaum auf Outputs oder Wirkungen bezögen. Insgesamt seien die bereitgestellten PI sowohl qualitativ als auch quantitativ für eine Steuerung der Budgets nicht ausreichend. Selbst Kennzahlensysteme mit einem hohen Detaillierungsgrad weisen systematische Schwächen auf, weil sie zum Beispiel überwiegend reine Finanzkennzahlen oder keine Werte aus der Vergangenheit ausweisen (Bleyen, Klimovsky, Bouckaert & Reichard, 2016). Andreas Burth und Dennis Hilgers (2012; 2014) stellen fest, dass das doppelte Rechnungswesen von Führungskräften kommunaler Finanzbereiche ebenso wie von Haushaltspolitikern und -politikern zwar insgesamt positiv eingeschätzt werde, konstatieren aber ebenfalls, dass die für Steuerung notwendigen Kennzahlen bislang überwiegend kaum in Haushaltsplänen zu finden seien. Ergebnisse einer Studie zum Strategischen Management in Kommunen, die auf der Auswertung von ca. 700 kommunalen Haushaltsplänen beruht, zeigen, dass maximal ca. 15%-20% der deutschen Kommunen über eine Form des Strategischen Managements verfügen könnten (Weiss, 2017).

Insgesamt deuten alle vorliegenden Ergebnisse darauf hin, dass in den Bundesländern, in denen eine entsprechende rechtliche Verpflichtung besteht, die formale Umstellung des Haushalts- und Rechnungswesens auf eine doppelte Systematik flächen-



deckend erfolgt ist. Nicht abschließend geklärt ist, welche Wirkungen die DOPPIK-Reform auf Management-Strukturen und -Prozesse in Verwaltungen sowie auf die Verwaltungssteuerung, verstanden als Interaktion zwischen Verwaltung und politischen Gremien, hatte. Hier gibt es zwar einzelne Befunde, bislang aber keine umfassende Einschätzung zum Reformergebnis.

### 3 Theoretische Grundlagen zur Erklärung der Implementierung managerialistischer Instrumente in öffentlichen Verwaltungen

#### 3.1 Performance Management und Strategisches Management – Stand der Forschung

Die Nutzung managerialistischer Instrumente wird verwaltungswissenschaftlich insbesondere mit Bezug auf die Konzepte *Performance Management* und *Strategisches Management* diskutiert. Unter Performance Management wird dabei die Nutzung von PI für Entscheidungen über den Einsatz von Ressourcen innerhalb der Verwaltung und durch politische Gremien verstanden (Moynihan, 2008; Proeller & Siegel, 2009; van Dooren, Bouckaert & Halligan, 2015). Strategisches Management wird üblicherweise als eine Form der Führung einer Organisation definiert, die sich an analytisch entwickelten, mittel- bis langfristigen Zielstellungen orientiert und zur Realisierung dieser Ziele systematisch Maßnahmen entwickelt, umsetzt sowie evaluiert. Breitere Definitionen betonen dabei die Relevanz einer zielorientierten Organisationskultur, welche die Kohärenz des Verwaltungshandelns verbessert und organisationale Lernprozesse fördert (Bryson, 2015; Ferlie & Ongaro, 2015; Joyce, 2015; Siegel, 2019).

Die Messung von Leistungen und die Bereitstellung von PI wird als eine mittlerweile auch in öffentlichen Organisationen weit verbreitete Praxis gesehen (van Dooren, Bouckaert & Halligan, 2015). Allerdings belegen empirische Ergebnisse, dass sowohl aus Sicht der Verwaltungen wie auch der gewählten Vertretungen Anreize bestehen, PI nicht zu erheben oder nicht zu veröffentlichen, um Diskussionen über schlechte Leistungen und entsprechende Schuldzuweisungen zu vermeiden (*blame avoidance*; Hood, 2007; Charbonneau & Bellavance, 2012; Hood, 2011).

Stehen PI zur Verfügung, so generieren diese unter Umständen Anreize für Akteure, ihr Handeln in nicht gewünschter Weise nur auf gemessene Outputs auszurichten oder Leistungsdaten gezielt zu manipulieren (*gaming* und *creaming*; Courty & Marschke, 2004; Bevan & Hood, 2006). Zur Verfügung stehende PI werden außerdem bei politischen Entscheidungen ignoriert, rein taktisch genutzt oder (bewusst) falsch interpretiert (Ammons & Rivenbark, 2008; Askim, 2009; Dimitrijevska-Markoski & French, 2019; McDavid & Huse, 2012; Moynihan & Pandey, 2010; Nielsen & Baekgaard, 2015; Rivenbark, Fasiello & Adamo, 2018; Saliterer & Korac, 2013). In einer Meta-Analyse von 25 Zeitschriftenbeiträgen identifiziert Alexander Kroll (2015) sieben Faktoren, die vergleichsweise stark positiv auf die Nutzung von PI für Management-Entscheidungen zu wirken scheinen: Reifegrad des Messsystems, Einbezug von Stakeholdern, Unterstützung des Systems durch die Führungskräfte, interne Ressourcen zur Unterstützung des Performance Managements, eine innovative Organisationskultur und

die Existenz klarer Ziele. Daneben existiert eine Vielzahl weiterer Faktoren, für die schwächere Wirkungen auf die Nutzung von PI nachgewiesen wurden.

Insgesamt belegen die vorliegenden Ergebnisse erhebliche Schwierigkeiten bei der Implementierung von Performance-Management-Systemen und eine tendenziell geringe Berücksichtigung von Performance-Informationen insbesondere bei politischen Entscheidungen, und dies auch in Ländern mit einer ausgeprägten *New Public Management* (NPM)-Tradition und rechtlichen Regelungen zur verbindlichen Implementierung von Performance Management. Theoretisch können wesentliche Teile dieser Ergebnisse mit Bezug auf die *blame-avoidance*-Theorie sowie einer Priorisierung politischer Ziele gegenüber Effizienz- bzw. Effektivitätszielen erklärt werden.

Die Forschung zum *Strategischen Management* auf kommunaler Ebene hatte überwiegend etwas andere Schwerpunkte. Theodore H. Poister und Gregory D. Streib (1994; 2005) berichten, dass ca. 40%-45% der US-amerikanischen Kommunen eine Form strategischer Planung implementiert haben, wobei sich dieser Wert von 1994 auf 2005 kaum verändert zu haben scheint. In der Untersuchung von Benedict S. Jimenez (2014) geben ca. 60% der US-amerikanischen Städte, die sich an der Umfrage beteiligten, an, einen „strategischen Plan“ zu haben. In einer 2005 in Florida durchgeführten Befragung konstatierten Myungjung Kwon, Frances Berry und Hee Soun Jang (2014), dass ungefähr 50% der Städte, die geantwortet hatten, keine Form der strategischen Planung implementiert hatten. Alvin E. Holliman und Marcia Bouchard (2015) berichten auf Basis einer anderen Umfrage, dass ca. 17% der befragten US-amerikanischen Kommunen angeben, über eine Form von *Management by Objectives* zu verfügen. In der Befragung von Åge Johnsen (2017) gaben ca. 40% der teilnehmenden norwegischen Kommunen an, nicht über strategische Planungsdokumente zu verfügen. Es kann also davon ausgegangen werden, dass auch in Ländern mit stärker ausgeprägten NPM-Reformen Kommunen nur teilweise ein Strategisches Management implementiert haben.

Die Bedeutung der Größe von Kommunen für die Umsetzung managerialistischer Instrumente wurde bereits früh von William C. Rivenbark und Janet M. Kelly (2003) herausgearbeitet. Auch Kwon, Berry und Jang (2014) kommen zu dem Ergebnis, dass größere Kommunen deutlich eher Formen eines Strategischen Managements implementieren als kleinere. Bert George, Steven van de Walle und Gerhard Hammerschmid (2019) kommen zu einem ähnlichen Ergebnis, unterstreichen aber gleichzeitig die Bedeutung von Homogenität bzw. Heterogenität der Aufgaben. Für die Einführung der DOPPIK in Kommunen in Bayern und Thüringen hat darüber hinaus Doris Böhme (2017) einen positiven Effekt der Größe nachgewiesen.

Untersuchungen zur Wirkung des Strategischen Managements auf die *Performance* von Kommunen finden teilweise keine, teilweise schwach positive Effekte (Andrews, Boyne, Law & Walker, 2012; Bryson, Crosby & Bryson, 2009; Jimenez, 2014; Pasha, Poister & Edwards, 2015; Walker & Andrews, 2015; George, Walker & Monster, 2019). Die Untersuchung von Johnsen (2017) zeigt, dass auch dann, wenn keine Wirkungen auf objektivierbare Messgrößen zu erkennen sind, die Praxis des Strategischen Managements durch die Führungskräfte einer Organisation typischerweise positiv bewertet wird. Strategisches Management scheint eine Form der Führung von Organisationen zu sein, die für bestimmte Praktikerinnen und Praktiker einen Wert an sich hat, weil sie zu deren Verständnis von Führung und normativen Vorstellungen von Verwaltungsmanagement passt. Gleichzeitig, so Bert George, Richard Walker und Joost Monster (2019, p. 285) sei festzustellen, „dass die Realitäten in öffentlichen Verwal-

tungen und insgesamt im öffentlichen Sektor die Entwicklung stärker strategisch ausgerichteter Formen von Führung verhindern“.

Weitere wichtige Faktoren hierfür scheinen die große Anzahl der in die Formulierung von Strategien einzubindenden Stakeholder sowie die Heterogenität der zu berücksichtigenden Interessen zu sein (Conteh, 2014). Ebenso wie beim Performance Management dürften auch beim Strategischen Management Aspekte von *blame avoidance* von Bedeutung sein, weil strategische Ziele immer das Risiko eines messbaren Scheiterns implizieren (Weiss, 2017).

Entscheidungen für oder gegen die Implementierung eines Strategischen Managements als Konsequenz von Aufwand-Nutzen-Kalkülen zu interpretieren (Ritz, 2005; Kwon, Berry & Jang 2014), führt zu der Frage, wie diese ex ante eingeschätzt werden bzw. eingeschätzt werden können. Vermutlich sind für solche Abschätzungen auch die Werte und Einstellung der Führungskräfte von Bedeutung (Saliterer & Korac, 2014).

### 3.2 Entwicklung des Analysemodells

Vor dem Hintergrund der rechtlichen Regelungen für die Kommunen in Deutschland ergibt sich das folgende theoretische Modell. Die Einführung einer doppelischen Buchhaltung ist in verschiedenen Bundesländern gesetzlich verbindlich geregelt. Die Regelungen zur Umsetzung von Aspekten des Performance Managements bzw. des Strategischen Managements sind deutlich zurückhaltender. Strategische Ziele werden in § 41, u der Gemeindeordnung (GO NRW) sowie in § 7 der Kommunalhaushaltsverordnung (KomHVO NRW) Nordrhein-Westfalens erwähnt. Das Niedersächsische Kommunalverfassungsgesetz (NKomVG) erwähnt in § 58 Abs. 1 Nr. 1 die Zuständigkeit der politischen Vertretung für Beschlüsse über grundlegende Entwicklungsziele der Kommune. Daraus folgt allerdings keine Verpflichtung, solche Ziele zu entwickeln. Vorgaben zu Zielen und Kennzahlen sind überwiegend lediglich Soll-Bestimmungen. Das heißt, dass die Implementierung managerialistischer Instrumente – jenseits der formalen Erfüllung schwacher Vorgaben zu einigen Instrumenten – in Kommunen freiwillig ist. Vor dem Hintergrund der oben angeführten Ergebnisse bisheriger Untersuchungen ist davon auszugehen, dass die Steuerung über (strategische) Ziele und Kennzahlen noch nicht umfassend verbreitet ist und die auf Isomorphismus resultierenden Anreize (Dimaggio & Powell, 1991), diese Instrumente einzuführen, gering sind. Akteure in Verwaltung oder den politisch gewählten Vertretungen, die eine Entscheidung für die Implementierung managerialistische Instrumente treffen, agieren eher als Pioniere und müssen die Erwartung haben, dass deren Nutzen zumindest mittelfristig größer ist, als der Aufwand zu deren Einführung. Entsprechende Nutzenerwartungen können auf Erwartungen an effizienz- und effektivitätssteigernden Wirkungen dieser Instrumente und auf der Möglichkeit beruhen, sich durch die Nutzung solcher Instrumente als „moderne Verwaltung“ zu profilieren.

Daher wurde bei der Entwicklung der Hypothesen davon ausgegangen, dass nur eine kleinere Zahl von Kommunen managerialistische Instrumente implementiert haben (Hypothese 1, H1) und dass Faktoren identifizierbar sind, welche die Wahrscheinlichkeit der Einführung solcher Instrumente positiv beeinflussen. Entsprechend der Ergebnisse von Rivenbark und Kelly (2003), Kwon, Berry und Jang (2014), Böhme (2017) sowie George, van de Walle und Hammerschmid (2019) ist zu erwarten, dass

die Größe der Kommunen, gemessen anhand der Bevölkerung, einen positiven Effekt auf die Wahrscheinlichkeit einer Implementierung hat (H2). Dies kann einerseits daran liegen, dass nur größere Kommunen überhaupt über Beschäftigte mit entsprechenden Kompetenzen verfügen, die dann auch spezialisiert für Managementaufgaben eingesetzt werden können. Andererseits ist es wahrscheinlich, dass aus Sicht kleinerer Kommunen die Möglichkeiten einer Durchsetzung einer eigenen strategischen Planung als geringer eingeschätzt werden, weil das Gestaltungspotenzial im Verhältnis zu den Entwicklungen in der Umwelt als sehr beschränkt wahrgenommen wird. Umgekehrt mag in kleineren Kommunen die Wahrnehmung dominieren, hier sei die Komplexität nicht so groß, als dass komplexe Steuerungsinstrumente notwendig seien.

Auf Basis der Ergebnisse von Kroll (2015) wird weiterhin ein positiver Einfluss der Länge des Zeitraums vermutet, über den sich Kommunen mit managerialistischen Instrumenten beschäftigen (H3). Dieser Aspekt wird über den Zeitpunkt des Beginns der DOPPIK-Einführung operationalisiert. Die Verfügbarkeit von Ressourcen (Kroll, 2015) sollte ebenfalls eine positive Wirkung haben (H4) und wird über die Haushaltssituation sowie Personal- bzw. Finanzprobleme im Einführungszeitraum operationalisiert. Für die Entwicklung von Hypothesen zur Wirkung von *blame avoidance* wurden Überlegungen von Jörg Bogumil, Lars Holtkamp, Martin Junkernheinrich und Uwe Wagschal (2014) sowie Jens Weiss (2017) einbezogen. Bogumil, Holtkamp, Junkernheinrich und Wagschal (2014) argumentieren, dass die Politisierung in konkurrenzdemokratischen Ratssystemen stärker ist und zu einer höheren Verschuldung von Kommunen führt, weil Parteien größere Anreize haben, sich über Ausgaben für bestimmte Gruppen politisch zu profilieren und die Koordinationsfunktionen der Bürgermeisterinnen und Bürgermeister sowie Finanzverantwortlichen schwächer sind. Daraus ergibt sich die Hypothese, dass in Ländern mit eher konkordanzdemokratischen Kommunalverfassungen Kommunen eher managerialistische Instrumente implementieren (H5). Weiss (2017) vermutet, dass *blame-avoidance*-Probleme bei der Implementierung eines Strategischen Managements in Kommunen mit einem höheren Vertrauen zwischen Rat und Verwaltung eher überwunden werden können, weil hier die Gefahr kleiner ist, dass bei der Nichterreichung von Zielen eine gegenseitige Schuldzuweisung stattfindet. Daher ist zu erwarten, dass Kommunen mit einem größeren Vertrauen und einem – aus Sicht der Beteiligten – besseren Verhältnis zwischen Rat und Verwaltung eher managerialistische Instrumente implementiert haben (H6).

Schließlich ist auf Basis der Ergebnisse von Kroll (2015) zu erwarten, dass es eine positive Wirkung bzw. eine Wechselwirkung zwischen einem Strategischen Management mit klaren Zielen und der Implementierung von Performance Management gibt (H7). Dieser Aspekt wird mit Fragen zum Verhältnis von Rat und Verwaltung operationalisiert. Sollten sich die in der normativen Management-Literatur behaupteten und teilweise nachgewiesenen, positiven Effekte managerialistischer Instrumente auf Effizienz und Effektivität auch in deutschen Kommunen einstellen, dann ist schließlich zu erwarten, dass Kommunen, die managerialistische Instrumente eingeführt haben, insgesamt mit der DOPPIK-Einführung zufriedener sind (H8). *Tabelle 1* gibt einen Überblick über die Hypothesen.

Tabelle 1: Übersicht über die Forschungshypothesen

|    | Hypothese                                                                                                                                    | Operationalisierung                                                                                            | Bezüge                                                                                                    |
|----|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| H1 | Managerialistische Instrumente werden bislang nur von einer geringen Anzahl deutscher Kommunen genutzt.                                      | Anzahl der Kommunen, die angeben über ein Performance Management oder ein Strategisches Management zu verfügen | Reichard, 2014; Weiß, 2014; Weiss, 2017                                                                   |
| H2 | Größere Kommunen nutzen eher managerialistische Instrumente.                                                                                 | Bevölkerung                                                                                                    | Rivenbark & Kelly, 2003; Kwon, Berry & Jang, 2014; Böhme, 2017; George, van de Walle & Hammerschmid, 2019 |
| H3 | Kommunen, die früher mit der Einführung der DOPPIK begonnen haben, haben eher managerialistische Instrumente implementiert.                  | Zeitraum seit Beginn der DOPPIK-Einführung                                                                     | Kroll, 2015                                                                                               |
| H4 | Die Verfügbarkeit von finanziellen (H4a) und Personalressourcen (H4b) wirkt positiv auf die Implementierung managerialistischer Instrumente. | Allgemeine Haushaltssituation, Finanz- und Personalprobleme im Einführungszeitraum                             | Kroll, 2015                                                                                               |
| H5 | In Bundesländern mit konkordanzdemokratischen Kommunalverfassungen implementieren Kommunen eher managerialistische Instrumente.              | Konkordanzindex nach Bogumil & Holtkamp, 2013                                                                  | Bogumil & Holtkamp, 2013                                                                                  |
| H6 | Kommunen mit einem positiven und vertrauensvollen Verhältnis zwischen Rat und Verwaltung verfügen eher über managerialistische Instrumente.  | Wahrnehmung des Verhältnisses zwischen Rat und Verwaltung durch die Befragten                                  | Weiss, 2017                                                                                               |
| H7 | Es besteht ein positiver Zusammenhang zwischen Performance Management und Strategischem Management.                                          | Anzahl der Kommunen, die angeben über ein Performance Management oder ein Strategisches Management zu verfügen | Kroll, 2015                                                                                               |
| H8 | Kommunen, die managerialistische Instrumente implementiert haben, sind insgesamt zufriedener mit der DOPPIK-Einführung.                      | Zufriedenheit der Befragten mit der DOPPIK-Einführung                                                          | NPM-Literatur                                                                                             |

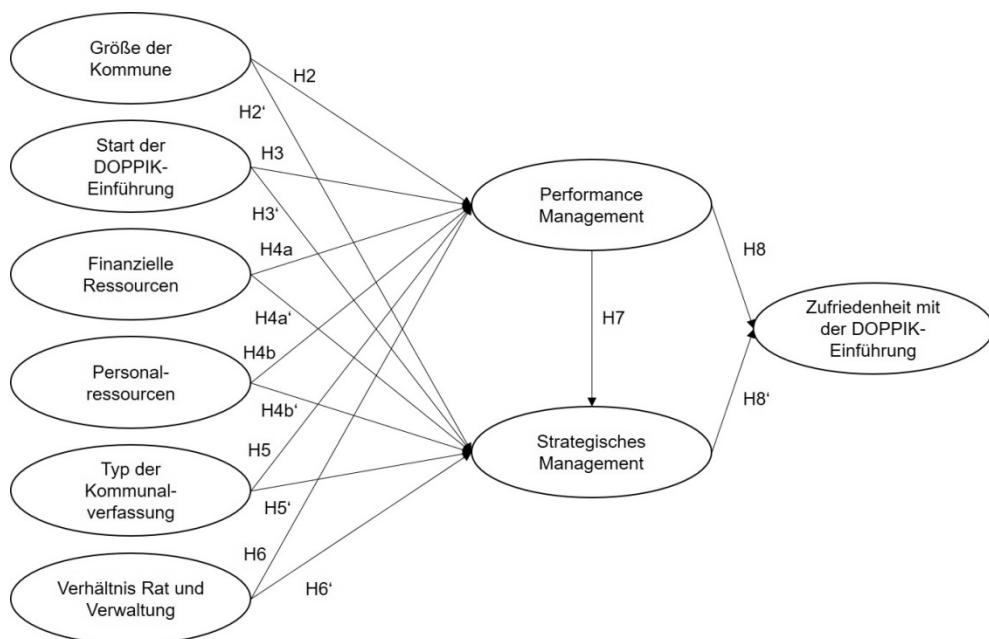
Quelle: Eigene Darstellung.

Zu verschiedenen dieser Hypothesen, lassen sich aus der vorliegenden Literatur Gegenhypothesen entwickeln. So könnte beispielsweise die Größe einer Kommune eine größere Heterogenität der Aufgaben mit sich bringen und negativ auf die Implementierung insbesondere eines Strategischen Managements wirken (George, van de Walle & Hammerschmid, 2019). Eine schwierige Haushaltssituation könnte auch die Anreize erhöhen, *Performance*-steigernde Instrumente zu implementieren, wie zum Beispiel im Fall der Stadt Lemgo dokumentiert (Tolkemitt, 2014). Außerdem variieren nicht nur die Kommunalverfassungen zwischen den Bundesländern, sondern auch die festgeleg-

ten Fristen zur Einführung der DOPPIK. Nordrhein-Westfalen hat beispielsweise eine konkurrenzdemokratische Kommunalverfassung (Bogumil & Holtkamp, 2013), hat aber am deutlichsten Elemente des Neuen Steuerungsmodells in rechtliche Regelungen übernommen und ist außerdem das Land mit dem frühesten Umstellungsstichtag (Adam & Mahnke, 2018).

Zur Entwicklung eines analytischen Modells haben wir von diesen Aspekten zunächst abgesehen, sie werden in der Diskussion der Ergebnisse wieder aufgegriffen. *Abbildung 1* zeigt das der Auswertung zu Grunde gelegte Modell auf Basis der Hypothesen aus *Tabelle 1*. Die Einzelhypothesen werden für beide Instrumente - Performance Management und Strategisches Management - getrennt überprüft.

*Abbildung 1:* Modell zur Erklärung der Implementierung von Performance Management und Strategischem Management



*Quelle:* Eigene Darstellung.

## 4 Methodisches Vorgehen zur Datenerhebung und Auswertung

Die Daten für diese Untersuchung wurden im Mai und Juni 2019 mit einer Online-Befragung über das Befragungsportal Questback ([www.unipark.de](http://www.unipark.de)) erhoben. Als Stichprobe wurden die Kommunen in Niedersachsen (Umstellung auf doppischen Haushaltsplan zum 01.01.2012), Nordrhein-Westfalen (Umstellung zum 01.01.2009) und Sachsen-Anhalt (Umstellung zum 01.01.2013 bzw. mit Ausnahmegenehmigung zum 01.01.2014) gewählt. Die Stichprobe umfasst damit Kommunen mit vergleichs-

weise frühen und späteren Einführungsstichtagen und unterschiedlichen Kommunalverfassungen. Angeschrieben wurden Führungskräfte aus den Finanzbereichen aller 1.003 Kommunen der drei Länder. Die Einladung zur Teilnahme an der Befragung erfolgte per E-Mail an die öffentlich zugänglichen Mailadressen der Finanzbereiche bzw. soweit möglich direkt an die Mailadressen der Führungskräfte. Teilweise mussten die von Kommunen zur Verfügung gestellten Online-Kontaktformulare genutzt werden.

438 Kommunen haben auf die Anfrage zur Beteiligung an der Umfrage reagiert. Davon haben 80 Befragte die Datenschutzerklärung abgelehnt und nicht an der Befragung teilgenommen. 274 Befragungen wurden abgeschlossen. Nach dem Aussortieren von Datensätzen mit nur wenigen beantworteten Fragen und offensichtlich falschen Angaben verblieben 266 verwertbare Datensätze, was einem Rücklauf von 26,5% entspricht. Im Vergleich zur Gesamtstichprobe, die ca. 44% Kommunen aus Niedersachsen, ca. 43% aus Nordrhein-Westfalen und ca. 13% aus Sachsen-Anhalt umfasst, sind die Kommunen aus Nordrhein-Westfalen in den erhobenen Daten mit einem Anteil von ca. 34% leicht unterrepräsentiert, während die Kommunen aus den beiden anderen Ländern ganz leicht überrepräsentiert sind. Ca. 50% der Kommunen, die an der Befragung teilgenommen haben, liegen im Bereich zwischen 25.000 bis 50.000 Einwohnerinnen und Einwohnern. Sehr kleine Kommunen mit unter 10.000 Einwohnerinnen und Einwohnern sind insgesamt mit ca. 25% etwas unterrepräsentiert, ebenso wie Kommunen mit 50.000 bis 100.000 Einwohnerinnen und Einwohnern (ca. 5%). Etwa 5% der eingegangenen Antworten kommen aus Landkreisen. Insgesamt ist die Beteiligung zwar kleiner als in der vergleichbaren Befragung von Burth und Hilgers (2012), dafür ist die vergleichsweise große Gruppe der Kommunen mit weniger als 10.000 Einwohnerinnen und Einwohnern besser repräsentiert.

Ca. 87% der Personen, die teilgenommen haben, sind Führungskräfte aus den Finanzbereichen oder andere Führungskräfte, in diesem Fall überwiegend Bürgermeisterinnen und Bürgermeister. Die übrigen Teilnehmerinnen und Teilnehmer sind fast ausschließlich Beschäftigte aus Finanzbereichen. Vermutlich handelt es sich in diesen Fällen um die Vertretungen der angeschriebenen Führungskräfte, an die die Anfragen automatisiert oder manuell weitergeleitet wurden.

Nicht zuletzt auf Basis einiger per E-Mail eingegangener Rückmeldung ist zu vermuten, dass sich Kommunen, die die Einführung der DOPPIK als kritisch oder falsch bewerten, an der Befragung eher nicht teilgenommen haben. Insgesamt erlauben die erhobenen Daten aber sinnvolle Aussagen über die Gesamtsituation nach der DOPPIK-Einführung in den drei untersuchten Bundesländern, und zentrale Ergebnisse sollten auch auf andere Bundesländer mit vergleichbaren rechtlichen Regelungen extrapolierbar sein.

Der Fragebogen umfasste allgemeine Fragen zu Lage, Art und Größe der Kommune, den antwortenden Personen sowie geschlossene Fragen zu Daten der DOPPIK-Einführung, zum Beispiel zum Beginn des Einführungsprozesses, dem Zeitpunkt der Vorlage der Eröffnungsbilanz, der Haushaltssituation und der Umsetzung einer Kosten- und Leistungsrechnung (KLR). Die Fragen zu Problemen im Einführungsprozess sowie zum Stand der Einführung von Performance Management und Strategischen Managements wurden in Form einer Aussage mit einer fünfteiligen Likert-Skala – „trifft nicht zu“ bis „trifft voll zu“ – gestellt. *Tabelle 2* gibt einen Überblick über die genutzten Fragen und Indizes, die in verschiedenen Vorprojekten getestet wurden.

**Tabelle 2:** Fragen und Indizes zur Messung von Performance Management und Strategischem Management

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>PM1: Wir erfassen systematisch Daten und verarbeiten sie in Form von Kennzahlen zu verschiedenen Sachverhalten (z.B. Projekten, Produkten, Maßnahmen).</p> <p>PM2: Die Kennzahlen werden zur Steuerung von Ausgaben und Leistungen auf der operativen Ebene (innerhalb der Verwaltung) genutzt.</p> <p>PM3: Die operativen Kennzahlen werden an die Politik berichtet.</p> <p>PM4: Die operativen Kennzahlen werden regelmäßig bei politischen Entscheidungen berücksichtigt.</p> <p>IPM: Index Performance Management, Mittelwert aus PM1 bis PM4</p> |
| <p>SM1: Unsere Kommune hat klar definierte strategische Ziele.</p> <p>SM2: Die Politik steuert unsere Kommune über diese strategischen Ziele.</p> <p>SM3: Die strategischen Ziele werden in regelmäßigen Abständen systematisch überprüft und bei Bedarf angepasst.</p> <p>SM4: Wir haben Kennzahlen zur Messung der Erreichung der strategischen Ziele.</p> <p>ISM: Index Strategisches Management, Mittelwert aus SM1 bis SM4</p>                                                                                                                       |

Quelle: Eigene Darstellung.

Die Daten wurden deskriptiv ausgewertet und auf Grundlage des analytischen Modells (*Abbildung 1*) mit einem *Partial Least Squares* Strukturgleichungsmodell (Hair, Sarstedt, Pieper & Ringle, 2012; Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017) in der Software SmartPLS ([www.smartpls.com](http://www.smartpls.com), Ringle, da Silva & Bido, 2015) analysiert.

## 5 Ergebnisse

Alle 266 Kommunen, die an der Befragung teilgenommen haben, hatten spätestens 2015 auf einen doppischen Haushaltsplan umgestellt, über die Hälfte bereits vor dem Jahr 2010. Ca. 80% der Kommunen haben bis 2015 eine Eröffnungsbilanz vorgelegt, ca. 5% geben an, derzeit noch keine Eröffnungsbilanz fertiggestellt zu haben. Ungefähr 60% der Kommunen verfügen über einen aktuellen Jahresabschluss für die Jahre 2017 oder 2018, allerdings geben viele Kommunen aus Sachsen-Anhalt an, nicht über einen Jahresabschluss aus den letzten Jahren zu verfügen. Ca. 45% der Kommunen verfügen über mindestens einen konsolidierten Gesamtabschluss. Wie zu erwarten war, haben die Kommunen, die an der Befragung teilgenommen haben, die Umstellung auf eine doppische Haushaltsführung weitgehend vollzogen, wobei bei der Umsetzung von Gesamtabschlüssen noch Probleme zu bestehen scheinen, teilweise wurden entsprechende Regelungen aber auch mittlerweile gelockert (Ministerium des Innern des Landes Nordrhein-Westfalen, 2018).

*Tabelle 3* zeigt die Bewertungen der DOPPIK-Reform durch die Befragten. Knapp über 50% der Befragten stimmen den Aussagen zu, die DOPPIK ermögliche einen realistischeren Blick auf die Haushaltssituation und die Einführung sei prinzipiell sinnvoll gewesen zu. Auch die Akzeptanz der DOPPIK bei den Ratsmitgliedern wird leicht überwiegend positiv bewertet, wobei nur ungefähr ein Viertel der Befragten davon ausgeht, dass die Ratsmitglieder die „Vorteile der DOPPIK verstanden haben“.



Tabelle 3: Bewertung der DOPPIK

|                                                                                                                     | Mittelwert | Standard-<br>abweichung | % min. tw.<br>Zustimmung | Anzahl volle<br>Zustimmung |
|---------------------------------------------------------------------------------------------------------------------|------------|-------------------------|--------------------------|----------------------------|
| Die DOPPIK hat insgesamt dazu geführt, dass wir einen realistischen Überblick über unsere Haushaltssituation haben. | 3.44       | 1.008                   | 51.1%                    | 36                         |
| Die DOPPIK wird von den Ratsmitgliedern akzeptiert.                                                                 | 3.44       | .947                    | 51.3%                    | 28                         |
| Unsere Ratsmitglieder haben die Vorteile der DOPPIK verstanden.                                                     | 2.86       | .967                    | 23.9%                    | 10                         |
| Insgesamt ist die Umstellung auf die DOPPIK für uns eine sinnvolle Maßnahme gewesen.                                | 3.40       | 1.071                   | 52.1%                    | 37                         |

n=266; 5-stufige Likert-Skala: keine Zustimmung = 1 / volle Zustimmung = 5

Quelle: Eigene Daten.

Tabelle 4 zeigt die deskriptive Auswertung der Daten zur Umsetzung der beiden untersuchten managerialistischen Instrumente Performance Management und Strategisches Management. Insgesamt ist die Implementierung managerialistischer Instrumente unerwartet gering. Nur eine Kommune stimmt der Aussage, operative Kennzahlen würden bei politischen Entscheidungen berücksichtigt, voll zu (PM4). Auch für operative Entscheidungen scheint die Nutzung von Kennzahlen eher selten (PM2). Während noch vergleichsweise viele Kommunen angeben, zumindest teilweise über strategische Ziele zu verfügen (SM1) ist auch hier die Relevanz für politische Entscheidungen gering (SM2).

Tabelle 4: Ergebnisse zur Umsetzung von Performance Management und Strategischem Management

|                                                                                                                                                        | Mittelwert | Standard-<br>abweichung | % min. tw.<br>Zustimmung | Anzahl volle<br>Zustimmung |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-------------------------|--------------------------|----------------------------|
| PM1: Wir erfassen systematisch Daten und verarbeiten sie in Form von Kennzahlen zu verschiedenen Sachverhalten (z.B. Projekten, Produkten, Maßnahmen). | 2.45       | 1.081                   | 17.3%                    | 9                          |
| PM2: Die Kennzahlen werden zur Steuerung von Ausgaben und Leistungen auf der operativen Ebene (innerhalb der Verwaltung) genutzt.                      | 2.121      | .913                    | 5.3%                     | 5                          |
| PM3: Die operativen Kennzahlen werden an die Politik berichtet.                                                                                        | 2.42       | 1.251                   | 21.4%                    | 18                         |
| PM4: Die operativen Kennzahlen werden regelmäßig bei politischen Entscheidungen berücksichtigt.                                                        | 1.88       | .817                    | 3.4%                     | 1                          |
| SM1: Unsere Kommune hat klar definierte strategische Ziele.                                                                                            | 2.52       | 1.127                   | 19.6%                    | 13                         |
| SM2: Die Politik steuert unsere Kommune über diese strategischen Ziele.                                                                                | 1.98       | .967                    | 7.1%                     | 2                          |
| SM3: Die strategischen Ziele werden in regelmäßigen Abständen systematisch überprüft und bei Bedarf angepasst.                                         | 2.15       | 1.069                   | 7.1%                     | 4                          |
| SM4: Wir haben Kennzahlen zur Messung der Erreichung der strategischen Ziele                                                                           | 1.93       | .997                    | 1.9%                     | 2                          |

n=266; 5-stufige Likert-Skala: keine Zustimmung = 1 / volle Zustimmung = 5

Quelle: Eigene Daten.

Mit diesen Ergebnissen lässt sich Hypothese 1 vorläufig bestätigen: Nur ein geringer Teil der befragten Kommunen hat managerialistische Instrumente implementiert. Ca. 17% erheben nach eigenen Angaben PI, ca. 20% geben an, über strategische Ziele zu verfügen. Die Relevanz von PI und strategischen Zielen für politische Entscheidungen wird als sehr gering eingeschätzt.

Die Auswertung der Daten ergibt für die gebildeten Indizes eine hohe Konsistenz mit einem Cronbachs Alpha von .877 für den IPM ( $AVE=.729$ ) und .890 für den ISM ( $AVE=.750$ ) und es gibt keine zu starken Kollinearitätseffekte. Die Analyse des Pfadmodells (*Abbildung 1*) erfolgte mit dem einfachen PLS-SEM Algorithmus der SmartPLS Software bei pfadbezogenen Gewichtungen (Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017), die Überprüfung mit der SmartPLS *Bootstrapping*-Funktion. Fehlende Werte wurden paarweise ausgeschlossen. Die maximale Anzahl der auf ein Konstrukt gerichteten, zu analysierenden Beziehungen im Modell ist sieben für den ISM. Bei 266 Datensätze kann auch bei kleinen Werten von  $R^2$  ein Signifikanzniveau von 1% erwartet werden (Cohen, 1992; Hair, Hult, Ringle, Sarstedt, Richter & Hauff, 2017).

*Tabelle 5: Ergebnisse der PLS-SEM-Analyse auf Basis des Analysemodells aus Abbildung 1*

|                             | Hypothesen | erwartete Wirkungen | Pfad Koeffizienten nach PLS-SEM Schätzung |                           |                            | Ergebnis            |
|-----------------------------|------------|---------------------|-------------------------------------------|---------------------------|----------------------------|---------------------|
|                             |            |                     | Performance Management                    | Strategisches Management  | DOPPIK-Einführung sinnvoll |                     |
| Größe der Kommune           | H2, H2'    | +                   | 0.173**<br>( $f^2=.031$ )                 | 0.115**<br>( $f^2=.018$ ) |                            | vorläufig bestätigt |
| Start der Doppik Einführung | H3, H3'    | +                   | 0.123                                     | 0.016                     |                            | nicht bestätigt     |
| Finanzielle Probleme        | H4a, H4a'  | -                   | -0.027                                    | 0.091                     |                            | nicht bestätigt     |
| Personalprobleme            | H4b, H4b'  | -                   | -0.132                                    | -0.023                    |                            | nicht bestätigt     |
| Konkordanzindex             | H5, H5'    | +                   | -0.061                                    | 0.006                     |                            | nicht bestätigt     |
| Vertrauen Rat Verwaltung    | H6, H6'    | +                   | -0.028                                    | 0.061                     |                            | nicht bestätigt     |
| Performance Management      | H7, H8     | +/+                 |                                           | 0.532**<br>( $f^2=.384$ ) | 0.317**<br>( $f^2=.084$ )  | vorläufig bestätigt |
| Strategisches Management    | H8         | +                   |                                           |                           | 0.156<br>( $f^2=.020$ )    | vorläufig bestätigt |

\*\* signifikant auf .05 Niveau  
alle nicht angegebenen  $f^2$ -Werte  $<.020$

|       | Performance Management | Strategisches Management | DOPPIK-Einführung sinnvoll |
|-------|------------------------|--------------------------|----------------------------|
| $R^2$ | .108                   | .342                     | .180                       |
| $Q^2$ | .065                   | .222                     | .160                       |

Quelle: Eigene Daten.

*Tabelle 5* gibt einen Überblick über die Ergebnisse der PLS-SEM Schätzung. Im Gesamtmodell ergibt sich für den IPM ein  $R^2=.108$ , für den ISM  $R^2=.341$  und für die Einschätzung der Befragten zur Sinnhaftigkeit der DOPPIK-Einführung  $R^2=.180$ . Entsprechend Joseph F. Hair, Tomas M. Hult, Christian M. Ringle, Marko Sarstedt, Nicole F.

Richter und Sven Hauff (2017) werden  $f^2$ -Werte unter .020 als Hinweis darauf interpretiert, dass kein relevanter Effekt besteht.  $f^2$ -Werte über .020 sind in *Tabelle 5* angegeben. Nachgewiesene, positive Effekte bestehen zwischen der Größe der Kommune und der Implementierung eines Performance Managements, zwischen Performance Management und Strategischem Management sowie zwischen den beiden untersuchten managerialistischen Instrumenten und der Einschätzung der DOPPIK-Einführung als „sinnvoll“ durch die Befragten. Der deutlich stärkste Zusammenhang besteht zwischen Performance Management und Strategischem Management, wobei die Berechnung mit einem inversen Modell zeigt, dass Performance Management als Voraussetzung für die Implementierung eines Strategischen Managements interpretiert werden kann.

Damit sind die Hypothesen H1, H2, H7 und H8 vorläufig bestätigt. Für alle anderen, in der Literatur genannten Faktoren kann kein Einfluss auf die Implementierung managerialistischer Praxen auf Basis der vorliegenden Daten nachgewiesen werden.

## 6 Diskussion der Ergebnisse

Die Ergebnisse bestätigen vielfach vorliegende Befunde, nach denen die Einführung eines doppelten Rechnungswesens in den Bundesländern, in denen entsprechende rechtliche Verpflichtungen bestehen, als weitgehend abgeschlossen gelten kann (Burth & Hilgers, 2012; KGSt, 2015; Burth, 2017). Allerdings ist nur knapp die Hälfte der Befragten zumindest teilweise der Meinung, dass die DOPPIK die Haushaltssituation realistischer darstellt und dass die Reform sinnvoll war.

Zu konstatieren ist eine unerwartet schwache Implementierung managerialistischer Instrumente in den Kommunen der Stichprobe. Über 90% der befragten Kommunen haben auch nach teilweise deutlich mehr als vier Jahren doppelter Haushaltsführung keine Instrumente für eine „neue Steuerung“ implementiert. Die Werte zwischen den Bundesländern unterscheiden sich dabei nicht signifikant. Mit Blick auf die Tatsache, dass die Stichprobe auch Kommunen aus Nordrhein-Westfalen umfasst, wo Aspekte des Neuen Steuerungsmodells noch am deutlichsten mit der DOPPIK-Einführung verbunden wurden, gibt es wenig Anlass, nicht davon auszugehen, dass dieses Ergebnis zur Verbreitung von managerialistischen Instrumenten auf die anderen Bundesländer übertragbar ist. Die bekannten Praxen strategischer Steuerung zum Beispiel in Kempten, Lemgo, Mannheim sowie den Landkreisen Ebersberg, Potsdam-Mittelmark, Soest usw. (Isabella Proeller in Ferlie & Ongaro, 2015; Burth, 2016; Bayerischer Landkreistag, 2017; Weiss, 2017; Weiß, 2019b) stehen mit großer Wahrscheinlichkeit nicht an der Spitze einer breiten Entwicklung, sondern sind als Einzelfälle zu interpretieren.

Das SEM-Modell hat insgesamt einen niedrigen Erklärungswert in Bezug auf die Implementierung managerialistischer Instrumente ( $R^2=.108$ ) und die positive Bewertung der Reform durch die Befragten ( $R^2=.180$ ). Der höhere Erklärungswert für die Einführung eines strategischen Managements ( $R^2=.342$ ) resultiert im Wesentlichen aus dem Zusammenhang zwischen Performance Management und strategischem Management. Bei beiden Indikatoren sind die Koeffizienten für die Verfügbarkeit von Kennzahlen bzw. Zielen deutlich größer als die für die Relevanz dieser für die operative oder strategische Steuerung. Die überwiegende Zahl der Kommunen, die überhaupt mit managerialistischen Instrumenten arbeiten, tun dies also ohne relevanten Einfluss von Zielen und Kennzahlen auf Managemententscheidungen.

Insgesamt reichen die verfügbaren Theorien zur Implementierung von Performance Management und Strategischem Management, die in Abschnitt 3 referiert wurden, nicht aus, um die Implementierung managerialistischer Instrumente für die zu beobachtenden Fälle hinreichend zu erklären. Zwar stützten die Ergebnisse dieser Untersuchung die Vermutung, dass managerialistische Instrumente eher in größeren Kommunen implementiert werden, aber dieser Faktor hat nur eine sehr geringe Erklärungskraft. Die stärkere Implementierung in größeren Kommunen könnte dabei auch darauf zurückzuführen sein, dass mit zunehmender Größe der Organisation die Ausdifferenzierung der Verwaltungsgliederungs- und Produktrahmenpläne steigt und rechtliche Verpflichtungen für die Einrichtung eines fachbezogenen Controllings, z. B. für Aufgaben im Sozialbereich, greifen.

Aus den vorliegenden Daten ergibt sich insbesondere weder ein signifikanter noch ein starker Reifeffekt. Dabei ist zu beachten, dass in der hier genutzten Stichprobe die Effekte einer früheren Einführung in Nordrhein-Westfalen durch die dort zu erwartende größere Intensität konkurrenzdemokratischer Konflikte überlagert werden könnten. Insgesamt ergeben sich aber keine Hinweise darauf, dass die, von den Befürworterinnen und Befürworthern der DOPPIK-Reform erwartete, zweite Stufe der Einführung mit einer umfassenderen Implementierung managerialistischer Instrumente derzeit oder zukünftig erreicht wird. Die Vermutung, dass eine umfassende Veränderung von Steuerungs- bzw. Managementprozessen und -verhalten nicht durch – auch noch sehr unkonkrete – formale Regeln gesteuert werden kann (Weiß, 2013a), wird damit gestützt.

Kritisch zu berücksichtigen sind die Beschränkung auf eine Stichprobe von Kommunen aus drei Bundesländern und die Datenerhebung über Selbstauskünfte der Kommunen, die einen *Common Source Bias* implizieren könnten (Meier & O'Toole, 2013a; 2013b). Mit Blick auf mögliche Selbstselektionseffekte und die erheblichen Differenzen, die zwischen realen managerialistischen Praxen in deutschen Kommunen nachgewiesen wurden (Weiss, 2019a), könnten die Ergebnisse zur Verbreitung managerialistischer Praxen im Vergleich zur Gesamtzahl der Kommunen insgesamt eher etwas zu hoch liegen. Insbesondere ist davon auszugehen, dass Kommunen, welche die DOPPIK-Einführung grundsätzlich negativ bewerten und daher auch keine Ziele oder Kennzahlen eingeführt haben, tendenziell weniger an der Befragung teilgenommen haben. Vor dem Hintergrund der in der Stichprobe angelegten Unterschiede, insbesondere zwischen den gesetzlichen Regelungen in den Bundesländern, scheinen uns die Ergebnisse hinsichtlich der gefundenen bzw. nicht nachweisbaren Kausalbeziehungen dagegen als relativ robust. Diese sollten auf Basis von Daten aus weiteren Bundesländern überprüft werden. Warum und unter welchen Bedingungen managerialistische Instrumente in deutschen Kommunen implementiert und genutzt werden, bleibt eine, nicht nur verwaltungswissenschaftlich, sondern auch für zukünftige *policies*, die das Ziel haben, den öffentlichen Sektor zu transformieren, zentrale Frage.

## 7 Zusammenfassung

Mit Blick auf die vorliegenden Ergebnisse wird deutlich, dass der von der Ständigen Konferenz der Innenminister und -senatoren der Länder (2003) formulierte und im Reformprozess von vielen Akteuren kommunizierte Zusammenhang zwischen einer Umstellung der Buchführung bzw. Rechnungslegung einerseits und der Reform der Steue-

rungs- bzw. Managementstrukturen andererseits weder logisch zwingend ist noch für die Praxis überzeugend war. Die Umstellung der Rechnungslegung konnte im Wesentlichen durch die Erarbeitung einer neuen Kontenstruktur und die Einführung einer entsprechenden Software realisiert werden. Das Ergebnis ist für die Aufsichtsbehörden einfach zu überprüfen. Die geforderten Veränderungen in der Verwaltungssteuerung erfordern dagegen umfassende Veränderungen von Organisationskultur, Prozessen und Managementkompetenzen. Die formale Erfüllung der Vorgaben zur Ausweisung von Zielen und Kennzahlen ließe sich zwar einfach überprüfen, bei der Durchsicht von Haushaltsplänen offensichtliche Defizite scheinen von den Aufsichtsbehörden aber nicht nachdringlich beanstandet zu werden. Es ist unwahrscheinlich, dass sich im Bereich der Verwaltungssteuerung in den nächsten Jahren umfangreichere, durch die DOPPIK-Einführung initiierte Veränderungen zeigen werden. Die vorliegenden Ergebnisse stützen damit die bereits seit einigen Jahren gestellte Diagnose, dass der zweite Teil der Reform - die Veränderung der Verwaltungssteuerung durch die Einführung managerialistischer Instrumente – gescheitert ist (Holtkamp, 2008; Bogumil & Holtkamp, 2012; Schmid, 2012). Dies scheint auch einer der Gründe dafür zu sein, dass nur ca. die Hälfte der befragten Kommunen die DOPPIK-Einführung insgesamt für sinnvoll hält. Die erhobenen Daten zeigen, dass die Implementierung von Ansätzen des Performance Managements oder des Strategischen Managements einen Einfluss auf die Zufriedenheit mit der DOPPIK-Reform insgesamt hat.

Vor diesem Hintergrund ist es bemerkenswert, dass die für einige Länder vorliegenden, als Evaluationen bezeichneten Umsetzungsberichte für die Landesparlamente das Scheitern der Zielstellungen für eine Veränderung der Verwaltungsführung kaum aufgreifen und in diesem Bereich keinen Bedarf für Verbesserungen oder Nachsteuerung konstatieren. Die formalen rechtlichen Anforderungen waren sicher unzureichend, um einen Wandel der Führungskultur und der Verwaltungsteuerung zu initiieren (Weiß, 2013b). Ob das Eigeninteresse der Kommunen an managerialistischen Instrumenten aus Sicht der die Reform steuernden Ministerien und Aufsichtsbehörden überschätzt wurde oder die „neue Steuerung“ eher zufällig mit der DOPPIK-Einführung verbunden und *top down* nie ernsthaft betrieben wurde, ist eine für die *Policy*-Forschung interessante Frage, die letztlich auf Differenzen zwischen narrativen Strukturen und materiellen Inhalten von Reformen verweist. Verwaltungswissenschaftlich wie auch für die Praxis der kommunalen Verwaltungen relevant bleibt die Frage, welche Rahmenbedingungen und internen Faktoren die Einführung managerialistischer Instrumente fördern oder erleichtern und wie diese Instrumente langfristig in deutschen Kommunen wirken.

## Literatur

- Adam, Berit & Mahnke, Matthias (2018). Bilanzierung in deutschen Gebietskörperschaften. In Karl Petersen & Christian Zwirner (Hrsg.), *Handbuch Bilanzrecht*. Köln: Bundesanzeiger Verlag.
- Ammons, David N. & Rivenbark, William C. (2008). Factors influencing the use of performance data to improve municipal services. Evidence from the North Carolina benchmarking project. *Public Administration Review*, 68, 304-318.
- Andrews, Rhys, Boyne, George, Law, Jennifer & Walker, Richard M. (2012). *Strategic management and public service performance*. Basingstoke, Hampshire: Palgrave Macmillan.
- Askim, Jostein (2009). The demand side of performance measurement. Explaining councillors' utilization of performance information in policymaking. *International Public Management Journal*, 12, 24-47.

- Bayerischer Landkreistag (2017). Leitfaden zur zielorientierten Steuerung und zum strategischen Management. Teil 1 – Leitfaden. München.
- Bevan, Gwyn & Hood, Christopher (2006). What's measured is what matters. Targets and gaming in the English public health care system. *Public Administration*, 84, 517-538.
- Bleyen, Pieter, Klimovský, Daniel, Bouckaert, Geert & Reichard, Christoph (2016). Linking budgeting to results? Evidence about performance budgets in European municipalities based on a comparative analytical model. *Public Management Review*, 19(7), 1-22.
- Böhme, Doris (2017). Die Diffusion von Verwaltungsreformen. Eine Analyse der Reform des kommunalen Haushalts- und Rechnungswesens aus neoinstitutionalistischer Perspektive. Bamberg: University of Bamberg Press.
- Böhme, Doris & Heindl, Markus (2014). Die Reform des Haushalts- und Rechnungswesens - Neue Herausforderungen für die Koordination im föderalen Deutschland. *dms - der moderne staat*, 7(1), 173-191.
- Bogumil, Jörg, Ebinger, Falk & Holtkamp, Lars (2011). Vom Versuch, das Neue Steuerungsmodell verpflichtend einzuführen. Wirkungen des Neuen Kommunalen Finanzmanagements in NRW. *VM Verwaltung & Management*, 17(4), 171-180.
- Bogumil, Jörg, Ebinger, Falk & Holtkamp, Lars (2012). Vom Versuch, unerfreuliche Ergebnisse als normative Wissenschaft abzutun. Eine Replik auf die Replik von Christoph Reichard. *VM Verwaltung & Management*, 18(1), 1-6.
- Bogumil, Jörg, Holtkamp, Lars, Junkernheinrich, Martin & Wagschal, Uwe (2014). Ursachen kommunaler Haushaltsdefizite. *Politische Vierteljahresschrift*, 55(4), 614-647.
- Bogumil, Jörg & Holtkamp, Lars (2012). Doppik in der Praxis: Bisher vor allem intransparent und ineffizient. *VM Verwaltung & Management*, 18(3), 115-117, Stand: 14.05.2014.
- Bogumil, Jörg & Holtkamp, Lars (2013). *Kommunalpolitik und Kommunalverwaltung. Eine praxisorientierte Einführung*. Bonn: bpb Bundeszentrale für Politische Bildung.
- Bryson, John M., Crosby, Barbara C. & Bryson, John K. (2009). Understanding Strategic Planning and the Formulation and Implementation of Strategic Plans as a Way of Knowing. The Contributions of Actor-Network Theory. *International Public Management Journal*, 12, 172-207.
- Bryson, John M. (2015). Strategic planning for public and nonprofit organizations. In James D. Wright (Ed.), *International Encyclopedia of the Social & Behavioural Sciences* (pp. 515-521). Elsevier.
- Burth, Andreas (2016). *Strategische Ziele deutscher Kommunen im Bereich der Finanzen*. Online verfügbar unter: <https://www.haushaltssteuerung.de/weblog-strategische-ziele-deutscher-kommunen-im-bereich-der-finanzen.html> [21.09.2019].
- Burth, Andreas (2017). *Verbreitung der kommunalen Doppik in Deutschland*. Online verfügbar unter: <https://www.haushaltssteuerung.de/weblog-verbreitung-der-kommunalen-doppik-in-deutschland.html> [02.09.2019].
- Burth, Andreas & Hilgers, Dennis (2012). Welchen Mehrwert bringt die kommunale Doppik?. *VM Verwaltung & Management*, 18(3), 141-146.
- Burth, Andreas & Hilgers, Dennis (2014). Cui bono? Depicting the benefits of the new municipal budgeting and accounting regime in Germany. *Journal of Business Economics*, 84, 531-570. Online verfügbar unter: <https://link.springer.com/article/10.1007/s11573-013-0698-9> [20.09.2019].
- Charbonneau, Etienne & Bellavance, François (2012). Blame Avoidance in Public Reporting. *Public Performance & Management Review*, 35, 399-421.
- Cohen, Jacob (1992). A power primer. *Psychological bulletin*, 112, 155-159.
- Conteh, Charles (2014). Strategic Management at the Local Level: City-Regions Managing Global Economic Change. In Paul Joyce, John M. Bryson & Marc Holzer (Eds.), *Developments in Strategic and Public Management. Studies in the US and Europe* (pp. 196-210). Palgrave Macmillan.
- Courty, Pascal & Marschke, Gerald (2004). An Empirical Investigation of Gaming Responses to Explicit Performance Incentives. *Journal of Labor Economics*, 22, 23-56.
- Dimaggio, Paul J. & Powell, Walter W. (1991). The iron cage revisited: Institutional Isomorphism and collective rationality in organizational fields. In Walter W. Powell & Paul J. Dimaggio (Eds.), *The New Institutionalism in Organizational Analysis* (pp. 63-82). University of Chicago press.

- Dimitrijevska-Markoski, Tamara & French, P. Edward (2019). Determinants of Public Administrators' Use of Performance Information: Evidence from Local Governments in Florida. *Public Administration Review*, 79(5), 699-709. doi:10.1111/puar.13036.
- Ferlie, Ewan & Ongaro, Edoardo (2015). Strategic management in public services organizations. Concepts, schools and contemporary issues. New York: Routledge.
- Fischer, Edmund (2012). Neues Haushalts- und Rechnungswesen und Neue Steuerung: Funktioniert es und wenn ja, warum nicht?. *VM Verwaltung & Management*, 18(3), 147-150.
- George, Bert, Walker, Richard M. & Monster, Joost (2019). Does Strategic Planning Improve Organizational Performance? A Meta-Analysis. *Public Administration Review*, 79(6), 810-819.
- George, Bert, Van de Walle, Steven & Hammerschmid, Gerhard (2019). Institutions or Contingencies? A Cross-Country Analysis of Management Tool Use by Public Sector Executives. *Public Administration Review*, 79(3), 330-342.
- Hair, Joseph F., Sarstedt, Marko, Pieper, Torsten M. & Ringle, Christian M. (2012). The Use of Partial Least Squares Structural Equation Modeling in Strategic Management Research: A Review of Past Practices and Recommendations for Future Applications. *Long range planning*, 45(5/6), 320-340.
- Hair, Joseph F., Hult, Tomas M., Ringle, Christian M., Sarstedt, Marko, Richter, Nicole F. & Hauff, Sven (2017). *Partial Least Squares Strukturgleichungsmodellierung (PLS-SEM). Eine anwendungsorientierte Einführung*. München: Franz Vahlen.
- Hilgers, Dennis & Burth, Andreas (2012). *Der Mehrwert der kommunalen Doppik in Deutschland aus Sicht von Kämmerern und Haushaltspolitikern*. Online verfügbar unter: <http://www.doppik-studie.de/> [15.05.2014].
- Holliman, Alvin E. & Bouchard, Marcia (2015). The Use of Management by Objectives in Municipalities: Still Alive?. *Review of Public Administration and Management*, 3(1). Online verfügbar unter: <https://www.longdom.org/open-access/the-use-of-management-by-objectives-in-municipalities-still-alive-2315-7844-1000150.pdf> [29.09.2019].
- Holtkamp, Lars (2008). Das Scheitern des Neuen Steuerungsmodells. *dms – der moderne staat*, 1(2), 423-446.
- Hood, Christopher (2007). What happens when transparency meets blame-avoidance?. *Public Management Review*, 9(2), 191-210.
- Hood, Christopher (2011). *The blame game. Spin, bureaucracy, and self-preservation in government*. Princeton NJ u. a.: Princeton University Press.
- Innenministerium des Landes Nordrhein-Westfalen & steria mummert consulting (2008). *NKF – Ergebnisse der NKF-Befragung 2008*. o. O.
- Jimenez, Benedict S. (2014). City Government: Strategic Management and Fiscal Health In Paul Joyce, John M. Bryson & Marc Holzer (Eds.), *Developments in Strategic and Public Management. Studies in the US and Europe* (pp. 179-195). Palgrave Macmillan.
- Johnsen, Åge (2017). Impacts of strategic planning and management in municipal government. An analysis of subjective survey and objective production and efficiency measures in Norway. *Public Management Review*, 20, 397-420.
- Joyce, Paul (2015). *Strategic management in the public sector*. New York: Routledge.
- Kommunale Gemeinschaftsstelle für Verwaltungsmodernisierung (KGSt) (1993). *Das Neue Steuerungsmodell*. Begründung. Konturen. Umsetzung, Köln.
- KGSt (2008). Stand der Einführung des neuen Haushalts- und Rechnungswesens. Ergebnisse einer bundesweiten Umfrage. Köln.
- KGSt (2010). Stand der Einführung des neuen Haushalts- und Rechnungswesens. Ergebnisse einer bundesweiten Umfrage. Köln.
- KGSt (2015). *Rechtsgrundlagen und Stand der Umsetzung der Doppik in den Bundesländern*. Online verfügbar unter: [http://www.doppikvergleich.de/\\_Resources/Persistent/0d8c3a9b6d6c34370bfae7595d7fbc04585a1505/Umsetzungsstand-Doppik.pdf](http://www.doppikvergleich.de/_Resources/Persistent/0d8c3a9b6d6c34370bfae7595d7fbc04585a1505/Umsetzungsstand-Doppik.pdf) [02.09.2019].
- Kroll, Alexander (2012). Why Public Managers Use Performance Information. Concepts, Theory, and Empirical Analysis. Potsdam.

- Kroll, Alexander (2015). Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research. *Public Performance & Management Review*, 38, 459-486. <http://dx.doi.org/10.1080/15309576.2015.1006469>.
- Kroll, Alexander & Proeller, Isabella (2012). Steuerung mit Kennzahlen in den kreisfreien Städten. Ergebnisse einer empirischen Studie. Gütersloh.
- Kwon, Myungjung, Berry, Frances S. & Jang, Hee Soun (2014). To Use or Not to Use Strategic Planning: Factors City Leaders Consider to Make This Choice. In Paul Joyce, John M. Bryson & Marc Holzer (Eds.), *Developments in Strategic and Public Management. Studies in the US and Europe* (pp. 163-178). Palgrave Macmillan.
- Landtag Brandenburg (2012). Bericht zur Evaluierung der Kommunalverfassung für das Land Brandenburg. Drucksache 5/4747. Potsdam.
- Landtag Rheinland-Pfalz (2015). *Bericht der Enquete-Kommission 16/1 „Kommunale Finanzen“*. Drucksache 16/5250. Mainz.
- Landtag von Sachsen-Anhalt (2014). Einführung des Neuen Kommunalen Haushalts- und Rechnungswesens (NKHR). Drucksache 6/3707. Magdeburg.
- Lüder, Klaus (Hrsg.) (1996). Konzeptionelle Grundlagen des Neuen Kommunalen Rechnungswesens. Speyerer Verfahren. Stuttgart.
- Lüder, Klaus (2001). Neues öffentliches Haushalts- und Rechnungswesen. Anforderungen, Konzept, Perspektiven. Berlin: Ed. Sigma.
- McDavid, James C. & Huse, Irene (2012). Legislator Uses of Public Performance Reports: Findings from a Five-Year Study. *American Journal of Evaluation*, 33, 7-25.
- Meier, Kenneth J. & O'Toole, Laurence J. (2013a). Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships. *Journal of Public Administration Research and Theory*, 23, 429-456.
- Meier, Kenneth J. & O'Toole, Laurence J. (2013b). I Think (I Am Doing Well), Therefore I Am: Assessing the Validity of Administrators' Self-Assessments of Performance. *International Public Management Journal*, 16, 1-27. <https://doi.org/10.1080/10967494.2013.796253> [21.09.2019].
- Ministerium des Innern des Landes Nordrhein-Westfalen (2018). Zweites Gesetz zur Weiterentwicklung des Neuen Kommunalen Finanzmanagements für Gemeinden und Gemeindeverbände im Land Nordrhein-Westfalen und weiterer kommunalrechtlicher Vorschriften. 2. NKFWG NRW.
- Ministerium für Heimat, Kommunales, Bau und Gleichstellung des Landes Nordrhein-Westfalen (2017). Zweite Evaluation des Neuen Kommunalen Finanzmanagements (NKF) in Nordrhein-Westfalen – Zwischenbericht zur Evaluation auf der Grundlage von Artikel 10 § 1 des 1. NKF-Weiterentwicklungsgesetzes. Düsseldorf.
- Moynihan, Donald P. (2008). *The dynamics of performance management. Constructing information and reform*. Washington, D.C: Georgetown University Press.
- Moynihan, Donald P. & Pandey, Sanjay K. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory*, 20, 849-866, Stand: 26.04.2016.
- Mühlenkamp, Holger & Glöckner, Andreas (2009). Rechtsvergleich kommunale Doppik: eine Synopse und Analyse ausgewählter Themenfelder des neuen, doppelischen kommunalen Haushaltsrechts der Bundesländer. Speyer: Deutsches Forschungsinstitut für Öffentliche Verwaltung.
- Nielsen, P. A. & Baekgaard, Martin (2015). Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform. Evidence from an Experiment. *Journal of Public Administration Research and Theory*, 25, 545-569.
- Pasha, Obed Q., Poister, Theodore H. & Edwards, Lauren H. (2015). Mutual Relationship of Strategic Stances and Formulation Methods, and Their Impacts on Performance in Public Local Transit Agencies. *Administration & Society*, 50, 884-910.
- Poister, Theodore H. & Streib, Gregory D. (1994). Municipal management tools from 1976 to 1993: An overview and update. *Public Productivity & Management Review*, 115-125.
- Poister, Theodore H. & Streib, Gregory D. (2005). Elements of strategic planning and management in municipal government: Status after two decades. *Public Administration Review*, 65, 45-56.



- Polzer, Tobias (2010). Verwendung von Performance-Informationen in der öffentlichen Verwaltung. Eine Untersuchung der Berliner Sozialhilfeverwaltung. Potsdam: Universitätsverlag Potsdam.
- Proeller, Isabella & Siegel, John Philipp (2009). Performance Management in der deutschen Verwaltung – Eine explorative Einschätzung. *dms – der moderne staat*, 2(2). Online verfügbar unter: <https://www.budrich-journals.de/index.php/dms/article/view/2889/2424>.
- Rechnungshof Rheinland-Pfalz (2011). *Kommunalbericht 2011*. Speyer.
- Reichard, Christoph (2011). Vom Versuch, NSM und NKF in einen Topf zu werfen. *VM Verwaltung & Management*, 17, 283-287.
- Reichard, Christoph (2012). Umsetzung und Praxis des neuen kommunalen Haushalts- und Rechnungswesens. *VM Verwaltung & Management*, 18, 118-121.
- Reichard, Christoph (2014). Leistungsinformationen im neuen Kommunalhaushalt – welche Rolle spielen diese Daten in der Praxis?. *VM Verwaltung & Management*, 20, 125-129.
- Ringle, Christian, da Silva, Dirceu & Bido, Diogenes (2015). Structural Equation Modeling with the SmartPLS. *Brazilian Journal Of Marketing*, 13, 56-73. Online verfügbar unter: 10.5585/remark.v13i2.2717 [20.09.2019].
- Ritz, Adrian (2005). 10 Jahre Verwaltungsreform in den Schweizer Kantonen. In: Andreas Lienhard (Hrsg.), *10 Jahre New Public Management in der Schweiz. Bilanz, Irrtümer und Erfolgsfaktoren* (S. 47-67). Bern: Haupt.
- Rivenbark, William C., Fasiello, Roberta & Adamo, Stefano (2018). Exploring Performance Management in Italian Local Government. The Necessity of Outcome Measures and Citizen Participation. *The American Review of Public Administration*, 49(5), 545-553. <https://doi.org/10.1177/0275074018775125>.
- Rivenbark, William C. & Kelly, Janet M. (2003). Management innovation in smaller municipal government. *State and Local Government Review*, 35, 196-205.
- Sächsisches Staatsministerium des Innern (2016). Bericht des Sächsischen Staatsministeriums des Innern über die Evaluierung der Regelungen zur Erforderlichkeit eines Haushaltsstrukturkonzeptes sowie zum Haushaltsausgleich im Neuen Kommunalen Haushalts- und Rechnungswesen (Kommunale Doppik). In *Drucksache 6/6016* (S. 22-44). Dresden.
- Saliterer, Iris & Korac, Sanja (2013). Performance information use by politicians and public managers for internal control and external accountability purposes. *Critical Perspectives on Accounting*, 24, 502-517.
- Saliterer, Iris & Korac, Sanja (2014). The discretionary use of performance information by different local government actors – analysing and comparing the predictive power of three factor sets. *International Review of Administrative Sciences*, 80, 637-658.
- Schmid, Andreas (2012). *Verwaltungsreform im politischen System*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Schmid, Andreas (2014). Die Reform des Haushalts- und Rechnungswesens nimmt keinen Einfluss auf die politische Steuerung - Ergebnisse einer empirischen Analyse und Bewertung. *dms – der moderne staat*, 7(1), 193-214. <https://doi.org/10.3224/dms.v7i1.16243>.
- Siegel, John Philipp (2019). Strategisches Management. In Sylvia Veit, Christoph Reichard & Göttrik Wewer (Hrsg.), *Handbuch zur Verwaltungsreform* (S. 1-12). Bd. 40. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-21571-2>.
- Ständige Konferenz der Innenminister und -senatoren der Länder (IMK) (2003). Beschlussniederschrift über die 173. Sitzung der Ständigen Konferenz der Innenminister und -senatoren der Länder am 21. November 2003 in Jena.
- Tolkemitt, Dirk (2014). Strategische Konsolidierung am Beispiel der Alten Hansestadt Lemgo. In Jens Weiß (Hrsg.), *Strategische Haushaltskonsolidierung in Kommunen* (S. 229-254). Wiesbaden: Springer Fachmedien Wiesbaden.
- van Dooren, Wouter, Bouckaert, Geert & Halligan, John (2015). *Performance management in the public sector*. London, New York, NY: Routledge.
- Walker, R. M. & Andrews, Rhys (2015). Local Government Management and Performance. A Review of Evidence. *Journal of Public Administration Research and Theory*, 25, 101-133.

- Weiß, Jens (2013a). Die Transformation der kommunalen Verwaltung unter dem Einfluss des Neuen Steuerungsmodells. *dms – der moderne staat*, 6(1), 235-255.
- Weiß, Jens (2013b). Kennzahlen zur Steuerung kommunaler Verwaltungen. In Franz-Reinhard Habel & Jürgen Stember (Hrsg.), *Wissenstransfer zwischen Hochschule und Kommunen* (S. 323-351). Wien: Lit-Verl..
- Weiß, Jens (2014). Evaluation der DOPPIK-Einführung in den Kommunen des Landes Sachsen-Anhalt. Halberstadt: DEU.
- Weiss, Jens (2017). Trust as a key for strategic management? The relevance of council-administration relations for NPM-related reforms in German local governments. *Public Management Review*, 19, 1399-1414.
- Weiss, Jens (2019a). Managing Performance and Strategy: Managerial Practices in German Local Governments. *Public Performance & Management Review*, 29, published online 12/2019, <https://doi.org/10.1080/15309576.2019.1702064>.
- Weiß, Jens (2019b). Strategisches Management und wirkungsorientierte Steuerung in Kommunen. Wiesbaden: Springer.

*Anschriften der Autoren:*

Prof. Dr. Jens Weiß, Hochschule Harz, Domplatz 16, 38820 Halberstadt,  
E-Mail: [jweiss@hs-harz.de](mailto:jweiss@hs-harz.de).

Prof. Dr. Dino André Schubert, Hochschule Harz, Domplatz 16, 38820 Halberstadt,  
E-Mail: [dinoschubert@hs-harz.de](mailto:dinoschubert@hs-harz.de).

Svenja Bauer-Blaschkowski

## Politik nach Präferenzlage oder pragmatisches Problemlösen? Landesregierungen und die Unterbringung von Asylbewerbern (1995-2016)

### Zusammenfassung

Während asylpolitische Entscheidungen des Bundes als ideologiegetrieben beschrieben werden, gelten die Kommunen seit Jahrzehnten als Vorreiter einer pragmatischen Asylpolitik. Unklar ist bislang, wodurch die Asylpolitiken der Bundesländer geprägt werden. Dieser Fragestellung geht das vorliegende Papier anhand der Einflussfaktoren auf die Unterbringungsart von Asylbewerbern nach. Es wird zunächst angenommen, dass Unterbringungspolitiken durch länderspezifischen sozioökonomischen Problemdruck geprägt werden. Kontrastierend wird ausgehend von der Parteiendifferenzhypothese argumentiert, dass unterschiedliche Landesregierungen die Unterbringungsart von Asylbewerbern gemäß ihren *policy*- und *vote-seeking*-Motiven verschiedenartig ausgestalten, diese im Wählerstimmenwettbewerb mit rechtspopulistischen Parteien aber restriktiver ausrichten. Die Befunde *gepooled* Zeitreihenanalysen von 1995 bis 2016 zeigen, dass demografischer Problemdruck, steigende Asylbewerberzahlen und wirtschaftliche Ressourcen die Art der Unterbringung nur teilweise erklären. Vielmehr wird sie insbesondere durch SPD und Union entsprechend ihren ideologischen Motiven beeinflusst. Auf asylkritischere Einstellungen ihrer Stammwählerschaft reagieren beide mit einer restriktiveren Unterbringungspolitik. Im Wettbewerb mit der AfD verändern die Unionsparteien ihre Unterbringungspolitik restriktiv, die SPD dagegen permissiv.

**Schlagworte:** Asylpolitik; Unterbringung von Asylbewerbern; Bundesländer; Parteien

### Abstract

*Party Politics or pragmatic problem-solving? Länder governments and the accommodation of asylum seekers (1995-2016)*

Which motives shape the asylum policies of the German *Länder*? I investigate this question by focusing on accommodation of asylum seekers. First, I assume that accommodation policy depends on socioeconomic pressures and resources. Contrasting this, I argue that *Länder* governments shape accommodation policy according to their respective policy- and vote-seeking preferences. However, in competition with right-wing parties, they uniformly shift towards more restrictive measures. The results of pooled time series analyses (1995-2016) show that demographic pressure, increasing numbers of asylum seekers and economic resources only partly explain different accommodation policies. Rather, SPD and Union basically implement their policy preferences, but react differently to pressure from their electorate and from other parties. If their core electorate is xenophobic, both adopt more restrictive policies. Competing with the AfD, the SPD pursues more permissive policies, while the Union adapts restrictively.

**Keywords:** asylum policy; accommodation of asylum seekers; German *Länder*; party politics

## 1 Einleitung

In der medialen und politischen Debatte rund um die „Flüchtlingskrise“<sup>1</sup> galt die Unterbringung von Asylbewerbern<sup>2</sup> als Paradebeispiel für problemdruckinduziertes Handeln staatlicher Stellen: Angesichts drastisch angestiegener Asylbewerberzahlen, allgegenwärtigen Wohnraummangels und einer finanziellen Überlastung der öffentlichen Haushalte sei eine Unterbringung von Flüchtlingen in großen Sammelunterkünften unumgänglich und eine Unterbringung in kleineren Wohneinheiten schlichtweg nicht realisierbar. Gleichwohl wird die Unterbringung von Asylbewerbern aufgrund ihrer Auswirkungen als Ausdruck unterschiedlicher asylpolitischer Zielvorstellungen angesehen. Die Unterbringung in so genannten „Gemeinschaftsunterkünften“ gilt wegen ihrer separierenden und abschreckenden Wirkung als restriktive Asylpolitik, die sogenannte „dezentrale Unterbringung“ in eigenen Wohneinheiten wegen ihrer integrativen, das Wohlbefinden von Asylbewerbern steigernden Wirkung als permissive Asylpolitik (u. a. Aumüller & Bretl, 2008; Johansson, 2016).

Maßgeblich ausgestaltet wird die Unterbringung von Asylbewerbern durch die Bundesländer. Ein Blick in die Empirie zeigt hierbei erhebliche Unterschiede: Während Asylbewerber in Bundesländern wie Brandenburg oder Bayern seit Jahren überwiegend in Gemeinschaftsunterkünften leben, werden in Niedersachsen oder Schleswig-Holstein mehr als 90% der Asylbewerber dezentral untergebracht. Auch über Zeit lassen sich keine bundesweit einheitlichen Trends in der Art der Unterbringung ausmachen: Während der Anteil dezentraler Unterbringung in Baden-Württemberg oder dem Saarland in den letzten Jahren deutlich sank, ist er in Thüringen oder Bayern angewachsen.

Aber wodurch lässt sich diese Varianz erklären? Eine Betrachtung der Forschungsliteratur zeigt, dass sich bislang keine Antwort auf diese Frage finden lässt. Trotz ihrer Relevanz im asylpolitischen Kompetenzgeflecht ist die Unterbringungspolitik der Bundesländer ein politikwissenschaftlich vernachlässigtes Feld. Bisherige Untersuchungen zu Bestimmungsfaktoren der Asylpolitik der Bundesländer befassen sich mit der Zusammensetzung und Arbeit von Härtefallkommissionen (Wolf, 2011), Asylanerkennungswerten (Riedel & Schneider, 2017), Ausgabenhöhen oder Abschiebequoten (Hörisch, 2018) und, folgt man einem weiteren Verständnis von Asylpolitik als Teil von Migrationspolitik, mit Integration (Münch, 2016; Kostner, 2017; Träger & Neumann, 2017) oder Einbürgerung (Henkes, 2008). Sie liefern unterschiedliche, zum Teil widersprüchliche, Befunde zur Relevanz von Problemdruck und finanziellen Ressourcen oder der parteipolitischen Zusammensetzung von Landesregierungen. Konsens über die Bestimmungsfaktoren von Asylpolitiken besteht bislang lediglich für die kommunale und die Bundesebene: Erstere gilt als pragmatisch und problemlösungsorientiert, letztere als ideologiegetrieben und politisiert.

Ausgehend von diesen möglichen Motiven für Regierungshandeln geht das vorliegende Papier daher folgenden Fragestellungen nach: Reagieren die Länder bei der Unterbringung von Asylbewerbern, ebenso wie die deutschen Kommunen, vorrangig auf regionale Problemlagen, wie Wohnraummangel oder steigende Asylbewerberzahlen, und versuchen diese entsprechend den vorhandenen Ressourcen zu bewältigen? Oder gestalten auch Landesregierungen – ähnlich wie Bundesregierungen – Unterbringungspolitiken gemäß eigenen politischen Zielvorstellungen oder wahlstrategischen Erwägungen permissiv oder restriktiv aus? Zur Beantwortung dieser Fragestellungen

erscheint die Art der Unterbringung von Asylbewerbern als empirisches Beispiel besonders geeignet, da sie zentral durch die Bundesländer ausgestaltet werden kann und durch diesen Handlungsspielraum die Motive für Regierungshandeln besonders deutlich werden sollten.

Im zweiten Teil des Papiers soll zunächst die Kompetenzverteilung und Ausgestaltung der Unterbringungspolitik in Deutschland dargestellt werden, um den Handlungsrahmen zu skizzieren, in dem Landesregierungen bei der Unterbringungspolitik agieren. Im dritten Teil werden unterschiedliche Motive für das Handeln von Landesregierungen und deren Auswirkungen auf die Art der Unterbringung theoretisch plausibilisiert und Hypothesen formuliert. Im vierten Teil wird das Forschungsdesign vorgestellt. Darauf folgen die Ergebnisse der empirischen Analyse mithilfe von *gepoolten* Zeitreihenanalysen von 1995 bis 2016. Das Papier schließt im sechsten Teil mit einem Fazit und möglichen Anknüpfungspunkten für weitere Forschung.

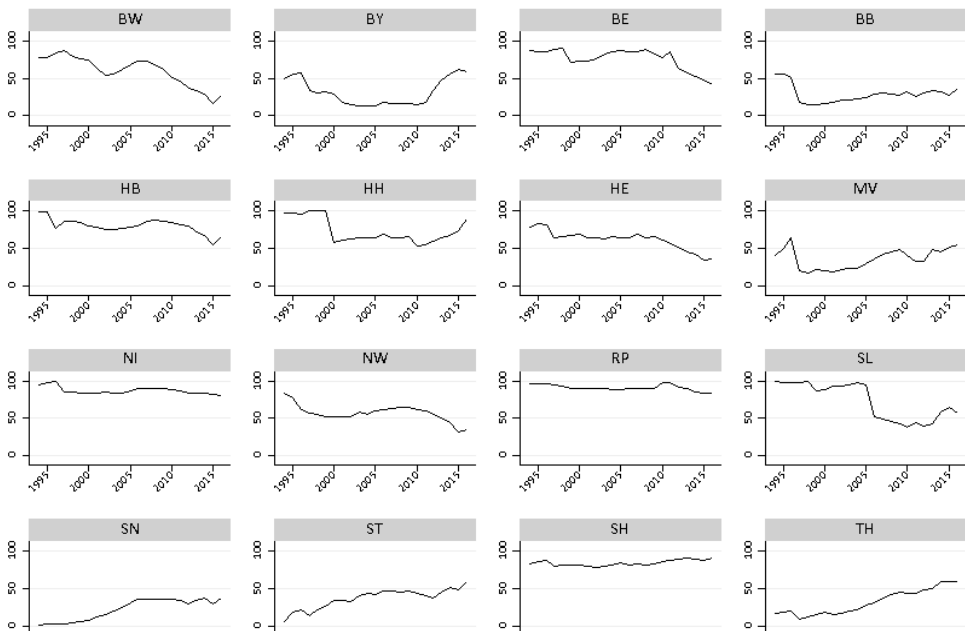
## 2 Unterbringung von Asylbewerbern: Kompetenzverteilung und politischer Gestaltungsspielraum

Die Unterbringung von Asylbewerbern in Deutschland wird im Rahmen geteilter Zuständigkeiten durch Regelungen auf der Bundes-, der Bundesländer- und der Kommunalebene bestimmt. Auf Bundesebene finden sich zentrale Regelungen der Unterbringung von Asylbewerbern im Asylgesetz (AsylG), im Aufenthaltsgesetz und im Asylbewerberleistungsgesetz (AsylbLG) (Wendel, 2014, S. 8). Der Bund regelt die Aufnahme, das bundesweite Verfahren zur Verteilung von Asylbewerbern und den Aufenthalt in sogenannten Erstaufnahmeeinrichtungen verbindlich. Zunächst werden Asylsuchende nach ihrer Ankunft gemäß dem Königsteiner Schlüssel auf die Bundesländer verteilt (§§45, 46 AsylG) und in sogenannten (Erst-) Aufnahmeeinrichtungen untergebracht, in denen auch die Erstanthörung im Asylverfahren stattfindet (Aumüller, Daphi & Biesenkamp, 2015, S. 21). In diesen sollen sie mindestens sechs Wochen, längstens aber drei Monate, verbleiben (Bogumil, Hafner & Kuhlmann, 2016, S. 129). Mit dem Asylverfahrensbeschleunigungsgesetz (2015) wurde festgelegt, dass Asylbewerber aus so genannten sicheren Herkunftsstaaten bis zum Ende des Asylverfahrens, aber maximal sechs Monate, in den Erstaufnahmeeinrichtungen verbleiben (Aumüller, 2018, S. 179f.). Zuständig dafür, die Erstaufnahmeeinrichtungen zu schaffen, zu unterhalten und eine entsprechende Anzahl an Unterbringungsplätzen bereitzustellen, sind nach §44 Abs. 1 AsylG die Bundesländer. Die Größe, Beschaffenheit oder die Ausstattung der Erstaufnahmeeinrichtung sind im AsylG nicht festgelegt (Wendel, 2014, S. 8).

Nach der Unterbringung in einer Erstaufnahmeeinrichtung werden Asylbewerber innerhalb eines Bundeslandes weiter verteilt (§50 AsylG). Dort „sollen“ sie laut dem AsylG „in der Regel in Gemeinschaftsunterkünften untergebracht werden“ (§53 Abs. 1 AsylG). Weitere Vorgaben zur Art und Ausgestaltung der sogenannten Anschlussunterbringung macht das AsylG nicht. Die Finanzierung obliegt ebenfalls den Bundesländern, seit Januar 2016 unterstützt der Bund die Länder allerdings mit einer Pauschale von 670 Euro pro Asylbewerber und Monat (Bundesregierung, 2018a). Aus dieser bundesgesetzlichen Regelung ergibt sich ein Ermessensspielraum: Die Länder sind nicht zur Errichtung und dem Betreiben von Gemeinschaftsunterkünften verpflichtet (Bergmann & Dienelt, 2016, Rn. 4-8). Die genaue Verteilung innerhalb des Bundes-

landes, die Art und Ausgestaltung der Anschlussunterbringung, behördliche Zuständigkeiten und Verfahren zur Finanzierung werden von den Bundesländern im Rahmen von (Flüchtlings-) Aufnahmegesetzen eigenständig geregelt (Aumüller, 2018, S. 180). Neben den Flüchtlingsaufnahmegesetzen werden in immer mehr Bundesländern Landesintegrationskonzepte und mittlerweile auch –gesetze verabschiedet, in denen sich ebenfalls Regelungen bzw. Zielvorstellungen in Zusammenhang mit der Anschlussunterbringung von Asylbewerbern finden (Aumüller, Daphi & Biesenkamp, 2015, S. 24). Die bestehenden Regelungen unterscheiden sich sowohl inhaltlich als auch in der Verbindlichkeit erheblich (Cremer, 2014, S. 6). Ein Blick in die Empirie zeigt, dass sich diese Varianz auch im Politik-*Outcome*, gemessen über den Anteil dezentraler Unterbringung, niederschlägt (Abbildung 1):

Abbildung 1: Anteil dezentraler Unterbringung in den Bundesländern 1994 bis 2016



Quelle: Eigene Darstellung basierend auf Bundesamt für Statistik, 2018h.

In den meisten Bundesländern werden Asylbewerber aus der vom Land betriebenen Erstaufnahmeeinrichtung auf die Kommunen (Landkreise und kreisfreie Städte) verteilt. Ausnahmen bilden hierbei Bayern, da dort die Regierungsbezirke für die Anschlussunterbringung zuständig sind, sowie die Stadtstaaten Hamburg und Berlin, in denen sämtliche Verantwortlichkeiten bei der Landesregierung liegen (Aumüller, 2018, S. 180). Kommunen übernehmen daher in den meisten Bundesländern eine ausführende Funktion bei der Anschlussunterbringung. Dementsprechend ist es möglich, dass auch Kommunen die dargestellte *Outcome*-Varianz beeinflussen. Dies sollte bei der Interpretation der Befunde im vorliegenden Papier berücksichtigt werden, stellt aber nicht den Fokus der Untersuchung dar. Über die Reichweite gesetzlicher Vorgaben, die Gliedrigkeit des Unterbringungssystems und die Art der Finanzierung können die Lan-

desregierungen aber Einfluss darauf nehmen, inwiefern die Kommunen hierbei einen eigenen Gestaltungsspielraum haben, oder an politische Vorstellungen der Landesregierung gebunden sind.

### 3 Theoretischer Rahmen

#### 3.1 Forschungsstand

Welchen Motiven folgen die Landesregierungen also bei der Ausgestaltung der Unterbringung von Asylbewerbern, einem ihrer zentralen asylpolitischen Kompetenzbereiche? Bislang findet sich in der Forschungsliteratur keine eindeutige Antwort auf diese Frage. Konsens besteht bislang nur über die Prägung von Asylpolitiken auf Bundes- und kommunaler Ebene: Übereinstimmend wird für die Asylpolitik der Bundesrepublik konstatiert, dass sich diese spätestens seit den 1980er Jahren angesichts steigender Flüchtlingszahlen und sich verändernder *Cleavage*-Strukturen durch ein zunehmendes Maß an parteipolitischer Ideologisierung, Polarisierung und Politisierung auszeichnet (Boswell, 2000, p. 550; Thränhardt, 1995, p. 330f.; Thränhardt, 2002, S. 223; Rotte, 2000, S. 377; Green, 2001, p. 83; Minkenberg, 2003, p. 224). Gegenläufig dazu wird argumentiert, dass sich die Asylpolitik auf der lokalen Ebene durch ein hohes Maß an Pragmatismus auszeichnet, da Problemen vor Ort unmittelbar und lösungsorientiert begegnet werden muss (Schmidtke, 2017, p. 510; Aumüller, 2018, S. 181f.). Analysen zu Bestimmungsfaktoren der Asylpolitik auf Bundesländerebene untersuchen sowohl ökonomische und soziale, als auch parteipolitische Einflussfaktoren, liefern aber bisher widersprüchliche Befunde. Während Frieder Wolf (2011) und Felix Hörisch und Matthias Heiken (2018) zeigen können, dass ökonomische Faktoren wie eine hohe Arbeitslosenquote und ein niedriges Bruttoinlandsprodukt zu einer restriktiveren Ausrichtung von Asylpolitiken führen, finden Felix Hörisch (2018) und Lisa Riedel und Gerald Schneider (2017) keine bzw. gegenläufige Effekte dieser Variablen. Hörisch und Heiken (2018) können darüber hinaus aufzeigen, dass ein höherer Ausländeranteil zu permissiveren Asylpolitiken führt, während Wolf (2011) und Riedel und Schneider (2017) dafür keine bzw. gemischte Evidenz finden. Das Bild zum Einfluss von Parteiendifferenzen ist ebenfalls uneinheitlich: Hendrik Träger und Alexandra Neumann (2018) und Sandra Kostner (2017) zeigen für Thüringen und Baden-Württemberg deutliche Effekte linker Landesregierungen auf eine permissivere Ausrichtung der Asylpolitik auf. Hörisch und Heiken (2018), Hörisch (2018) und Ursula Münch (2016) weisen diesen Effekt auch im Bundesländervergleich nach. Die Befunde von Hörisch (2018) und Christian Henkes (2008) deuten zudem darauf hin, dass bürgerliche Regierungen eher bereit sind, Asylbewerber nach Afghanistan abzuschieben bzw. systematisch niedrigere Einbürgerungsquoten verantworten. Dahingegen findet Wolf (2011) keinen Einfluss bürgerlicher Regierungen und einen restriktiven Effekt linker Parteien auf die Erfolgsquote von Härtefallkommissionen. Riedel und Schneider (2017) können in ihrer umfassenden Analyse zu Anerkennungsquoten ebenfalls keinerlei Parteieneffekte feststellen und kommen zu dem Schluss, dass die Höhe von Anerkennungsquoten vor allem durch sozioökonomische Parameter geprägt wird. Sämtliche der genannten Untersuchungen prüfen aber nur direkte Einflüsse der Regierungszusammensetzung und sozioökonomi-

scher Faktoren auf die Ausgestaltung von Asylpolitiken. Internationale Untersuchungen zur Asylpolitik gehen differenzierter vor: Zahlreiche Analysen gehen von der Überlegung aus, dass Asylpolitiken nicht nur durch sozioökonomische Kontextbedingungen und jeweilige Werte und *Policy*-Ziele unterschiedlicher Parteifamilien direkt geprägt werden, sondern Regierungsparteien auch Wählerpräferenzen oder parteienwettbewerbliche Erwägungen berücksichtigen und ihre Politiken gegebenenfalls kontextspezifisch justieren (u. a. Squire, 2008). So konnte beispielsweise Kyung Joon Han (2015) nachweisen, dass linke Parteien nur dann permissivere Asylpolitiken verabschieden, wenn sie sich nicht allein auf ungelernte Arbeiter, sondern auf eine breite Wählerbasis stützen. Einige Untersuchungen befassen sich zudem mit den Effekten von rechtspopulistischen Parteien. Sie zeigen beispielsweise, dass sowohl linke als auch rechte Parteien angesichts steigender Wahlerfolge von rechtspopulistischen Parteien ihre Politiken restriktiver ausrichten (Bale, Green-Pedersen, Krouwel, Luther & Sitter, 2010; Bolin, Lidén & Nyhlén, 2014).

Eine differenziertere Betrachtung erscheint daher auch für die Asylpolitik der Bundesländer lohnenswert, um etwaige Veränderungen direkter Effekte unterschiedlicher Landesregierungen aufzuspüren. Im Folgenden sollen wegen diesem unklaren Bild beide Motive für Regierungshandeln – Reaktion auf sozioökonomischen Problemdruck sowie ideologische Präferenzen von Regierungsparteien – theoretisch plausibilisiert und um *vote*- und *office-seeking*-Motive von Regierungsparteien ergänzt werden.

### 3.2 Unterbringung als Ergebnis bundeslandspezifischen Problemdrucks und vorhandener Ressourcen

Sozioökonomische Ansätze gehen davon aus, dass Politiken eine Reaktion auf sozioökonomische Modernisierungsprozesse und damit verbundene Herausforderungen und Problemlagen darstellen, sofern genügend finanzielle Ressourcen oder Möglichkeiten der Problembearbeitung zur Verfügung stehen (Obinger, 2015, S. 35). Landesregierungen sollten sich daher bei ihrer Unterbringungspolitik zum einen an den konkreten Problemlagen und dem sozioökonomischen Kontext ihres Bundeslandes orientieren (Riedel & Schneider, 2017, S. 31 ff.). Als besondere Herausforderungen bei der Unterbringung von Asylbewerbern können steigende Asylbewerberzahlen sowie die Unterbringung einer hohen Anzahl besonders schutzbedürftiger Asylbewerber gelten. Angesichts steigender Asylbewerberzahlen sollten Landesregierungen Asylbewerber in Sammelunterkünften unterbringen, da eine dezentrale Unterbringung bei stark erhöhten Zuzugszahlen logistisch kaum zu bewältigen ist. Als besonders schutzbedürftig, und in Gemeinschaftsunterkünften durch Gewalt und Übergriffe gefährdet, gelten weibliche Asylbewerber. Mit einem hohen Frauenanteil an Asylbewerbern sollte der Druck auf die Landesregierungen steigen, vermehrt dezentral unterzubringen, um gewaltsame Übergriffe zu verhindern:

*H1a: Steigende Asylbewerberzahlen und wenige weibliche Asylbewerber sollten einen niedrigen Anteil dezentraler Unterbringung bewirken.*

Inwiefern Landesregierungen Problemlagen bei der Unterbringung von Asylbewerbern begegnen können, sollte von ihren sozioökonomischen Ressourcen wie vorhandenem Wohnraum bzw. Wohnungsleerstand, ihrem finanziellen Spielraum, aber auch ihrem



„Sozialkapital“ in der Bevölkerung beeinflusst werden. Wohnraumknappheit im Bundesland sollte die dezentrale Unterbringung von Asylbewerbern erschweren, da notwendige Ressourcen für eine dezentrale Unterbringung fehlen und die Unterbringung in Einzelwohnungen kostspielig machen (Inforadio rbb, 2015). Dahingegen ist zu erwarten, dass Landesregierungen in Bundesländern, die mit einer Überalterung ihrer Bevölkerung, Landflucht und Wohnungsleerstand zu kämpfen haben, Asylbewerber vermehrt dezentral unterbringen werden, um vorhandenen Wohnraum zu nutzen, der Entvölkerung ganzer Landstriche entgegenzuwirken und Asylbewerber zur dauerhaften Ansiedlung zu bewegen (Busch, 2015 zit. nach Aumüller, Daphi & Biesenkamp, 2015, S. 62). Landesregierungen sollten sich zudem an den finanziellen Ressourcen ihres Bundeslandes orientieren. Denn vielfach wird argumentiert, dass begrenzte finanzielle Ressourcen zu restriktiven Asylpolitiken und damit zu einem niedrigen Anteil dezentraler Unterbringung führen (u. a. Breunig & Luedke, 2008; Han, 2013; Toshkov, 2014; Hörisch, 2018).<sup>3</sup> Schließlich kann angenommen werden, dass Landesregierungen sich bei ihrer Unterbringungspolitik auch am gesellschaftlichen Kontext, dem „Sozialkapital“ ihres Bundeslandes, orientieren. So ist zu erwarten, dass Asylbewerber in den Bundesländern, die schon länger durch Migration geprägt werden und in denen Erfahrung im Zusammenleben mit Ausländern besteht, entsprechend der Kontakthypothese routinemäßig in Gesellschaften aufgenommen werden. Dies sollte sich positiv auf den Anteil dezentraler Unterbringung auswirken (Weber, 2015, S. 422). Zudem ist anzunehmen, dass auch Jahre nach der Wiedervereinigung Differenzen zwischen Ost und West in Bezug auf Berührungspunkte mit Ausländern und Integrationserfahrung, aber auch in Hinblick auf autoritär-libertäre Wertvorstellungen bestehen und deshalb in Ostdeutschland Asylbewerber weniger dezentral untergebracht werden (Abedi, 2017, p. 473; Heckmann, 2015, S. 175 f.; Clark & Legge, 1997, p. 901; Hörisch, 2018):

*H1b: Wohnraumknappheit, knappe finanzielle Ressourcen, eine frühere Zugehörigkeit zur DDR und fehlende Erfahrung im Umgang mit Migration in einem Bundesland sollten einen niedrigen Anteil dezentraler Unterbringung bewirken.*

### 3.3 Parteien und Unterbringungspolitik

Parteien können als *policy*-, *vote*- oder *office-seeker* agieren. Die einzelnen Motive von Parteien wirken nicht isoliert, sondern in einem Zusammenspiel: Parteien evaluieren ihre ideologischen Ziele unter Berücksichtigung strategischer Überlegungen wie dem Kampf um Wählerstimmen im Parteienwettbewerb und passen ihre *Policies* ggf. an (Hinnfors, Spehar & Bucken-Knapp, 2012, pp. 588 f.). Dementsprechend ist es möglich, dass sich je nach Wählerschaft und Parteienlandschaft eines Bundeslandes Unterschiede in den Politiken von Landesregierungen einer Partei zeigen, wenngleich diese über gemeinsame ideologische Präferenzen verfügen (Joppke, 1997, S. 276; Schneider, 2010, S. 154, 171; Münch, 2014, S. 70 f.; Träger & Neumann, 2018, S. 435; Reutter, 2006, p. 287). Nachfolgend werden daher sowohl *policy*-, als auch *vote*- und *office-seeking*-Motive der unterschiedlichen Parteien dargestellt und in ihrem Zusammenspiel betrachtet.

*Linke Parteien* vertreten Werte wie internationale Solidarität, Gleichheit und kulturellen Pluralismus (Otto & Steinhardt, 2014, p. 69; Han, 2015, p. 602; Hinnfors, Spehar & Bucken-Knapp, 2012, p. 589). Sie sprechen sich deshalb für die Aufnahme von Asyl-

bedürftigen und ihre menschenwürdige Behandlung unter dem Aspekt der Gleichheit aller Menschen aus. Linke Parteien mit einem besonders libertären Wertekanon vertreten zudem Politiken, die ein friedliches und zugewandtes Zusammenleben mit Asylbewerbern im Rahmen einer multikulturellen Gesellschaft zum Ziel haben:

*H2: Je höher der Anteil linker Parteien an der Regierung ist, desto höher ist der Anteil dezentraler Unterbringung in einem Bundesland.*

Es ist aber zu erwarten, dass sich linke Parteien bei ihren *Policies* nicht nur an ihren Wertvorstellungen, sondern allen voran auch an ihrer klassischen Wählerklientel orientieren. Diese sollte sich sowohl in ihren Wertvorstellungen als auch in ihren sozioökonomischen Präferenzen und der daraus resultierenden Haltung zu Asylbewerbern zwischen den verschiedenen linken Parteien und zwischen den Bundesländern unterscheiden. Linke Parteien verfügen mitunter über eine klassische Wählerklientel, die eher gering qualifiziert ist, über ein niedriges Einkommen verfügt und deren soziale Position durch drohende Erwerbslosigkeit oder Wohlstandsverluste gefährdet ist. Viele dieser Wähler fühlen sich durch Asylbewerber in ihrer ökonomischen Existenz bedroht und in einem Wettbewerb mit diesen um Arbeitsplätze und finanzielle Ressourcen (Boswell, 2000, p. 553; Rippl & Seipel, 2018, S. 239 f.). Asylkritische Haltungen in dieser Wählerschaft werden auch durch Sorgen um traditionelle Identitätsbezüge und Wertesysteme befördert (Weber, 2015, S. 401 f.). Derartige ökonomische und kulturelle Bedrohungsgefühle und die daraus erwachsenden asylkritischen Haltungen sollten von denjenigen linken Parteien berücksichtigt werden, die sich dieser Wählerklientel besonders verbunden fühlen (Koning, 2017, p. 632; Han, 2015, p. 603). Linke Parteien mit dieser Wählerklientel verfolgen keine gänzlich asylablehnende Linie, sie verabschieden aber Maßnahmen, die sich gegen einen vermeintlichen Missbrauch der Sozialsysteme durch Asylbewerber richten oder durch die Leistungen weniger generös und am Sachleistungsprinzip orientiert ausgestaltet werden, und damit vermutlich eine etwas weniger permissive Unterbringungspolitik (Boswell, 2000, p. 553). Dahingegen sollten linke Parteien, deren Wählerklientel gut ausgebildet und finanziell abgesichert ist, eher libertäre, inklusive Werte vertritt und dementsprechend asylfreundlich eingestellt ist, von asylkritischen Haltungen absehen und eine permissive Unterbringungspolitik verfolgen (von Dirke, 1994, p. 526):

*H3: Je asylkritischer die Kernwählerschaft einer linken Regierungspartei eingestellt ist, desto schwächer ist der permissive Effekt linker Regierungsparteien auf den Anteil dezentraler Unterbringung.*

*Christdemokratische Parteien* vertreten vor allem traditionelle, kulturkonservative Wertvorstellungen und sprechen sich eher für eine nationale, homogene Kultur aus (Boswell & Hough, 2008, p. 339). Der Aufnahme von und dem Zusammenleben mit Menschen aus fremden Kulturen und mit anderer Religion stehen sie daher eher skeptisch gegenüber (Schmidtke, 2017, p. 499). Stattdessen präferieren sie Politiken, die darauf abzielen, potenzielle Asylbewerber von einer Einreise nach Deutschland abzuschrecken oder vor Ort die Berührungspunkte mit Asylbewerbern möglichst gering zu halten. Gleichwohl sind die Wertvorstellungen christdemokratischer Parteien auch durch ihren christlichen Wertekanon und die Grundsätze der Nächstenliebe und Humanität geprägt, die politischen Zielen der nationalen Abschottung und der gesellschaftlichen Separation von Asylbewerbern entgegenstehen (Boswell & Hough, 2008; Bale,

2008, p. 324; van Kersbergen & Krouwel, 2008, p. 399). Christdemokratische Parteien befinden sich daher in der Asylpolitik in einem besonderen Spannungsverhältnis konfligierender innerparteilicher Wert- und Zielvorstellungen. Vielfach findet sich jedoch in der Literatur die Auffassung, dass christdemokratische Parteien in der Asylpolitik eher einer restriktiven Position zuneigen und sich ähnlich wie konservative Parteien verhalten (Boswell & Hough, 2008, pp. 335 f.):

*H4: Je höher der Anteil christdemokratischer Parteien an der Regierung ist, desto niedriger ist der Anteil dezentraler Unterbringung in einem Bundesland.*

Das Spannungsverhältnis unterschiedlicher Wertvorstellungen spiegelt sich auch in der klassischen Wählerschaft christdemokratischer Parteien wider (Boswell & Hough, 2008, p. 333). In Bundesländern, in denen die klassische Wählerschaft weniger christlich, sondern vor allem traditionell und kulturkonservativ eingestellt ist und deshalb eher asylkritischen Positionen zuneigt (Ceobanu & Escandell, 2010, p. 321), sollte dieses zusätzliche *vote-seeking*-Motiv christdemokratische Parteien dazu bewegen, ihre Unterbringungspolitik deutlich restriktiv auszurichten:

*H5: Je asylkritischer die Kernwählerschaft einer christdemokratischen Partei in einem Bundesland eingestellt ist, desto stärker ist der restriktive Effekt christdemokratischer Regierungsparteien auf den Anteil dezentraler Unterbringung.*

*Liberalen Parteien* sollten sich bei Asylpolitiken vorrangig an einem ihrer zentralen Werte, der individuellen Freiheit, orientieren (Boswell & Hough, 2008, p. 336). Zugleich sollten sie sich von Überlegungen zu ökonomischen Vorteilen und hierbei von einer utilitaristischen Logik leiten lassen (Minkenberg, 2003, p. 232). Für liberale Parteien ist daher davon auszugehen, dass sie sich eher für eine dezentrale Unterbringung aussprechen, weil diese Asylbewerbern mehr Freiraum zur individuellen Lebensgestaltung bietet, eine schnellere gesellschaftliche Integration und dadurch mittelbar auch eine schnellere Integration in den Arbeitsmarkt ermöglicht (Koning, 2017, p. 632):

*H6: Je höher der Anteil liberaler Parteien an der Regierung ist, desto höher ist der Anteil dezentraler Unterbringung.*

Angesichts sich auflösender Schichten und Milieus scheint es in Erweiterung der Hibbs'schen Herleitung angezeigt, davon auszugehen, dass Parteien nicht nur die Präferenzen ihrer klassischen Wählerklientel, sondern auch die allgemeine Verteilung der Wählerschaft und ihrer Präferenzen berücksichtigen, um ihre Wiederwahl im Wettbewerb mit anderen Parteien zu sichern (Zohlhöfer, 2013, S. 268). Ob Regierungsparteien parteienwettbewerbliche Erwägungen tatsächlich in ihre Politikentscheidungen einbeziehen, hängt unter anderem davon ab, ob es im Politikfeld andere Parteien gibt, die alternative *Policy*-Positionen anbieten und diese von der Wählerschaft auch als glaubwürdig erachtet werden.<sup>4</sup> Im Politikfeld Asylpolitik sind vor allem rechtspopulistische und rechtsextreme Parteien<sup>5</sup> relevante Wettbewerber um Stimmenanteile. Die Asyl- und Migrationspolitik gehört zu ihrem Markenkern (Bolin, Lidén & Nyhlén, 2014, p. 324; Akkerman, 2012, p. 511). Rechte Parteien vertreten besonders deutliche Forderungen nach einer Verschärfung von Asylpolitiken und schüren bewusst Ressentiments gegenüber Asylbewerbern (Lewandowsky, 2017, S. 5 f.): Einerseits stellen sie sie als eine Bedrohung für die „Identität des Volkes“ dar und warnen vor einer Gefährdung der inneren Sicherheit durch ebendiese (Spier, 2010, S. 256; Lewandowsky, 2017, S. 5;

Akkerman, 2012, p. 516). Andererseits stellen rechte Parteien Asylbewerber wahlweise als unfaire Wettbewerber um Arbeitsplätze oder als „Sozialschmarotzer“ dar, die aufgrund ihrer anderen Nationalität als weniger zu Sozialleistungen berechtigt gelten und eine Bürde für den Wohlfahrtsstaat darstellen (Keskinen, 2016, p. 359). Rechte Parteien befinden sich mit ihren Positionen in einem Wettbewerb mit christdemokratischen Parteien um nationalistische, kulturkonservative, auf Sicherheit und Ordnung bedachte Wählerstimmen (Bale, 2008, p. 319; Boswell & Hough, 2008, p. 332, 345). Mit steigenden Wahlergebnissen rechter Parteien sollten christdemokratische Parteien daher ihre Positionierung justieren und tendenziell restriktiver ausrichten (Lewandowsky, 2017, S. 8):

*H7: Je höher der Stimmenanteil rechtsextremer/rechtspopulistischer Parteien ist, desto restriktiver ist der Effekt christdemokratischer Parteien auf den Anteil dezentraler Unterbringung.*

Zudem ist davon auszugehen, dass rechte Parteien auch linke Parteien durch die Betonung von Asyl und Immigration unter Druck setzen können, da diese Parteien im Politikfeld als wenig kompetent wahrgenommen werden (fehlende *issue ownership*) und eine derartige Themensetzung in der Regel nicht elektoral für sich nutzen können (Bale, Green-Pedersen, Krouwel, Luther & Sitter, 2010, p. 412). Zunehmende Wahlerfolge rechter Parteien bringen insbesondere auch die linken Parteien in Bedrängnis, die über eine eher asylkritische Wählerschaft verfügen. Rechte Parteien machen ihnen den Teil der Wählerklientel streitig, der (subjektive) wirtschaftliche Abstiegsängste hat (Bale, Green-Pedersen, Krouwel, Luther & Sitter, 2010, p. 412; Tutic & von Hermann, 2018, S. 284):

*H8: Je höher der Stimmenanteil rechtsextremer/rechtspopulistischer Parteien ist, desto weniger permissiv ist der Effekt linker Parteien, die über eine asylkritische Wählerschaft verfügen, auf den Anteil dezentraler Unterbringung.*

Abbildung 2: Hypothesenübersicht

| Hypothesen                                                                      | Vermuteter Einfluss auf den Anteil dezentraler Unterbringung |
|---------------------------------------------------------------------------------|--------------------------------------------------------------|
| H <sub>1a</sub> Sozioökonomischer Problemdruck                                  | Negativ                                                      |
| H <sub>1b</sub> Finanzielle Ressourcen                                          | Positiv                                                      |
| H <sub>2</sub> Linke Parteien                                                   | Positiv                                                      |
| H <sub>3</sub> Linke Parteien x asylkritische Wählerklientel                    | Positiv, schwächer werdend                                   |
| H <sub>4</sub> Christdemokratische Parteien                                     | Negativ                                                      |
| H <sub>5</sub> Christdemokratische Parteien x asylkritische Wählerklientel      | Negativ, stärker werdend                                     |
| H <sub>6</sub> Liberale Parteien                                                | Positiv                                                      |
| H <sub>7</sub> Rechte Parteien x christdemokratische Parteien                   | Negativ, stärker werdend                                     |
| H <sub>8</sub> Rechte Parteien x linke Parteien mit asylkritischer Wählerschaft | Positiv, schwächer werdend                                   |

Quelle: Eigene Darstellung.

## 4 Forschungsdesign

### 4.1 Fallauswahl

In der nachfolgenden Analyse soll die Art der Unterbringung von Asylbewerbern in den Bundesländern von 1994 bis 2016 betrachtet werden. Dieser Zeitraum wurde gewählt, weil eine ausschließliche Untersuchung eines Jahres während der „Flüchtlingskrise“ ungeeignet erscheint, um allgemeinere Aussagen über Einflussfaktoren auf die Unterbringungspolitik der Bundesländer treffen zu können (siehe auch die bisherige Untersuchung von Hörisch & Heiken, 2018). 1994 wurde als Untersuchungsbeginn gewählt, um eine einheitliche Erfassung der abhängigen Variablen, der Art der Unterbringung, zu ermöglichen. Im Zuge des „Asylkompromisses“ wurde 1992 das Asylrecht in der Bundesrepublik sowohl im Grundgesetz als auch einfachgesetzlich umfassend neu geregelt, sodass das Jahr als Zäsur in der deutschen Asylpolitik gilt. Im Rahmen der Reformen wurde auch das Asylbewerberleistungsgesetz eingeführt, auf dessen Definition von Leistungsberechtigten die Zahlen zur Unterbringung basieren (Bünthe, 2014, S. 219). Aus methodischen Gründen muss das Jahr 1994 in der Analyse allerdings aus den Berechnungen ausgeschlossen werden, sodass 1995 das Anfangsjahr ist (vgl. Kap. 4.3). Somit ergibt sich eine gesamte Fallzahl von 352 Länderjahren.

### 4.2 Operationalisierung

Als abhängige Variable (AV)<sup>6</sup> soll die Art der Unterbringung von Asylbewerbern in den Bundesländern, gemessen über den *Anteil dezentraler Unterbringung*, untersucht werden. Der Anteil dezentraler Unterbringung orientiert sich an der Erhebungspraxis der Bundesländer und des Statistischen Bundesamtes. Entsprechend bemisst er sich als Anteil derjenigen Empfänger von Leistungen nach §§2, 3 AsylbLG, die in dezentralen Unterkünften untergebracht sind, an der Gesamtzahl sämtlicher Leistungsempfänger in der Anschlussunterbringung. Als berechnete Leistungsempfänger nach §1 AsylbLG gelten Menschen, deren Asylantrag (noch) nicht positiv beschieden wurde. Asylberechtigte, Flüchtlinge nach der Genfer Menschenrechtskonvention und subsidiär Schutzberechtigte erhalten Sozialleistungen nach SGB II oder SGB XII (Wissenschaftliche Dienste des Bundestages, 2016, S. 5). Sie sind deshalb nicht durch die abhängige Variable erfasst. Dementsprechend können in der vorliegenden Arbeit lediglich Aussagen über die Unterbringung der Gruppe von „Asylbewerbern“ (vgl. Kapitel 1) getroffen werden. Laut der Definition des Statistischen Bundesamtes wird unter dezentrale Unterbringung diejenige Unterbringung gefasst, bei der „die Möglichkeit zur Nutzung einer eigenen Küche bzw. eines eigenen Sanitärbereichs mit den entsprechenden Entfaltungsmöglichkeiten besteht“, „auch dann, wenn es sich dabei um einen eigenen, abgeschlossenen Wohnbereich innerhalb einer nur von Asylbewerbern genutzten Unterkunft handelt“. Bundesweit einheitliche Mindeststandards zur dezentralen Unterbringung gibt es nicht. Insgesamt variieren damit das Verständnis und die lebensweltliche Ausgestaltung von dezentraler Unterbringung zwischen den Bundesländern, was bei der Interpretation der Befunde berücksichtigt werden muss.

Die unabhängigen Variablen werden wie folgt operationalisiert: Problemdruck, der aus steigenden Asylbewerberzahlen oder einem hohen Frauenanteil resultiert, wird

über die *Veränderung des Anteils von Leistungsempfängern* nach §§2,3 AsylbLG an der Bevölkerung eines Bundeslandes und den *Frauenanteil* an ebenjener Leistungsempfängergruppe erfasst. Um den Einfluss wirtschaftlicher Ressourcen abzubilden, werden entsprechend dem Vorgehen in der Forschungsliteratur das *BIP pro Kopf*, die *Arbeitslosenquote* und die *Kaufkraft* in die Modellierungen einbezogen. *Stadtstaaten-Dummies* und die *Bevölkerungsdichte* dienen dazu, Wohnraum-Ressourcen zu erfassen. Demographischer Problemdruck soll mithilfe der *Geburtenrate* abgebildet werden. *Ost-West*-Unterschiede werden über eine Dummy-Variable erfasst, Kontakt zu Ausländern über den *Ausländeranteil* eines Bundeslandes.<sup>7</sup> Partieneffekte werden mithilfe von *Kabinettsitzanteilen* (KSA) abgebildet. Da anzunehmen ist, dass Parteien an der Regierung ihren Einfluss auf den Anteil dezentraler Unterbringung (*outcome*) nicht unmittelbar, sondern erst mit gewisser zeitlicher Verzögerung durch die Implementierung entsprechender *Policies* ausüben können, werden die Kabinettsitzanteile in die nachfolgende Untersuchung mit einem *time-lag* von einem Jahr (bzw. t-2 und t-5 für die Robustheitschecks) einbezogen. Entsprechend der vorgenommenen Differenzierung in den Hypothesen gelten SPD, Bündnis 90/Die Grünen und PDS/Die Linke als linke Parteien, CDU und CSU als christdemokratische Parteien und die FDP als liberale Partei. Da die genannten linken Parteien sich in der Reichweite libertär-inklusive Wertvorstellungen und in ihrer Kernwählerschaft unterscheiden, werden sie nachfolgend gesondert betrachtet. Die *asylkritische Haltung von Kernwählerschaften*, die die Effekte von linken und christdemokratischen Parteien bundeslandspezifisch moderieren sollte (Hypothese 3 und Hypothese 5), wird mithilfe von Einstellungsdaten von Wählern mit Parteibindung erfasst und als Anteil an sämtlichen Wählern mit der jeweiligen Parteibindung in einem Bundesland wiedergegeben. Da asylkritische Haltungen lediglich zu Beginn der 1990er Jahre, 2015 und 2016 abgefragt wurden, muss auf Daten zu ausländerfeindlichen Einstellungen als *Proxy* zurückgegriffen werden. Der parteienwettbewerbliche Einfluss *rechtsextremer/-populistischer Parteien*, der in Hypothese 7 und Hypothese 8 vermutet wird, wird über den aufaddierten Stimmenanteil der Parteien NPD, DVU, REP und AfD bei Landtagswahlen erfasst (Andersen & Woyke, 2013; Franzmann, 2016, S. 457).<sup>8</sup> Die Variablen, für die ein Interaktionseffekt mit den Parteienvariablen vermutet wurde, gehen theoretisch plausibilisiert mit einem *time-lag* (t-1) in die Interaktionsmodellierungen ein.

### 4.3 Methode

Da sich das theoretische Interesse der vorliegenden Arbeit auf Zusammenhänge über Raum und Zeit hinweg erstreckt, soll die Überprüfung der Hypothesen mithilfe einer Regressionsanalyse *gepoolter* Zeitreihen erfolgen. *Gepoolte* Daten liegen gruppiert vor, zwischen den Makroeinheiten herrscht Heterogenität und es ist in der Regel davon auszugehen, dass ein Jahr mit dem Folgejahr kausal zusammenhängt. Daraus resultieren verschiedene Problematiken wie Nicht-Stationarität, serielle Autokorrelation, Heterogenität und Heteroskedastizität, denen durch ergänzende Modellspezifikationen begegnet werden muss (Wenzelburger, Jäckle & König, 2014, S. 123).

Begrenzten Problemen mit Nicht-Stationarität bei der abhängigen Variablen konnte durch den Ausschluss des Jahres 1994 aus den Berechnungen begegnet werden. Die Variable *Arbeitslosenquote* geht wegen deutlicher Probleme mit Nicht-Stationarität als

Veränderungsvariable in die Modelle ein. Wegen Multikollinearität wurden die Variablen *Ausländeranteil* sowie der *Ost-West-Dummy* aus den Modellierungen ausgeschlossen. Die *Kabinettsitzanteile von Unionsparteien* und *SPD* wurden wegen hoher Korrelation getrennt in die Modellierungen einbezogen. Im Folgenden wurde eine Modellierung geschätzt, mit der sich die direkten Effekte von Parteien überprüfen lassen und die sich durch eine möglichst hohe Erklärungskraft auszeichnet (nachfolgend: *Best-Fit-Modell*).

Mithilfe von *Panel Corrected Standard Errors* sowie alternativ durch die Aufnahme einer weiteren Variablen in das Modell, *Kontakt zu Ausländern* (nachfolgend: *Erweitertes Modell*), konnte leichten Problemen mit Heteroskedastizität begegnet werden. Tests auf serielle Autokorrelation zeigten für das *Best-Fit-Modell* mittelstarke Probleme. Methodische Vorbehalte (Kittel, 2005, S. 105; Wenzelburger, Jäckle & König, 2014, S. 135 f.) und die ähnlich lautenden Befunde unterschiedlicher Modellspezifikationen (*Best-Fit-Modell*, ADL-Modell, Modell mit dynamisierter AV, Prais-Winsten-Modell, *Erweitertes Modell*) sprachen dafür, die empirische Analyse auf dem *Best-Fit-Modell* mit PCSE zu basieren. Berechnungen der Modellierungen mit *Länderdummies* und der zugehörige F-Test zur Überprüfung auf Heterogenität zeigten, dass *fixed effects* für die Modellierungen benötigt werden. Allerdings interessierten bei der vorliegenden Analyse insbesondere Länderunterschiede, die durch *fixed effects* absorbiert würden. Stattdessen wurden systematisch Länder aus der Analyse ausgeschlossen (*Jackknife*).<sup>9</sup>

## 5 Ergebnisse und Diskussion

Zunächst sollen die *Best-Fit-Modelle* und die um die Variable *Kontakt zu Ausländern* *Erweiterten Modelle*, jeweils getrennt für die Kabinettsitzanteile von SPD (Modell 1 und Modell 3) bzw. der Unionsparteien (Modell 2 und Modell 4) betrachtet werden, um Aussagen über die direkten Effekte von sozioökonomischem Problemdruck und ideologischen Motiven von Parteien treffen zu können. Sämtliche Modelle weisen mit einem  $R^2$  zwischen 0.61 und 0.66 eine gute Gesamterklärungsgüte auf.

Für die verschiedenen Variablen, die sozioökonomischen Problemdruck abbilden, zeigt sich ein differenziertes Bild: Entsprechend der vermuteten Wirkrichtung von Hypothese 1a geht ein steigender *Anteil von weiblichen Asylbewerbern* statistisch signifikant mit einem steigenden Anteil dezentraler Unterbringung einher. Überraschenderweise liefern die Modelle keine Evidenz dafür, dass eine drastische *Zunahme des Anteils von Asylbewerbern* eine dezentrale Unterbringung erschwert oder gar unmöglich macht. Finanzielle und gesellschaftliche Ressourcen scheinen ebenfalls in unterschiedlichem Maße relevant für die Art der Unterbringung zu sein: Für die Variablen *Arbeitslosenquote* und *Konsum pro Kopf* ergeben sich sowohl positive als auch negative, aber durchgängig statistisch insignifikante, Effekte. Entgegen den Erwartungen in Hypothese 1b geht eine höhere *Geburtenrate* mit einem höheren Anteil dezentraler Unterbringung einher, allerdings ist der Effekt nur in zwei Modellierungen statistisch signifikant. Auch für die Argumentation, dass knapper Wohnraum bedingt, dass Asylbewerber weniger dezentral untergebracht werden können, findet sich in den Modellierungen keine Evidenz: Vielmehr zeigen sie, dass *Stadtstaaten* einen statistisch signifikant höheren Anteil an dezentraler Unterbringung aufweisen. Dieser Befund lässt sich

gegebenenfalls damit erklären, dass sämtliche Stadtstaaten in den vergangenen Jahren trotz Wohnraumknappheit Beschlüsse gefasst haben, Asylbewerber vorzugsweise bzw. sogar ausschließlich (Berlin) in Wohnungen unterzubringen (Aumüller, Daphi & Biesenkamp, 2015, S. 39; Flüchtlingsrat Bremen, 2015). Berechnet man die Modelle testweise für frühere Zeiträume, verschwindet der statistisch signifikante *Stadtstaaten*-Effekt. Insgesamt betrachtet zeigt sich damit, dass sozioökonomischer Problemdruck und vorhandene Ressourcen eines Bundeslandes die Art der Unterbringung von Asylbewerbern in den Bundesländern nur zu einem geringen Teil erklären können.

*Tabelle 1:* Bestimmungsfaktoren des Anteils dezentraler Unterbringung 1995-2016

|                                | Modell 1             | Modell 2             | Modell 3              | Modell 4              |
|--------------------------------|----------------------|----------------------|-----------------------|-----------------------|
| $\Delta$ Arbeitslosenquote     | 0.73<br>(0.96)       | 0.78<br>(0.95)       | -1.33<br>(1.87)       | -0.98<br>(1.84)       |
| $\Delta$ Asylbewerberanteil    | 4.05<br>(4.30)       | 3.29<br>(4.25)       | 14.68<br>(16.02)      | 15.23<br>(15.82)      |
| Anteil weiblicher Asylbewerber | 2.39 ***<br>(0.16)   | 2.44 ***<br>(0.16)   | 2.49 ***<br>(0.26)    | 2.55 ***<br>(0.25)    |
| Konsum pro Kopf                | 0.00<br>(0.00)       | 0.00<br>(0.00)       | -0.00<br>(0.00)       | -0.00<br>(0.00)       |
| Stadtstaat                     | 13.73 ***<br>(2.65)  | 11.61 ***<br>(2.66)  | 14.74 ***<br>(4.44)   | 12.97 ***<br>(4.44)   |
| Geburtenrate                   | 24.32 ***<br>(8.73)  | 23.29 ***<br>(8.67)  | 22.71<br>(15.09)      | 23.24<br>(14.84)      |
| Kontakt zu Ausländern          |                      |                      | 0.21<br>(0.15)        | 0.20<br>(0.15)        |
| KSA Union $t_{-1}$             |                      | -0.26 ***<br>(0.03)  |                       | -0.26 ***<br>(0.04)   |
| KSA SPD $t_{-1}$               | 0.24 ***<br>(0.03)   |                      | 0.25 ***<br>(0.05)    |                       |
| KSA Grüne $t_{-1}$             | -0.08<br>(0.09)      | -0.32 ***<br>(0.10)  | -0.17<br>(0.15)       | -0.42 **<br>(0.17)    |
| KSA Linke $t_{-1}$             | -0.24 ***<br>(0.09)  | -0.48 ***<br>(0.10)  | -0.27 *<br>(0.10)     | -0.53 ***<br>(0.16)   |
| KSA FDP $t_{-1}$               | 0.51 ***<br>(0.10)   | 0.27 ***<br>(0.09)   | 0.62 ***<br>(0.14)    | 0.37 ***<br>(0.14)    |
| Konstante                      | -84.20 ***<br>(8.43) | -59.45 ***<br>(8.89) | -88.13 ***<br>(14.89) | -62.98 ***<br>(15.41) |
| Fallzahl                       | 352                  | 352                  | 160                   | 160                   |
| $R^2$ (adj.)                   | 0.61                 | 0.62                 | 0.65                  | 0.66                  |

Anmerkung: Angaben in Klammern sind Standardfehler (Modelle 3 und 4) bzw. PCSE (Modelle 1 und 2).

\*  $p \leq 0.10$ ; \*\*  $p \leq 0.05$ ; \*\*\*  $p \leq 0.01$

Quelle: Eigene Darstellung.

Stattdessen verdeutlichen die Befunde, dass Regierungsparteien (ebenfalls) eine gewichtige Rolle bei der Ausgestaltung der Unterbringung von Asylbewerbern einnehmen. Insbesondere bei den beiden Volksparteien zeigen sich klare und robuste Parteidifferenzen: Entsprechend Hypothese 4 zeigen die Modelle 2 und 4, dass mit höherem *Kabinettsitzanteil der Unionsparteien* der Anteil dezentraler Unterbringung sinkt. Der Effekt ist auf dem 1%-Niveau statistisch signifikant. Die Unterbringungs politik der Unionsparteien scheint damit, wie theoretisch dargestellt, tatsächlich weniger durch christlich-inklusive, sondern durch kulturkonservative, exklusive Wertvorstellungen

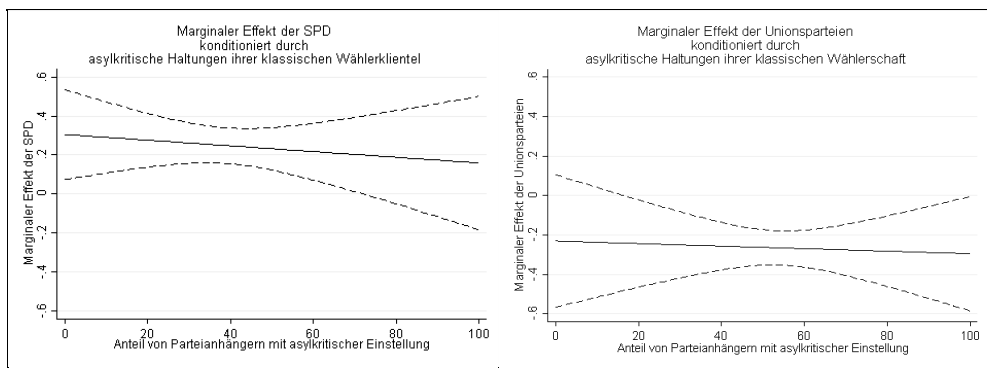


geprägt. Ein erhöhter *Kabinettsitzanteil der SPD* geht hingegen mit einem erhöhten Anteil dezentraler Unterbringung einher. Die Befunde sind ebenfalls auf einem 1%-Niveau statistisch signifikant. Robustheitschecks mithilfe der dynamisierten Modelle, der Modelle unter Ausschluss der Jahre 1994-2010, mit *fixed effects*, mit *time-lags* von zwei bzw. fünf Jahren für die Parteieneffekte, unter Einbezug der *gelagerten* abhängigen Variable sowie mithilfe der Befunde eines systematischen *Jackknife* von Jahren und Ländern stärken die vorgestellten Befunde: Sie bleiben auf einem 1%-Niveau entsprechend der vermuteten Wirkrichtung statistisch signifikant. Sowohl *Union* als auch *SPD* scheinen damit Asylpolitiken auf Bundesländerebene entsprechend ihren ideologischen Vorstellungen auf unterschiedliche Weise prägen zu können. Für die „kleineren Parteien“ lässt sich ein ambivalentes Bild zeichnen. Ein steigender *Kabinettsitzanteil der FDP* übt entsprechend Hypothese 6 einen statistisch signifikanten positiven Effekt auf den Anteil dezentraler Unterbringung aus, der in den meisten Robustheitschecks Bestand hat. Steigende *Kabinettsitzanteile von Grünen* und der *Linken* führen wider Erwarten zu einem sinkenden Anteil dezentraler Unterbringung. Besonders überraschend ist dieser Befund für *Bündnis 90/Die Grünen*, da diese aufgrund ihres besonders libertär-inklusive Wertekanons erwartungsgemäß die permissivste Unterbringungspolitik verfolgen sollten. Allerdings zeigen die verschiedenen Modelle und nachfolgende Robustheitschecks auch, dass die Befunde instabil sind: Die Effekte von *Grünen* und *Linken* schwanken je nach Modellierung zwischen völliger Insignifikanz und einem 1%-Signifikanz-Niveau. Wie lassen sich diese abweichenden und insignifikanten Ergebnisse erklären? Eine Ursachenforschung mithilfe des systematischen *Jackknife* deckt auf, dass der statistisch signifikante negative Effekt der Linkspartei offenbar maßgeblich durch bestimmte ostdeutsche Fälle geprägt ist. Sobald Brandenburg, Sachsen und/oder Mecklenburg-Vorpommern aus der Analyse ausgeschlossen werden, wird der Effekt insignifikant und mitunter positiv. Eine bivariate Korrelation der abhängigen Variablen *Anteil dezentraler Unterbringung* und des *Kabinettsitzanteils von Bündnis90/Die Grünen* zeigt, dass entsprechend der theoretischen Erwartung ein positiver Zusammenhang mit der abhängigen Variablen besteht. Dieser ist allerdings mit 0.15 sehr gering. Ursächlich für die insignifikanten Effekte von *Bündnis90/Die Grünen* könnte der Umstand sein, dass die Partei im Untersuchungszeitraum in höchst unterschiedlichen Koalitionen mit Parteien regierte, die sowohl über gleichgerichtete permissive, aber auch entgegengesetzte restriktive asylpolitische Präferenzen verfügten. In ihrer regelmäßigen Rolle als „Juniorpartner“ ohne relevante Ressortverantwortung vermochte die Partei vermutlich keine gleichlautende, permissive Linie in der Asylpolitik durchzusetzen. *Bündnis90/Die Grünen* führten im Untersuchungszeitraum lediglich in Rheinland-Pfalz und Bremen ab 2011 die Integrationsministerien mit Zuständigkeit für Asyl, sowie in Baden-Württemberg im letzten Untersuchungsjahr das Ministerium für Soziales und Integration. Für Innen- oder weitere Sozialministerien trugen sie keine Ressortverantwortung.<sup>10</sup> Da zu erwarten ist, dass sich auch bei den konditionierenden Effekten von Wählerpräferenzen und rechten Parteien für *Bündnis 90/Die Grünen* und *Die Linke* insignifikante bzw. durch die Fallauswahl geprägte gegenläufige Effekte zeigen, werden diese im Folgenden lediglich für SPD und Union fokussiert betrachtet.<sup>11</sup>

Um den moderierenden Einfluss der *asylkritischen Einstellungen der klassischen Wählerklientel* auf *linke* und *christdemokratische Parteien* und ihre Unterbringungspolitik zu untersuchen, wurden Interaktionsmodelle berechnet, die auf den *Best-Fit*

*Modellen* basieren. Um die Interpretation der Ergebnisse zu erleichtern, werden die Effekte grafisch mithilfe von *Marginal Effect Plots* betrachtet (Abbildung 3). Wie in Hypothese 5 vermutet, wird der restriktive Effekt der *Unionsparteien* etwas stärker, je asylkritischer die eigene Wählerschaft eingestellt ist. Dieser Effekt ist für einen Wertebereich von ca. 20% bis 100% statistisch signifikant. Entsprechend Hypothese 3 zeigt sich, dass der positive Effekt der *SPD* auf den Anteil dezentraler Unterbringung schwächer wird, wenn der Anteil von Parteianhängern mit asylkritischer Haltung ansteigt. Dieser Effekt ist für einen asylkritischen Anteil von Parteianhängern zwischen 0% und rund 70% statistisch signifikant.

**Abbildung 3:** Marginale Partieneffekte konditioniert durch asylkritische Haltungen der klassischen Wählerklientel



*Anmerkung:* Basis der Modellierungen: Best-Fit-Modell mit SPD bzw. Best-Fit-Modell mit Union. Gestrichelte Linien: 95%-Konfidenzintervalle.

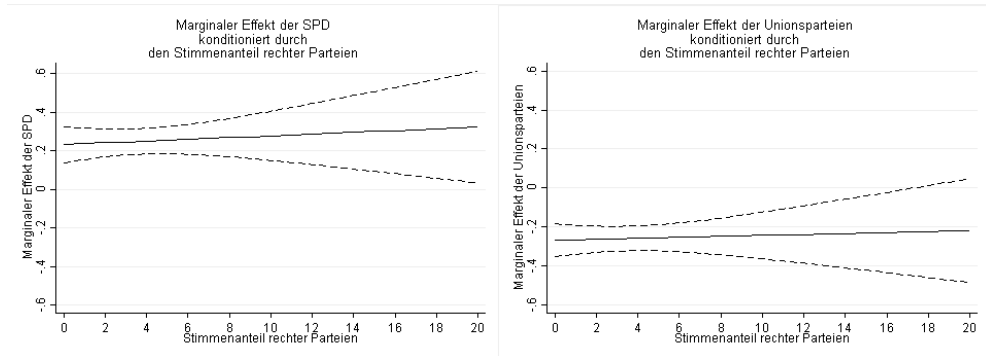
*Quelle:* Eigene Darstellung.

Beide „*Volksparteien*“ orientieren sich nach den Ergebnissen der vorliegenden Untersuchung also an asylkritischen Einstellungen ihrer Kernwählerschaften – obwohl davon auszugehen ist, dass die Kernwählerschaften von *SPD* und *Unionsparteien* deutlich heterogener in ihrer Präferenzlage sind als diejenigen „kleiner Parteien“, was ihre Berücksichtigung eigentlich erschweren sollte, und selbst dann, wenn nicht etwa die Mehrheit der Kernwählerschaft asylkritische Einstellungen vertritt, sondern bereits kleine Teile. Dieser Befund könnte darauf hindeuten, dass beide Parteien auf asylpolitische Präferenzen ihrer Kernwählerschaft besonders responsiv reagieren und negative Stimmungsverlagerungen rasch aufnehmen.

In einem nächsten Schritt werden die Befunde zum Parteienverhalten aufgrund von Parteienwettbewerb dargestellt. In den Hypothesen 7 und 8 wurde vermutet, dass sowohl *christdemokratische* als auch *linke Parteien*, die über eine asylkritisch eingestellte Wählerschaft verfügen, angesichts zunehmender Wahlerfolge rechter Parteien ihre asylpolitischen Positionen justieren und restriktivere Asylpolitiken verfolgen. Zur vereinfachten Interpretation der Ergebnisse dienen wieder *Marginal Effects Plots* (Abbildung 4). Erneut zeigen sich Befunde, die überraschen: Die *SPD* übt einen durchweg statistisch signifikanten, positiven Effekt auf den Anteil dezentraler Unterbringung aus, der sich mit steigendem Stimmenanteil rechter Parteien sogar noch verstärkt. Mit einem zunehmenden Stimmenanteil rechter Parteien wird der restriktive Effekt der *Unionsparteien* nicht, wie

in Hypothese 7 vermutet, stärker, sondern schwächer. Dieser Effekt ist nahezu für den gesamten empirisch vorkommenden Wertebereich statistisch signifikant.

**Abbildung 4:** Marginale Partieneffekte konditioniert durch den Stimmenanteil rechter Parteien

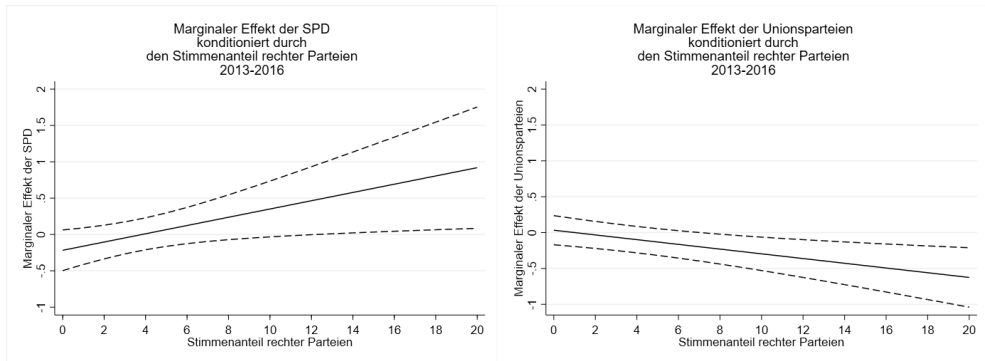


*Anmerkung:* Basis der Modellierungen: Best-Fit-Modell mit SPD bzw. Best-Fit-Modell mit Union. Gestrichelte Linien: 95%-Konfidenzintervalle.

*Quelle:* Eigene Darstellung.

Inwieweit lassen sich diese Befunde theoretisch erklären? Der überraschende Befund zur *SPD* kann möglicherweise mithilfe der Argumentation von Tim Bale, Christoffer Green-Pedersen, André Krouwel, Kurt Richard Luther und Nick Sitter (2010) eingeordnet werden: Diese machen in ihrer Untersuchung verschiedene Reaktionsstrategien linker Parteien auf das Aufkommen von rechtspopulistischen Parteien aus und stellen unter anderem dar, dass linke Parteien bisweilen mit einer stärkeren Betonung der eigenen Werte und nicht etwa mit Anpassung reagieren. Erklärungspotenzial für die widersprüchlichen Befunde zur *Union* ergibt sich aus der Überlegung, dass diese womöglich nicht sämtliche rechtsextremen und rechtspopulistischen Parteien als relevante Wettbewerber um Wählerstimmen ansieht. Wählerwanderungen verschiedener Landtagswahlen im Untersuchungszeitraum zeigen zwar, dass die Unionsparteien bisweilen Wähler an NPD, DVU oder auch die Republikaner verloren haben. Umgekehrte Wählerwanderungen in nachfolgenden Jahren und stetig zurückgehende Mitgliedszahlen verdeutlichen aber, dass das Wählerpotenzial von DVU, NPD und Republikanern flüchtig ist (Neu, 2004, S. 4). Wählerwanderungen hin zur rechtspopulistischen AfD erfolgten bei sämtlichen Wahlen seit 2013 in deutlich umfangreicherem Maße. Allen voran verfolgt die AfD als rechtspopulistische Partei andere ideologische und kommunikative Strategien als rechtsextreme Parteien, die sie potenziell für deutlich mehr Wähler anschlussfähig und wählbar machen (Decker & Lewandowsky, 2017). Vermutlich führt dies dazu, dass christdemokratische Parteien sie als wesentlich gewichtigere elektorale Gefahr einschätzen. Eine testweise Berechnung der Modellierungen für die Jahre ab 2013, dem Gründungsjahr der AfD,<sup>12</sup> spricht für diese Überlegung: Der restriktive Effekt der *Unionsparteien* wird mit zunehmendem Wahlerfolg rechter Parteien nicht mehr schwächer, sondern vielmehr statistisch signifikant deutlich stärker (bei Werten zwischen 7% bis 20%). Der statistisch signifikante, stärker positiv werdende Effekt der SPD bleibt bei einem Stimmenanteil rechter Parteien zwischen 10% und 20% bestehen.

**Abbildung 5:** Marginale Partieneffekte konditioniert durch den Stimmenanteil rechter Parteien 2013-2016



*Anmerkung:* Basis der Modellierungen: Best-Fit-Modell mit SPD bzw. Best-Fit-Modell mit Union. Gestrichelte Linien: 95%-Konfidenzintervalle.

*Quelle:* Eigene Darstellung.

## 6 Fazit und Ausblick

Reagieren Landesregierungen bei der Art der Unterbringung von Asylbewerbern vorrangig auf regionale Problemlagen und versuchen diese entsprechend der vorhandenen Ressourcen zu bewältigen? Oder gestalten sie Unterbringungspolitiken gemäß eigenen politischen Zielvorstellungen oder wahlstrategischen Erwägungen permissiv oder restriktiv aus? Diesen Fragestellungen ist das vorliegende Papier anhand der Einflussfaktoren auf die Art der Unterbringung von Asylbewerbern zwischen 1995 und 2016 nachgegangen. Die Ergebnisse der *gepoolten* Zeitreihenanalyse ergaben insgesamt, dass Landesregierungen mit ihrer Unterbringungspolitik nicht nur auf *sozioökonomische Problemlagen* reagieren, wenngleich diese zur Gesamterklärung der Varianz in der Unterbringung beitragen. Vielmehr zeigte sich, dass insbesondere *SPD* und *Union* in der Lage sind, die Art der Unterbringung in den Bundesländern gemäß ihren pluralistisch-inklusiven bzw. konservativ-exklusiven Wertvorstellungen zu gestalten, Politiken aber unter Berücksichtigung asylkritischer, exklusiver Einstellungen ihrer klassischen Wählerschaft anzupassen. Des Weiteren wurde ersichtlich, dass sowohl *SPD* als auch *Unionsparteien* ihre Unterbringungspolitiken im Wettbewerb mit rechten Parteien justieren: Die *Unionsparteien* gestalten diese deutlich restriktiver – allerdings lediglich im Wettbewerb mit der AfD nach deren Parteigründung 2013. Die *SPD* reagiert, entgegen der theoretischen Vermutung, mit einer „permissiven Gegenstrategie“. Effekte der „kleinen Parteien“ erwiesen sich als deutlich modellierungsabhängiger und weniger robust.

Anknüpfungspotenzial für weitere Forschung zur Unterbringungs- und Asylpolitik der Bundesländer ergibt sich vor allem hinsichtlich der Relevanz der kommunalen Ebene für deren tatsächliche Ausgestaltung. Das Zusammenspiel von Kommunen und Bundesländern stellt bislang ein wenig beleuchtetes Feld dar. Für ein besseres Verständnis asylpolitischer Prozesse im Mehrebenensystem könnten daher beispielsweise Nacherhebungen der Finanzierungsart und der Gliederung der Unterbringung im Zeit-

verlauf für quantitative Untersuchungen kommunaler Einflussmöglichkeiten, aber auch Fallstudien zu einzelnen Bundesländern und ausgewählten Kommunen, hilfreich sein.

Das vorliegende Papier leistet zusammengefasst betrachtet durch seine Analyse folgenden Beitrag: Es kann zeigen, dass unterschiedliche Regierungsparteien, und hier allen voran die *SPD* und die *Unionsparteien*, einen unterschiedlichen Effekt auf die Ausgestaltung der Unterbringung von Asylbewerbern ausüben und dass sie hierbei sowohl von ideologischen Motiven geleitet werden, aber auch die Präferenzen ihrer Kernwählerschaft berücksichtigen und Positionen im Wettbewerb mit rechten Parteien justieren. Durch diese Ergebnisse wird ersichtlich, dass die Unterbringung von Asylbewerbern nicht allein durch sozioökonomische Faktoren bestimmt wird, sondern durch Regierungen gestaltet werden kann.

## Anmerkungen

- 1 Im juristischen Sinne sind Flüchtlinge nur diejenigen Personen, die eine Asylanerkennung nach Art. 16a GG, Flüchtlingsschutz als Konventionsflüchtlinge (§3 Abs. 1 AsylG) oder subsidiären Schutz (Art. 15 der EU-Qualifikationsrichtlinie) erhalten haben. Menschen, die nach Deutschland kommen, um Asyl zu beantragen, sind „Asylsuchende“, sobald sie beim Bundesamt für Migration und Flüchtlinge (BAMF) Asyl beantragt haben, „Asylbewerber“ (Bundesregierung, 2018b). Menschen, die sich nicht (mehr) im Asylverfahren befinden bzw. einen negativen Bescheid erhalten haben, bei denen aber die Abschiebung ausgesetzt wurde, sind „geduldet“ (Bundesamt für Migration und Flüchtlinge, 2018). In der vorliegenden Arbeit soll aus Gründen der Lesbarkeit von „Asylbewerbern“ gesprochen werden. Damit sind Personen gemeint, die Asyl in der Bundesrepublik ersuchen oder ersucht haben, aber (noch) keinen positiven Bescheid erhalten haben (vgl. hierzu auch die Operationalisierung der abhängigen Variablen des Papiers).
- 2 Aus Gründen der Lesbarkeit wird auf die gleichzeitige Verwendung weiblicher und männlicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für beide Geschlechter.
- 3 Bislang ist umstritten, ob die Unterbringung in dezentralen Unterkünften oder in Gemeinschaftsunterkünften kostengünstiger ist (Aumüller, Daphi & Biesenkamp, 2015, S. 66). Allerdings wird die dezentrale Unterbringung regelmäßig als generöser wahrgenommen, sodass der Zusammenhang hier, dem Vorgehen in der Literatur folgend, überprüft werden soll.
- 4 Parteienwettbewerbliche Erwägungen werden auch durch die Salienz eines Themas geprägt. Hochsaliante Themen sollten von Parteien besonders berücksichtigt werden, da sie wahlentscheidend sein können. Der Einfluss asylpolitischer Themensalienz auf Parteienverhalten in den Bundesländern kann aufgrund einer mangelhaften Datengrundlage im vorliegenden Papier aber nicht überprüft werden.
- 5 Nachfolgend aus Gründen der Lesbarkeit als „rechte Parteien“ bezeichnet.
- 6 Sämtliche Variablen, ihre Operationalisierung, die Datenverfügbarkeit und die zugehörige Datenquelle finden sich in einer tabellarischen Übersicht im Anhang.
- 7 Zusätzlich wäre es angebracht, kommunale Einflüsse als Kontrollvariablen einzubeziehen. Daten, beispielsweise zur Gliedrigkeit oder Finanzierung des Unterbringungssystems, liegen bislang allerdings nur als Querschnitt für das Jahr 2014 vor (Wendel, 2014).
- 8 Die Verortung der AfD als rechtspopulistisch ist in der Literatur umstritten. Das vorliegende Papier folgt Lewandowsky, Giebler & Wagner, 2016.
- 9 Die dargestellte Annahmeüberprüfung wurde gleichsam für sämtliche Interaktionsmodelle (Hypothesen 3, 5, 7 und 8) durchgeführt. Es ergaben sich keine weiteren Annahme-Verletzungen.
- 10 Ein anderes Bild ergibt sich für *FDP* und *Die Linke*. *FDP*-Minister führten beispielsweise in Nordrhein-Westfalen und Bremen das Innenministerium, in Mecklenburg-Vorpommern, Sachsen-Anhalt oder Schleswig-Holstein das Sozialministerium. *Linken*-Minister führten in Berlin, Bremen, Brandenburg, Thüringen und Mecklenburg-Vorpommern Sozialministerien.
- 11 Konditionierte Effekte für die *FDP* werden entsprechend den Hypothesen nicht überprüft.
- 12 Von einem Test für die *AfD* lässt sich sprechen, weil die anderen rechten Parteien seit ihrem Aufkommen in den meisten Bundesländern nahezu vollständig marginalisiert sind.

## Literatur

- Abedi, Amir (2017). We Are Not in Bonn Anymore: The Impact of German Unification on Party Systems at the Federal and Land Levels. *German Politics*, 26 (4), 457-479.  
<https://doi.org/10.1080/09644008.2017.1365135>.
- Akkerman, Tjitske (2012). Comparing Radical Right Parties in Government: Immigration and Integration Policies in Nine Countries (1996–2010). *West European Politics*, 35 (3), 511-529.  
<https://doi.org/10.1080/01402382.2012.665738>.
- Andersen, Uwe & Woyke, Wichard (2013). Rechtsextreme Parteien. In Uwe Andersen & Wichard Woyke (Hrsg.). *Handwörterbuch des politischen Systems der Bundesrepublik Deutschland*. Verfügbar unter: <http://www.bpb.de/nachschlagen/lexika/handwoerterbuch-politisches-system/202097/rechtsextreme-parteien> [20. September 2018].
- Aumüller, Jutta (2018). Die kommunale Integration von Flüchtlingen. In Frank Gesemann & Roland Roth (Hrsg.). *Handbuch Lokale Integrationspolitik* (S. 173-198). Wiesbaden: Springer VS.
- Aumüller, Jutta & Bretl, Carolin (2008). *Die kommunale Integration von Flüchtlingen in Deutschland*. Verfügbar unter: <https://www.desi-sozialforschung-berlin.de/migration-und-integration/> [01. Januar 2020].
- Aumüller, Jutta, Daphi, Priska & Biesenkamp, Celine (2015). *Die Aufnahme von Flüchtlingen in den Bundesländern und Kommunen. Behördliche Praxis und zivilgesellschaftliches Engagement*. Verfügbar unter: <https://www.bosch-stiftung.de/de/publikation/die-aufnahme-von-fluechtlingen-den-bundeslaendern-und-kommunen> [20. Oktober 2018].
- Bale, Tim (2008). Turning Round the Telescope: Centre-right Parties and Immigration and Integration Policy in Europe. *Journal of European Public Policy*, 15 (3), 315-330.  
<https://doi.org/10.1080/13501760701847341>.
- Bale, Tim, Green-Pedersen, Christoffer, Krouwel, André, Luther, Kurt Richard & Sitter, Nick (2010). If You Can't Beat them, Join them? Explaining Social Democratic Responses to the Challenge from the Populist Radical Right in Western Europe. *Political Studies*, 58 (3), 410-426.  
<https://doi.org/10.1111/j.1467-9248.2009.00783.x>.
- Bergmann, Jan & Dienelt, Klaus (2018). *Ausländerrecht. Kommentar* (12. Aufl.). München: Beck-Online.
- Bogumil, Jörg, Hafner, Jonas & Kuhlmann, Sabine (2016). Verwaltungshandeln in der Flüchtlingskrise: Die Erstaufnahmeeinrichtungen der Länder und die Zukunft des Verwaltungsvollzugssystems Asyl. *Verwaltung und Management*, 22 (3), 126-136.  
<https://doi.org/10.5771/0947-9856-2016-3-126>.
- Bolin, Niklas, Lidén, Gustav & Nyhlén, Jon (2014). Do Anti-immigration Parties Matter? The Case of the Sweden Democrats and Local Refugee Policy. *Scandinavian Political Studies*, 37 (3), 323-343. <https://doi.org/10.1111/1467-9477.12031>.
- Boswell, Christina (2000). European Values and the Asylum Crisis. *International Affairs*, 76 (3), 537-557. <https://doi.org/10.1111/1468-2346.00150>.
- Boswell, Christina & Hough, Dan (2008). Politicizing Migration: Opportunity or Liability for The Centre-Right in Germany? *Journal of European Public Policy*, 15 (3), 331-348.  
<https://doi.org/10.1080/13501760701847382>.
- Breunig, Christian & Luedtke, Adam (2008). What Motivates the Gatekeepers? Explaining Governing Party Preferences on Immigration. *Governance: An International Journal of Policy, Administration and Institutions*, 21 (1), 123-146. <https://doi.org/10.1111/j.1468-0491.2007.00388.x>.
- Bünthe, Rudolf (2014). Asyl und Arbeitsmarkt – zur sozialen Lage von Asylbewerbern. In Stefan Luft & Peter Schimany (Hrsg.). *20 Jahre Asylkompromiss. Bilanz und Perspektiven* (S. 219-237). Bielefeld: Transcript Verlag.
- Bundesamt für Migration und Flüchtlinge (2018). *FAQ: Zugang zum Arbeitsmarkt für geflüchtete Menschen*. Verfügbar unter: <http://www.bamf.de/DE/Infothek/FragenAntworten/ZugangArbeitFluechtlinge/zugang-arbeit-fluechtlinge-node.html> [20. Oktober 2018].
- Bundesamt für Statistik (2018a). Arbeitslose, Arbeitslosenquoten, Gemeldete Arbeitsstellen: Bundesländer, Jahre. Arbeitsmarktstatistik der Bundesagentur für Arbeit (Code 13211-0007).

- Bundesamt für Statistik (2018b). Ausländer: Bundesländer, Stichtag, Geschlecht/Altersjahre/Familienstand (Code 12521-0020).
- Bundesamt für Statistik (2018c). Bevölkerung, Erwerbstätige, Erwerbslose, Erwerbspersonen, Nichterwerbspersonen: Bundesländer, Jahre (Code 12211-0005).
- Bundesamt für Statistik (2018d). Bevölkerungsdichte nach Bundesländern 1980 bis 1994. Ergebnisse der Bevölkerungsfortschreibung. Zusammenstellung des Bundesamtes auf Anfrage.
- Bundesamt für Statistik (2018e). Bevölkerungsdichte nach Bundesländern. Auszüge aus dem Gemeindeverzeichnis 2015 und 2016. Zusammenstellung des Bundesamtes auf Anfrage.
- Bundesamt für Statistik (2018f). *Bevölkerungsdichte: Bundesländer, Stichtag* (Code 12411-0050).
- Bundesamt für Statistik (2018g). Empfänger von Asylbewerberregelleistungen nach dem Asylbewerberleistungsgesetz: Bundesländer, Stichtag, Geschlecht (Code 22221-0010).
- Bundesamt für Statistik (2018h). Länderübersicht der Leistungsempfängerinnen/-empfänger nach Grundleistungen (§3) bzw. von Hilfe zum Lebensunterhalt (§2) nach dem AsylbLG: Bundesländer, Stichtag, Art der Unterbringung. Zusammenstellung des Bundesamtes auf Anfrage.
- Bundesamt für Statistik (2018i). *Statistisches Jahrbuch – Ältere Ausgaben*. Verfügbar unter: [https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch\\_AeltereAusgaben.html;jsessionid=B428C6C2F3D0756E926058D5ABC6E55B.InternetLive2](https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch_AeltereAusgaben.html;jsessionid=B428C6C2F3D0756E926058D5ABC6E55B.InternetLive2) [21. Oktober 2018].
- Bundesamt für Statistik (2018j). *Zusammengefasste Geburtenziffer nach Bundesländern ab 1990*. Zusammenstellung des Bundesamtes auf Anfrage.
- Bundesamt für Statistik (2018k). Zusammengefasste Geburtenziffern (je Frau): Bundesländer, Jahre, Altersgruppe 15-50 (Code 12612-0104).
- Bundesregierung (2018a). *Integrationskosten: Bund entlastet Länder und Kommunen*. Verfügbar unter: <https://www.bundesregierung.de/breg-de/aktuelles/bund-entlastet-laender-und-kommunen-1535922> [07. April 2019].
- Bundesregierung (2018b). *Was ist der Unterschied zwischen einem Asylbewerber und einem Flüchtling?* Verfügbar unter: <https://www.deutschland-kann-das.de/dekd/politik/aktuelles/was-ist-der-unterschied-zwischen-einem-asylbewerber-und-einem-fluechtling--511666> [20. Oktober 2018].
- Bundeswahlleiter (2018). *Ergebnisse früherer Landtagswahlen*. Verfügbar unter: [https://www.bundeswahlleiter.de/dam/jcr/a333e523-0717-42ad-a772-d5ad7e7e97cc/ltw\\_erg\\_gesamt.pdf](https://www.bundeswahlleiter.de/dam/jcr/a333e523-0717-42ad-a772-d5ad7e7e97cc/ltw_erg_gesamt.pdf) [21. Oktober 2018].
- Busch, Christine (2015). Versorgung von Flüchtlingen in der Stadt Bergkamen. *Städte- und Gemeinderat*, 69 (1), 8-10.
- Ceobanu, Alin & Escandell, Xavier (2010). Comparative Analyses of Public Attitudes Toward Immigrants and Immigration Using Multinational Survey Data: A Review of Theories and Research. *Annual Review of Sociology*, 36, 309-328. <https://doi.org/10.1146/annurev.soc.012809.102651>.
- Clark, John A. & Legge, Jerome S. (1997). Economics, Racism, and Attitudes Toward Immigration in the New Germany. *Political Research Quarterly*, 50 (4), 901-917. <https://doi.org/10.1177%2F106591299705000409>.
- Cremer, Hendrik (2014). *Menschenrechtliche Verpflichtungen bei der Unterbringung von Flüchtlingen. Empfehlungen an die Länder, Kommunen und den Bund*. Verfügbar unter: [https://www.institut-fuer-menschenrechte.de/uploads/tx\\_commerce/Policy\\_Paper\\_26\\_Menschenrechtliche\\_Verpflichtungen\\_bei\\_der\\_Unterbringung\\_von\\_Fluechtlingen\\_01.pdf](https://www.institut-fuer-menschenrechte.de/uploads/tx_commerce/Policy_Paper_26_Menschenrechtliche_Verpflichtungen_bei_der_Unterbringung_von_Fluechtlingen_01.pdf) [20. November 2018].
- Decker, Frank & Lewandowsky, Marcel (2017). Rechtspopulismus: Erscheinungsformen, Ursachen und Gegenstrategien. In Bundeszentrale für Politische Bildung (Hrsg.). *Dossier Rechtspopulismus*. Verfügbar unter: <http://www.bpb.de/politik/extremismus/rechtspopulismus/240089/rechtspopulismus-erscheinungsformen-ursachen-und-gegenstrategien> [21. Oktober 2018].
- Dirke, Sabine von (1994). Multikulti: The German Debate on Multiculturalism. *German Studies Review*, 17 (3), 513-536. <https://doi.org/10.2307/1431898>.

- Flüchtlingsrat Bremen (2015). *Rückblick: Flüchtlinge setzen Rechte durch*. Verfügbar unter: <https://www.fluechtlingsrat-bremen.de/wohnen/kampf-um-wohnungen-fluechtlinge-setzen-rechte-durch/> [20. Oktober 2018].
- Forschungsgruppe Wahlen (2017). *Politbarometer 1977-2016 (Gesamtkumulation)*. ZA5100 Datenfile Version 1.0.0. Köln: GESIS Datenarchiv. (Variablen Vv34 & Vv35 nach Vx53).
- Franzmann, Simon (2016). Calling the Ghost of Populism: The AfD's Strategic and Tactical Agendas until the EP Election 2014. *German Politics*, 25 (4), 457-479. <https://doi.org/10.1080/09644008.2016.1201075>.
- Green, Simon (2001). Immigration, Asylum and Citizenship in Germany: The Impact of Unification and the Berlin Republic. *West European Politics*, 24 (4), 82-104. <https://doi.org/10.1080/01402380108425466>.
- Han, Kyung Joon (2013). Political Use of Asylum Policies: The effects of partisanship and election timing on government policies regarding asylum seeker's welfare benefits. *Comparative European Politics*, 11 (4), 383-405. <https://doi.org/10.1057/cep.2012.24>.
- Han, Kyung Joon (2015). When Will Left-Wing Governments Introduce Liberal Migration Policies? An Implication of Power Resources Theory. *International Studies Quarterly*, 59 (3), 602-614. <https://doi.org/10.1111/isqu.12156>.
- Heckmann, Friedrich (2015). *Integration von Migranten. Einwanderung und neue Nationenbildung*. Wiesbaden: VS Verlag.
- Henkes, Christian (2008). Integrationspolitik in den Bundesländern? In Achim Hildebrandt & Frieder Wolf (Hrsg.). *Die Politik der Bundesländer. Staatstätigkeit im Vergleich* (S. 113-135). Wiesbaden: VS Verlag.
- Hinnfors, Jonas, Spehar, Andrea & Bucken-Knapp, Gregg (2012). The Missing Factor: Why Social Democracy Can Lead to Restrictive Immigration Policy. *Journal of European Public Policy*, 19 (4), 585-603. <https://doi.org/10.1080/13501763.2011.607348>.
- Hörisch, Felix (2018). Asylpolitik im Bundesländervergleich: Eine fuzzy-set Qualitative Comparative Analysis. *Zeitschrift für Vergleichende Politikwissenschaft*, 12 (4), 783-803. <https://doi.org/10.1007/s12286-018-0399-4>.
- Hörisch, Felix & Heiken, Matthias (2018). Qualitative Comparative Analysis: Mit einer Anwendung zur Asylbewerberunterbringung im Bundesländervergleich. In Markus Tausendpfund (Hrsg.). *Fortgeschrittene Forschungsmethoden in den Sozialwissenschaften*. FernUniversität Hagen [i.E.].
- Inforadio rbb (2015). *Überforderte Städte – Kein Platz für Flüchtlinge?* Verfügbar unter: <https://www.inforadio.de/dossier/2016/grossprojekte/fluechtlings-dossiers/fluechtlingsunterbringung/ueberforderte-staedte--kein-platz-fuer-fluechtlinge-.html> [20. Oktober 2018].
- Johansson, Susanne (2016). Was wir über Flüchtlinge (nicht) wissen. Der wissenschaftliche Erkenntnisstand zur Lebenssituation von Flüchtlingen. Eine Expertise im Auftrag der Robert Bosch Stiftung und des SVR Forschungsbereichs. Verfügbar unter: <https://www.fh-dortmund.de/de/hs/medien/Was-wir-ueber-Fluechtlinge-nicht-wissen.pdf> [01. Januar 2020].
- Joppke, Christian (1997). Asylum and State Sovereignty. *Comparative Political Studies*, 30 (3), 259-298. <https://doi.org/10.1177%2F0010414097030003001>.
- Kersbergen, Kees van & Krouwel, André (2008). A Double-Edged Sword! The Dutch Centre-Right and the 'Foreigners Issue'. *Journal of European Public Policy*, 15 (3), 398-414. <https://doi.org/10.1080/13501760701847655>.
- Keskinen, Suvi (2016). From Welfare Nationalism to Welfare Chauvinism: Economic Rhetoric, the Welfare State and Changing Asylum Policies in Finland. *Critical Social Policy*, 36 (3), 352-370. <https://doi.org/10.1177%2F0261018315624170>.
- Kittel, Bernhard (2005). Pooled Analysis in der ländervergleichenden Forschung: Probleme und Potenziale. In Sabine Kropp & Michael Minkenberg (Hrsg.). *Vergleichen in der Politikwissenschaft* (S. 96-115). Wiesbaden: VS Verlag.



- Koning, Edward Anthony (2017). Selecting, Disentitling, or Investing? Exploring Party and Voter Responses to Immigrant Welfare Dependence in 15 West European Welfare States. *Comparative European Politics*, 15 (4), 628-660. <https://doi.org/10.1057/s41295-017-0097-7>.
- Kostner, Sandra (2017). Asyl und Integration: Aufbrüche in stürmische Zeiten. In Felix Hörisch & Stefan Wurster (Hrsg.). *Das grün-rote Experiment in Baden-Württemberg: Eine Bilanz der Landesregierung Kretschmann 2011-2016* (S. 185-222). Wiesbaden: VS Verlag.
- Lewandowsky, Marcel (2017). Was ist und wie wirkt Rechtspopulismus? *Bürger und Staat*, 67 (1), 4-11.
- Lewandowsky, Marcel, Giebler, Heiko & Wagner, Aiko (2016). Rechtspopulismus in Deutschland: Eine empirische Einordnung der Parteien zur Bundestagswahl 2013 unter besonderer Berücksichtigung der AfD. *Politische Vierteljahresschrift*, 57 (2), 247-275. <https://doi.org/10.5771/0032-3470-2016-2-247>.
- Minkenberg, Michael (2003). The Politics of Citizenship in the New Republic. *West European Politics*, 26 (4), 219-240. <https://doi.org/10.1080/01402380312331280758>.
- Münch, Ursula (2014). Asylpolitik in Deutschland – Akteure, Interessen, Strategien. In Stefan Luft & Peter Schimany (Hrsg.). *20 Jahre Asylkompromiss. Bilanz und Perspektiven* (S. 69-86). Bielefeld: Transcript Verlag.
- Münch, Ursula (2016). Integrationspolitik der Länder – dringliche Zukunftsaufgabe im Umbruch. In Achim Hildebrandt & Frieder Wolf (Hrsg.). *Die Politik der Bundesländer: Zwischen Föderalismusreform und Schuldenbremse* (S. 365-390). Wiesbaden: VS Verlag.
- Neu, Viola (2004). DVU – NPD: Perspektiven und Entwicklungen. *KAS-Arbeitspapiere* (140). Verfügbar unter: [https://www.kas.de/c/document\\_library/get\\_file?uuid=665b9e2f-a92c-01a1-6862-1394237fc33b&groupId=252038](https://www.kas.de/c/document_library/get_file?uuid=665b9e2f-a92c-01a1-6862-1394237fc33b&groupId=252038) [20. November 2018].
- Obinger, Herbert (2015). Funktionalismus. In Georg Wenzelburger & Reimut Zohnhöfer (Hrsg.). *Handbuch Policy-Forschung* (S. 35-54). Wiesbaden: VS Verlag.
- Otto, Alkis Henri & Steinhardt, Max Friedrich (2014). Immigration and election outcome: Evidence from city districts in Hamburg. *Regional Science and Urban Economics*, 45, 67-79. <https://doi.org/10.1016/j.regsciurbeco.2014.01.004>.
- Reutter, Werner (2006). The Transfer of Power Hypothesis and the German Länder: In Need of Modification. *Publius*, 36 (2), 277-301. <https://doi.org/10.1093/publius/pjj017>.
- Riedel, Lisa & Schneider, Gerald (2017). Dezentraler Asylvollzug diskriminiert: Anerkennungsquoten von Flüchtlingen im bundesdeutschen Vergleich, 2010-2015. *Politische Vierteljahresschrift*, 58 (1), 23-50. <https://dx.doi.org/10.5771/0032-3470-2017-1-23>.
- Rippl, Susanne & Seipel, Christian (2018). Modernisierungsverlierer, Cultural Backlash, Postdemokratie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 70 (2), 237-254. <https://doi.org/10.1007/s11577-018-0522-1>.
- Rotte, Ralph (2000). Immigration Control in United Germany: Toward a Broader Scope of National Policies. *The International Migration Review*, 34 (2), 357-389. <https://doi.org/10.1177/019791830003400201>.
- Schmidtke, Oliver (2017). Reinventing the Nation: Germany's Post-Unification Drive Towards Becoming a 'Country of Immigration'. *German Politics*, 26 (4), 498-515. <https://doi.org/10.1080/09644008.2017.1365137>.
- Schneider, Jan (2010). Modernes Regieren und Konsens: Kommissionen und Beratungsregime in der deutschen Migrationspolitik. Wiesbaden: VS Verlag.
- Spier, Tim (2010). Modernisierungsverlierer? Die Wählerschaft rechtspopulistischer Parteien in Westeuropa. Wiesbaden: VS Verlag.
- Squire, Vicki (2008). Accounting for the Dominance of Control: Inter-party dynamics and Restrictive Asylum Policy in Contemporary Britain. *British Politics*, 3 (2), 241-261. <https://doi.org/10.1057/bp.2008.4>.
- Thränhardt, Dietrich (1995). The Political Uses of Xenophobia in England, France and Germany. *Party Politics*, 1 (3), 323-345. <https://doi.org/10.1177/02F1354068895001003002>.
- Thränhardt, Dietrich (2002). Einwanderungs- und Integrationspolitik in Deutschland und den Niederlanden. *Leviathan*, 30 (2), 220-249. <https://doi.org/10.1007/s11578-002-0019-0>.

- Toshkov, Dimiter (2014). The Dynamic Relationship between Asylum Applications and Recognition Rates in Europe (1987-2010). *European Union Politics*, 15 (2), 192-214. <https://doi.org/10.1177%2F1465116513511710>.
- Träger, Hendrik & Neumann, Alexandra (2018). Integrations- und Flüchtlingspolitik. In Torsten Oppelland (Hrsg.). *Politik und Regieren in Thüringen* (S. 415-439). Wiesbaden: VS Verlag.
- Tutic, Andreas & Hermanni, Hagen von (2018). Sozioökonomischer Status, Deprivation und die Affinität zur AfD – Eine Forschungsnotiz. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 70 (2), 275-294. <https://doi.org/10.1007/s11577-018-0523-0>.
- Volkswirtschaftliche Gesamtrechnung der Länder (2018a). *Bruttoinlandsprodukt, Bruttowertschöpfung in den Ländern der Bundesrepublik Deutschland*. Verfügbar unter: <https://www.statistik-bw.de/VGRdL/tbIs/R0B0.jsp?tbl=R1B1> [21. Oktober 2018].
- Volkswirtschaftliche Gesamtrechnung der Länder (2018b). *Private Konsumausgaben in jeweiligen Preisen 1991 bis 2016*. Verfügbar unter: <https://www.statistik-bw.de/VGRdL/tbIs/> [21. Oktober 2018].
- Weber, Hannes (2015). Mehr Zuwanderer, mehr Fremdenangst? Ein Überblick über den Forschungsstand und ein Erklärungsversuch aktueller Entwicklungen in Deutschland. *Berliner Journal für Soziologie*, 25 (4), 397-428. <http://dx.doi.org/10.1007/s11609-016-0300-8>.
- Wendel, Kay (2014). *Unterbringung von Flüchtlingen in Deutschland: Regelungen und Praxis der Bundesländer im Vergleich*. Verfügbar unter: [https://www.proasyl.de/wp-content/uploads/2015/04/Laendervergleich\\_Unterbringung\\_2014-09-23\\_01.pdf](https://www.proasyl.de/wp-content/uploads/2015/04/Laendervergleich_Unterbringung_2014-09-23_01.pdf) [20. Oktober 2018].
- Wenzelburger, Georg, Jäckle, Sebastian & König, Pascal (2014). Weiterführende statistische Methoden für Politikwissenschaftler. Eine anwendungsbezogene Einführung in Stata. München: Oldenbourg Wissenschaftsverlag.
- Wissenschaftliche Dienste des Bundestages (2016). *Dokumentation. Sozialleistungen für Asylsuchende und Flüchtlinge in ausgewählten EU-Mitgliedstaaten*. Verfügbar unter: <https://www.bundestag.de/blob/437542/c0cefa93f919ae35ace1f89197bc41a7/wd-6-056-16-pdf-data.pdf> [20. Oktober 2018].
- Wolf, Frieder (2011). Die Einbindung zivilgesellschaftlicher Akteure in die Härtefallkommissionen der Bundesländer. Bandbreite, Ursachen, Folgen. *Zeitschrift für Politikwissenschaft*, 21 (2), 259–291. <https://doi.org/10.5771/1430-6387-2011-2-259>.
- Zohlnhöfer, Reimut (2013). Politische Parteien und öffentliche Politik: Zum Einfluss von Parteien auf die Staatstätigkeit. In Karl-Rudolf Korte & Timo Grunden (Hrsg.). *Handbuch Regierungsfor-schung* (S. 267-276). Wiesbaden: VS Verlag.

### *Anschrift der Autorin:*

Svenja Bauer-Blaschkowski, M.A., Technische Universität Darmstadt – Institut für Politikwissenschaft, Dolivostraße 15, 64293 Darmstadt,  
E-Mail: [bauer-blaschkowski@pg.tu-darmstadt.de](mailto:bauer-blaschkowski@pg.tu-darmstadt.de).

*Anhang 1: Variablen, Operationalisierung, Datenverfügbarkeit und Datenquellen*

| <b>Abhängige Variable</b>                                |                                                                                                                                                                       |                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dezentrale Unterbringung                                 | Anteil der in dezentralen Einrichtungen untergebrachten Leistungsempfängern nach §§2,3 AsylbLG an sämtlichen Leistungsempfängern in der Anschlussunterbringung (in %) | 1994-2016                                                  | <i>Eigene Berechnung basierend auf:</i><br>Bundesamt für Statistik (2018h): Länderübersicht der Leistungsempfängerinnen/-empfänger nach Grundleistungen (§3) bzw. von Hilfe zum Lebensunterhalt (§2) nach dem AsylbLG: Bundesländer, Stichtag, Art der Unterbringung, Zusammenstellung des Bundesamtes auf Anfrage.                                                                                                                                                                                             |
| <b>Unabhängige Variablen</b>                             |                                                                                                                                                                       |                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Kabinettsitzanteile Parteien                             | Kabinettsitzanteile von Bündnis 90/Die Grünen, SPD, Die Linke, FDP, CDU/CSU                                                                                           | 1994-2016                                                  | <i>Eigene Berechnung basierend auf:</i><br>Bundesamt für Statistik (2018i): Statistisches Jahrbuch – Ältere Ausgaben. Verfügbar unter: <a href="https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch_AeltereAusgaben.html;jsessionid=B428C6C2F3D0756E926058D5ABC6E55B.InternetLive2">https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch_AeltereAusgaben.html;jsessionid=B428C6C2F3D0756E926058D5ABC6E55B.InternetLive2</a> [21.Oktober 2018]. |
| Stimmenanteil rechtsextremer/-populistischer Parteien    | Aufaddierter Stimmenanteil von NPD, DVU, REP und AfD bei Landtagswahlen (in %)                                                                                        | 1994-2016                                                  | <i>Eigene Berechnung basierend auf:</i><br>Bundeswahlleiter (2018): Ergebnisse früherer Landtagswahlen. Verfügbar unter: <a href="https://www.bundeswahlleiter.de/dam/jcr/a333e523-0717-42ad-a772-d5ad7e7e97cc/ltw_erg_gesamt.pdf">https://www.bundeswahlleiter.de/dam/jcr/a333e523-0717-42ad-a772-d5ad7e7e97cc/ltw_erg_gesamt.pdf</a> [21.Oktober 2018].                                                                                                                                                       |
| Asylkritische Einstellungen der klassischen Wählerschaft | Anteil der Befragten mit Parteibindung mit negativer Einstellung zu Ausländern an allen Befragten mit jeweiliger Parteibindung in einem Bundesland (in %)             | 1995, 1998, 2000, 2001, 2002, 2004, 2006, 2007, 2010, 2014 | <i>Eigene Berechnung basierend auf:</i><br>Forschungsgruppe Wahlen (2017): Politbarometer 1977-2016 (Gesamtkumulation). ZA5100 Datenfile Version 1.0.0. Köln: GESIS Datenarchiv. ( <i>Variablen Vv34 &amp; Vv35 nach Vx53</i> )                                                                                                                                                                                                                                                                                 |
| Weibliche Leistungsempfänger                             | Anteil weiblicher Leistungsempfänger an allen Leistungsempfängern                                                                                                     | 1994-2016                                                  | <i>Eigene Berechnung basierend auf:</i><br>Bundesamt für Statistik (2018g): Empfänger von Asylbewerberregelleistungen nach dem Asylbewerberleistungsgesetz: Bundesländer, Stichtag, Geschlecht (Code 22221-0010).                                                                                                                                                                                                                                                                                               |
| Leistungsempfänger                                       | Veränderung des Anteils von Leistungsempfängern nach §§2,3 AsylbLG an der Bevölkerung eines Bundeslandes                                                              | 1994-2016                                                  | <i>Eigene Berechnung basierend auf:</i><br>Bundesamt für Statistik (2018g): Empfänger von Asylbewerberregelleistungen nach dem Asylbewerberleistungsgesetz: Bundesländer, Stichtag, Geschlecht (Code 22221-0010).<br>Bundesamt für Statistik (2018c): Bevölkerung, Erwerbstätige, Erwerbslose, Erwerbspersonen, Nichterwerbspersonen: Bundesländer, Jahre (Code 12211-0005).                                                                                                                                    |

| Unabhängige Variable |                                                                         |           |                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------|-------------------------------------------------------------------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bruttoinlandsprodukt | BIP preisbereinigt (pro Kopf)                                           | 1994-2016 | Volkswirtschaftliche Gesamtrechnung der Länder (2018a): Bruttoinlandsprodukt, Bruttowertschöpfung in den Ländern der Bundesrepublik Deutschland. Verfügbar unter: <a href="https://www.statistik-bw.de/VGRdL/tbls/R0B0.jsp?tbl=R1B1">https://www.statistik-bw.de/VGRdL/tbls/R0B0.jsp?tbl=R1B1</a> [21.Oktober 2018].                                                                                                                                   |
| Arbeitslosenquote    | Anteil Arbeitsloser an zivilen Erwerbspersonen (in %)                   | 1994-2016 | Bundesamt für Statistik (2018a): Arbeitslose, Arbeitslosenquoten, Gemeldete Arbeitsstellen: Bundesländer, Jahre. Arbeitsmarktstatistik der Bundesagentur für Arbeit (Code 13211-0007).                                                                                                                                                                                                                                                                 |
| Kaufkraft            | Private Konsumausgaben in jeweiligen Preisen (je Bundesland) (pro Kopf) | 1994-2016 | Volkswirtschaftliche Gesamtrechnung der Länder (2018b): Private Konsumausgaben in jeweiligen Preisen 1991 bis 2016. Verfügbar unter: <a href="https://www.statistik-bw.de/VGRdL/tbls/">https://www.statistik-bw.de/VGRdL/tbls/</a> [21.Oktober 2018].                                                                                                                                                                                                  |
| Stadtstaat           | Dummy-Variable Stadtstaat=1                                             | 1994-2016 | <i>Eigene Erhebung</i>                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Bevölkerungsdichte   | Einwohner je qkm                                                        | 1994-2016 | Bundesamt für Statistik (2018f): Bevölkerungsdichte: Bundesländer, Stichtag (Code 12411-0050). Bundesamt für Statistik (2018d): Bevölkerungsdichte nach Bundesländern 1980 bis 1994. Ergebnisse der Bevölkerungsfortschreibung. Zusammenstellung des Bundesamtes auf Anfrage. Bundesamt für Statistik (2018e): Bevölkerungsdichte nach Bundesländern. Auszüge aus dem Gemeindeverzeichnis 2015 und 2016. Zusammenstellung des Bundesamtes auf Anfrage. |
| Geburtenrate         | Geburtenziffer je Frau (15-50 Jahre)                                    | 1994-2016 | Bundesamt für Statistik (2018j): Zusammengefasste Geburtenziffer nach Bundesländern ab 1990. Zusammenstellung des Bundesamtes auf Anfrage. Bundesamt für Statistik (2018k): Zusammengefasste Geburtenziffern (je Frau): Bundesländer, Jahre, Altersgruppe 15-50 (Code 12612-0104).                                                                                                                                                                     |
| Ausländeranteil      | Anteil von Ausländern an der Bevölkerung eines Bundeslandes             | 1994-2016 | <i>Eigene Berechnung basierend auf:</i><br>Bundesamt für Statistik (2018b): Ausländer: Bundesländer, Stichtag, Geschlecht/Altersjahre/Familienstand (Code 12521-0020).<br>Bundesamt für Statistik (2018c): Bevölkerung, Erwerbstätige, Erwerbslose, Erwerbspersonen, Nichterwerbspersonen: Bundesländer, Jahre (Code 12211-0005).                                                                                                                      |
| Neue Bundesländer    | Dummy<br>1=Ost                                                          | 1994-2016 | <i>Eigene Erhebung</i>                                                                                                                                                                                                                                                                                                                                                                                                                                 |

Quelle: Eigene Darstellung.

Christian Person, René Geißler

## Ein Fass ohne Boden? Vier Jahrzehnte kommunaler Schuldenhilfen in Deutschland

### Zusammenfassung

Trotz Kommunalaufsicht und Finanzausgleich befinden sich viele Kommunen seit Jahrzehnten in einer Haushaltskrise. Um diese Haushaltsprobleme zu adressieren, stellten die Bundesländer wiederholt Schuldenhilfen bereit. Der Aufsatz beschreibt die Entwicklung dieser Schuldenhilfen, beginnend mit den ersten Entschuldungsmaßnahmen in den 1980ern bis hin zur flächendeckenden Verbreitung kommunaler Entschuldungsprogramme nach der globalen Finanzkrise. Anschließend werden unter Zuhilfenahme gängiger Analysekatoren der Policy-Analyse und der Neuen Institutionenökonomik Erklärungen für die unterschiedliche Ausgestaltung, Implementationszeitpunkte sowie Auswirkungen dieser Programme präsentiert. Im Hinblick auf ihre Ausgestaltungsformen werden Schuldenhilfen nach Zielen, Instrumenten und Reichweite kategorisiert. Die unterschiedliche Ausgestaltung und Implementationszeitpunkte von Entschuldungsprogrammen erklären sich primär durch variierenden sozioökonomischen Problemdruck, jeweilige politische Leitbilder und divergierende fiskalische Handlungsspielräume der Landesregierungen. In Bezug auf die Auswirkungen der Entschuldungsprogramme sind Erfolge bei Haushaltskonsolidierung und Schuldenabbau erkennbar, allerdings auf Kosten sinkender Investitionsausgaben und steigender Steuerbelastungen. Die eigentlichen Risiken liegen jedoch in der massiven Einschränkung kommunaler Selbstverwaltung.

**Schlagworte:** Kommunalfinanzen, Kommunalverschuldung, Haushaltskrise, Kassenkredite, Schuldenhilfen

### Abstract

*A bottomless pit? Forty years of municipal bailouts in Germany*

Despite state supervision and fiscal equalization many German municipalities have suffered from budgetary crisis for decades. To address these problems, most state governments implemented municipal bailouts. Therefore, we describe the historical development of these bailouts, ranging from the first program in the 1980s to the broad dissemination of bailout programs after the global financial crisis. Subsequently, we provide explanations on differences regarding their designs, time of implementation and impacts by using common analytical categories from policy analysis theory. With regard to design, we systematize bailouts according to their goals, instruments and scope. We explain different design and timing by socioeconomic problem pressure, political guiding principles and fiscal capacities of state governments. Concerning their impacts, one can observe improvements with regard to budget consolidation and debt reduction, but at the same time shrinking capital expenditure and higher taxation as well as massive interference with local self-government.

**Keywords:** Local public finances, local debt, budgetary crisis, cash credits, bailout grants

## 1 Einleitung

Das aktuelle Gesamtbild der Kommunalfinanzen in Deutschland erscheint auf den ersten Blick positiv. Seit 2015 weist die kommunale Ebene regelmäßig Finanzierungsüberschüsse auf. Die aggregierten Kennzahlen verbergen jedoch, dass viele Städte aufgrund von Strukturwandel, unzureichender Finanzausstattung sowie steigender Sozialausgaben seit Jahrzehnten mit Haushaltsproblemen kämpfen (Schwarting, 2019; Geißler, 2019). Die enormen finanziellen Einschränkungen, denen diese Kommunen unterliegen, engen ihre Gestaltungsspielräume stark ein. Sie bedrohen nicht nur die Bereitstellung öffentlicher Dienstleistungen, sondern gefährden auch die generelle Funktions- und Zukunftsfähigkeit der Kommunen an sich.

Die Existenz kommunaler Haushaltskrisen in Deutschland vermag zu überraschen, praktizieren doch alle Bundesländer einen kommunalen Finanzausgleich (KFA). Dieser soll den Kommunen eine bedarfsgerechte Finanzausstattung garantieren. Zusätzlich unterliegen Kommunalhaushalte strikten Regulierungen, die das Abgleiten in Haushaltskrisen verhindern sollen. Deren Einhaltung wird durch staatliche Aufsichtsbehörden überwacht (Ebinger, Zabler, Person, Niemann & Geißler, 2018; Geißler, 2019). Trotz dieser Rahmenbedingungen befinden sich viele Kommunen in finanziellen Schwierigkeiten. Schuldenhilfen der Länder sind eine mögliche Reaktion auf diese Problemlagen. Sie dienen dazu, Kommunen bei Haushaltskonsolidierung und Schuldenabbau zu unterstützen. Schuldenhilfen sind seit Jahrzehnten Realität in Deutschland. Aktuell stehen sie auch erstmals auf der Agenda der Bundesregierung.<sup>1</sup>

Schuldenhilfen an subnationale Gebietskörperschaften sind auch international ein bekanntes Phänomen. Folglich existiert eine umfassende Literatur, die sich im Rahmen der Neuen Institutionenökonomik und der Theorie Öffentlicher Finanzen theoretisch mit den Vor- und Nachteilen entsprechender Maßnahmen auseinandersetzt (Wildasin, 1997; Inman, 2003; für einen Überblick siehe Baskaran, 2017). Zudem wurden jüngst vermehrt Arbeiten vorgelegt, welche die Effekte auf die kommunale Haushaltswirtschaft empirisch untersuchen (Pettersson-Lidbom, 2010; Allers, 2015; Baskaran, 2017). Allerdings wurden hierbei Schuldenhilfen stets als einheitliches Instrument aufgefasst. Eine Untersuchung unterschiedlicher Ausgestaltungsmöglichkeiten und eine Differenzierung unterschiedlicher Arten von Schuldenhilfen erfolgte nicht. Auch eine Analyse der Ursachen, d. h. wann und warum Schuldenhilfen vergeben werden (oder nicht) bzw. unter welchen Bedingungen unterschiedliche Instrumente ausgewählt wurden, unterblieb. Mit Blick auf Deutschland lässt sich zudem feststellen, dass Schuldenhilfen an die Kommunen im wissenschaftlichen Diskurs wenig Beachtung fanden. Insbesondere bundesländerübergreifende Studien existieren kaum (Ausnahmen sind beispielsweise Stolzenberg & Heinelt, 2013; Geißler, 2009).

Um diese Forschungslücken zu adressieren, analysiert der Beitrag folgende Fragestellungen: Wie lassen sich die empirisch beobachtbaren, unterschiedlichen Varianten kommunaler Schuldenhilfen kategorisieren? Und warum wurden spezifische Arten von Schuldenhilfen zu bestimmten Zeitpunkten durch manche Bundesländer bereitgestellt, von anderen jedoch nicht? Zur Beantwortung der Leitfragen wird zuerst die Empirie kommunaler Schuldenhilfen im Zeitablauf beschrieben, um unterschiedliche Strategien der Länder im Umgang mit lokalen Haushaltskrisen zu identifizieren. Darauf aufbauend wird induktiv eine Kategorisierung von Schuldenhilfen anhand zentraler Ausgestaltungsmerkmale entwickelt. Anschließend wird unter Zuhilfenahme gängiger Analyse-

kategorien der Policy-Analyse erklärt, warum manche Länder zu einem gegebenen Zeitpunkt durch die Bereitstellung einer bestimmten Form von Schuldenhilfen auf kommunale Finanzprobleme reagierten, während in anderen Ländern entsprechende Reaktionen ausblieben. Die Analyse stützt sich auf qualitative Daten, welche durch Experteninterviews mit Mitarbeitern<sup>2</sup> von Aufsichtsbehörden, Kommunen und kommunalen Spitzenverbänden sowie Rechtsdokumenten, parlamentarischen Drucksachen, Plenarprotokollen und Sekundärliteratur erhoben wurden.

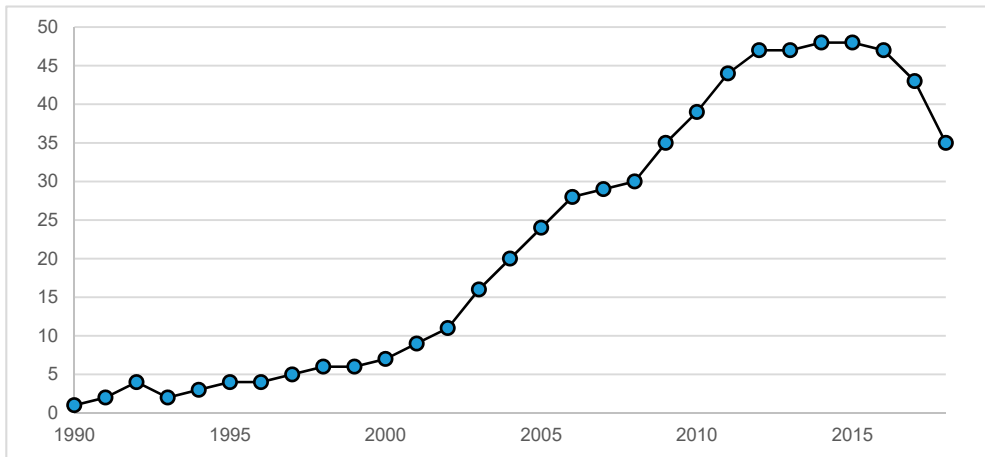
Der Beitrag gliedert sich wie folgt: Kapitel 2 skizziert die kommunale Haushaltssituation. Im dritten Kapitel wird der Begriff der Schuldenhilfe definiert sowie Vor- und Nachteile diskutiert. Das vierte Kapitel beschreibt die Historie kommunaler Schuldenhilfen. Darauf aufbauend werden in Kapitel 5 unterschiedliche Arten von Schuldenhilfen systematisiert, divergierende Ausgestaltungsformen und Implementationszeitpunkte kommunaler Entschuldungsprogramme erklärt sowie deren Auswirkungen auf die fiskalische Performanz und die Selbstverwaltung der Kommunen diskutiert. Das letzte Kapitel zieht ein kurzes Fazit.

## 2 Entwicklung der kommunalen Haushaltssituation

Auf den ersten Blick scheint die Lage der Kommunalhaushalte in Deutschland unproblematisch. Seit 2000 war der aggregierte Haushaltssaldo aller Kommunen lediglich in sieben Jahren negativ. Obwohl die Kommunalverschuldung in diesem Zeitraum um ein Drittel anstieg, machte sie 2018 nur sieben Prozent der gesamten Staatsverschuldung aus. Allerdings überdecken diese Zahlen die haushaltspolitische Realität vieler Kommunen. Diese ist seit Jahrzehnten von Haushaltskrisen geprägt. Der übliche Indikator für deren Erfassung sind die Kassenkredite.<sup>3</sup> Wie *Abbildung 1* verdeutlicht, stiegen diese seit 2000 dramatisch an. Ursachen waren u. a. die schwache wirtschaftliche Entwicklung und massive Steuersenkungen zu Beginn des Jahrtausends. Mitte der 2000er stabilisierte sich die Situation. Die globale Finanzkrise traf die Kommunen jedoch hart und führte zu einem erneuten, sprunghaften Anstieg der Kassenkredite.<sup>4</sup> Erst in der jüngeren Vergangenheit stabilisierte sich das Gesamtvolumen der Kassenkredite und es begann, nicht zuletzt infolge kommunaler Entschuldungsprogramme, zu sinken.

Ein Blick auf die regionale Verteilung der Kassenkredite verdeutlicht die dramatische Situation vieler Kommunen (*Abbildung 2*). Kassenkredite und Haushaltskrisen sind kein bundesweites Phänomen, sondern regional stark konzentriert. Im Jahr 2017 kumulierte die Hälfte der bundesweiten Kassenkredite in nur 20 Kreisen und kreisfreien Städten. Diese liegen v. a. in Nordrhein-Westfalen, Rheinland-Pfalz, Hessen und dem Saarland. In vielen Fällen handelt es sich um alte Industriestädte, die einen langfristigen wirtschaftlichen Niedergang durchlitten haben (Boettcher, Freier, Geißler, Schubert & Stollhoff, 2019, Teil D).

**Abbildung 1:** Bundesweites Volumen kommunaler Kassenkredite (Kernhaushalte), in Mrd. Euro



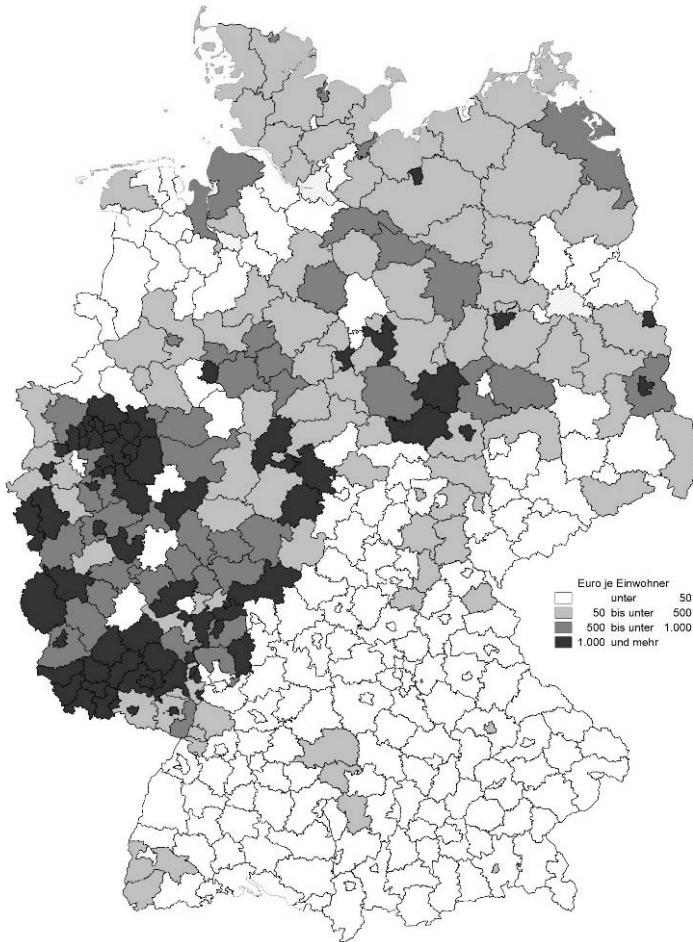
Quelle: DESTATIS, Schuldenstatistik.

Die haushaltspolitische Realität dieser Kommunen besteht aus permanentem Haushaltsdruck, anhaltenden Budgetdefiziten, steigenden Steuern, sinkenden Investitionsausgaben und Personalabbau. Diese Entwicklungen zeigen direkte Effekte auf die Qualität kommunaler Leistungen und werden seit vielen Jahren öffentlich thematisiert. Nicht zufällig stellte die Bundesregierung jüngst diverse Hilfsmaßnahmen bereit, um diesen Städten zu helfen (Boettcher & Geißler, 2017, S. 7 f.).

Zwar profitierten diese Städte neben Entlastungen durch den Bund in der jüngeren Vergangenheit auch von steigenden Steuereinnahmen und höheren Schlüsselzuweisungen. Die strukturellen Probleme wie beispielsweise das Zusammentreffen niedriger Steuerkraft und hoher Sozialausgaben oder die Unmöglichkeit der eigenständigen Tilgung der aufgelaufenen Kassenkredite bestehen jedoch fort. Zudem weisen diese Städte eine geringere Resilienz gegenüber konjunkturellen Abkühlungen auf, welche ihre Haushaltslage erneut verschärfen würden (Boettcher, Freier, Geißler, Schubert & Stollhoff, 2019, Teil E). Angesichts dieser Problemlagen wurde wiederholt über Schuldenhilfen als Instrument zur Überwindung kommunaler Haushaltskrisen diskutiert. Allerdings stellt dieses Instrument ein zweischneidiges Schwert dar, wie das nächste Kapitel aufzeigt.



**Abbildung 2:** Kassenkredite je Einwohner auf Ebene der Kreise und kreisfreien Städte, in Euro, 2017



Quelle: Wegweiser Kommune.

### 3 Vor- und Nachteile von Schuldenhilfen aus theoretischer Perspektive

Bevor die Praxis kommunaler Schuldenhilfen beleuchtet wird, ist es notwendig, den Begriff der Schuldenhilfe (im angelsächsischen Schrifttum als *bailout* geläufig) zu definieren. In Anlehnung an David Wildasin (2004, p. 253) versteht man unter Schuldenhilfen die Bereitstellung außergewöhnlicher, irregulärer Fiskaltransfers, welche notwendig sind, um eine unmittelbar drohende Haushaltskrise einer Kommune zu adressieren; eine Krise, die durch suboptimale fiskalische Entscheidungen und finanzielle Verschwendung angesichts der Existenz einer weichen Budgetbeschränkung hervorgerufen wird.<sup>5</sup> Sie stellen diskretionäre Maßnahmen übergeordneter Regierungen dar, um einer Kommune ad hoc zusätzliche Finanzmittel bereitzustellen, wenn diese ansonsten

unfähig wäre, ihren Verpflichtungen nachzukommen (Baskaran, 2017, p. 211; Rodden, Eskeland & Litvack, 2003, p. 8).

Hierbei wird auf Basis der Neuen Institutionenökonomik eine Prinzipal-Agenten-Beziehung zwischen Land und Kommunen unterstellt. Das Land als Prinzipal überträgt die Wahrnehmung vielfältiger Aufgaben auf die Kommunen und stellt Finanzmittel zu deren Erfüllung bereit. Bei der Aufgabenwahrnehmung unterliegen die Kommunen jedoch lediglich einer weichen Budgetbeschränkung (Kornai, 1986). Im Falle finanzieller Schwierigkeiten erwarten sie, von ihrem Prinzipal mit zusätzlichen Finanzmitteln unterstützt zu werden, da dieser kein Interesse an einer kommunalen Zahlungsunfähigkeit und den damit verbundenen negativen Auswirkungen hat. Dadurch entsteht ein *moral hazard*-Problem und die Anreize der Kommunen zu einer disziplinierten Haushaltspolitik sinken (Pettersson-Lidbom, 2010, pp. 154-155).

Zwar kann der Prinzipal eine *no-bailout*-Klausel postulieren, um diese Erwartungshaltung einzudämmen. Er unterliegt hierbei jedoch einem Glaubwürdigkeitsproblem (Problem der Zeitinkonsistenz). Das heute gegebene Versprechen eines *no-bailout* erscheint wenig glaubwürdig, wenn sich die Rahmenbedingungen aufgrund einer drohenden kommunalen Zahlungsunfähigkeit zukünftig verändern. Unter diesen Umständen kann sich der Prinzipal gezwungen sehen, seine ursprüngliche Position zu revidieren (Inman, 2003, p. 43; Allers, 2015, pp. 453-454).

Für die betroffenen Länder und Kommunen stellt die Bereitstellung von Schuldenhilfen eine bedeutsame Entscheidung dar. Folglich existiert in der ökonomischen Fachliteratur eine anhaltende theoretische Debatte über deren Vorzüge und Nachteile. Aus Sicht von Befürwortern sind Schuldenhilfen notwendig, um auch in Haushaltskrisen die Erbringung öffentlicher Leistungen zu gewährleisten (Rodden, 2002, p. 672). Außerdem können durch die Vergabe von Schuldenhilfen eine Zahlungsunfähigkeit lokaler Gebietskörperschaften und damit einhergehende negative Externalitäten wie ein flächendeckender Anstieg der Kreditrefinanzierungskosten für andere Gebietskörperschaften und finanzielle *spillover*-Effekte auf andere Wirtschaftssektoren vermieden werden (Seitz, 1999, p. 1; Brand, 2013, S. 89). Zudem würde die Zahlungsunfähigkeit größerer Städte auch die Bereitstellung zentralörtlicher Leistungen für ihr Umland gefährden (Wildasin, 1997). Darüber hinaus werden Schuldenhilfen als gerechtfertigt angesehen, wenn Gebietskörperschaften nur über begrenzte fiskalische Handlungsspielräume verfügen und ihre Einnahmehasis nicht autonom an ihre Ausgabenbedarfe anpassen können, sodass man sie nicht für ihre Finanzprobleme verantwortlich machen kann (Rodden, Eskeland & Litvack, 2003, p. 11). Des Weiteren wird argumentiert, dass übergeordnete Regierungen in der Pflicht stehen, Eigentumsrechte zu garantieren. Demzufolge müssten sie Gläubiger unterstützen, ihre vertraglich zugesicherten Zahlungen zu erhalten und Kreditverträge gegen zahlungsunfähige Gebietskörperschaften durchzusetzen. Gelingt dies nicht, erscheint es legitim, wenn sie deren Zahlungsverpflichtungen übernehmen (Rodden, Eskeland & Litvack, 2003, p. 11).

Im Gegensatz dazu betonen Kritiker von Schuldenhilfen deren hohe Kosten, insbesondere wenn die Finanzprobleme der Empfänger groß sind. Außerdem können Schuldenhilfen mit *moral hazard*-Problemen einhergehen. Wenn Gebietskörperschaften erwarten können, dass sie im Falle finanzieller Schieflagen gerettet werden, haben sie geringere Anreize, eine sorgsame, nachhaltige Haushaltspolitik zu betreiben (Seitz, 1999, S. 1; Baskaran, 2017; Allers, 2015). Aus Sicht einer Gebietskörperschaft reduziert die Wahrscheinlichkeit eines *bailouts* die wahrgenommenen Kosten der Kreditaufnahme

und setzt Anreize, sich zu überschulden (Inman, 2003, pp. 36-37). Auch Kreditgeber haben unter diesen Umständen keine Anreize, die fiskalische Performanz und Kreditwürdigkeit öffentlicher Schuldner sorgfältig zu prüfen, da sie erwarten, ihr Geld ohnehin zurückzubekommen (Brand, 2013, S. 89). Folglich stört die Vergabe von Schuldenhilfen die Wirkungsweise des Kapitalmarktmechanismus, da eine höhere Verschuldung nicht durch höhere Risikoaufschläge und steigende Refinanzierungskosten bestraft wird. Die Märkte fungieren nicht als effizientes Instrument zur Durchsetzung fiskalischer Disziplin (Canuto & Liu, 2013, p. 4).

Allerdings sind die empirischen Befunde bezüglich der postulierten Verschuldungsanreize infolge von Schuldenhilfen gemischt. Während Per Pettersson-Lidbom (2010) einen signifikanten Anstieg der Verschuldung nach der Durchführung von *bailouts* in schwedischen Kommunen beobachtet, stellt Thushyanthan Baskaran (2017) anhand der Gemeinden Hessens keine dauerhaft höhere Verschuldung, sondern einen Abbau von Verbindlichkeiten sowie strikte Haushaltskonsolidierung fest. Auch in den Niederlanden, in denen ein institutionalisiertes *bailout*-System für Kommunen existiert, lassen sich nach Maarten Allers (2015) keine Verschuldungsanreize nachweisen. Die konkrete Ausgestaltung der Schuldenhilfen scheint großen Einfluss darauf zu haben, ob sich die theoretisch erwarteten Vor- und Nachteile manifestieren. Umso überraschender ist daher, dass die Frage nach dem unterschiedlichen Design kommunaler Schuldenhilfen und den Rahmenbedingungen, unter denen diese gewährt werden, im wissenschaftlichen Diskurs nur eine nachgeordnete Rolle spielt. Um diese Forschungslücke zu adressieren, werden im Folgenden anhand der Praxiserfahrungen Deutschlands unterschiedliche Ausgestaltungsmöglichkeiten aufgezeigt und systematisiert sowie analysiert, welche Faktoren unterschiedliche Ausgestaltungsformen und Anwendungszeitpunkte beeinflussen.

## 4 Historische Entwicklung kommunaler Schuldenhilfen

Das vorliegende Kapitel gibt einen historischen Abriss über die Vergabe kommunaler Schuldenhilfen. Hierbei werden sowohl die Bereitstellung diskretionärer Bedarfszuweisungen zur Haushaltssicherung innerhalb der KFA-Systeme als auch spezielle, temporäre Entschuldungsprogramme betrachtet. Dieser deskriptiv-kursorische Überblick erlaubt die Identifikation unterschiedlicher Reaktionsmöglichkeiten auf kommunale Haushaltskrisen, von Mustern und spezifischen Entwicklungspfaden. Er bildet die empirische Grundlage, auf deren Basis in Kapitel 5 unterschiedliche Arten kommunaler Schuldenhilfen entsprechend ihrer relevanten Charakteristika kategorisiert werden. Gleichzeitig vermittelt er einen Eindruck davon, dass Schuldenhilfen zu unterschiedlichen Zeiten in unterschiedlichem Ausmaß von einzelnen Bundesländern genutzt wurden – ein Sachverhalt, der ebenfalls im nächsten Kapitel eingehender analysiert wird.

### 4.1 1980-1991: Erste Haushaltskrisen in Nordrhein-Westfalen

Anfang der 1980er gewann der Fachterminus „Kassenkredite“ in Diskussionen über die kommunale Haushaltssituation zunehmend an Prominenz. Im Jahr 1982 erreichten diese mit 1,2 Mrd. € einen ersten Höhepunkt. Die Entwicklung der Kassenkredite spie-

gelte die Entstehung beachtlicher Haushaltsdefizite wider, verursacht durch die Abschaffung der Lohnsummensteuer und den anhaltenden Strukturwandel. Diese Problemlagen trafen besonders die altindustriellen Zentren Nordrhein-Westfalens (Held, 1995, S. 59-60; Diemert, 2005, S. 78).

Der wachsende Problemdruck veranlasste Nordrhein-Westfalen, die im Gemeindefinanzierungsgesetz vorgesehenen Bedarfszuweisungen zum Ausgleich strukturell bedingter Haushaltsdefizite für größere Kommunen zu öffnen (Held, 1996, S. 74-75). Im Jahr 1987 konnten erstmals auch Städte über 25.000 Einwohnern diese Finanzhilfen beantragen, sofern sie spezifische Bedingungen erfüllten. Insgesamt nutzten sieben Städte diese Option. Im Zuge dessen erfind das Land mit dem Haushaltssicherungskonzept (HSK) ein Steuerungsinstrument, das in den kommenden Jahren weite Verbreitung fand. Mittels der HSK mussten Städte eigene Haushaltskonsolidierungsmaßnahmen umsetzen.

Diese Bedarfszuweisungen wiesen jedoch Nachteile auf: Sie wurden dem KFA entnommen und waren nach erfolgreicher Konsolidierung hälftig zurückzuzahlen (Held, 1995, S. 60). Die Aufsichtsbehörden mussten dem HSK zustimmen, konnten allerdings seine Umsetzung nicht effektiv durchsetzen. Im Falle der Nichteinhaltung verloren die Städte lediglich die Zuschüsse. Außerdem war die Inanspruchnahme freiwillig (Geißler, 2009, S. 20 f.). Folglich war die Bilanz dieser Schuldenhilfe gemischt: Die Haushaltssicherungshilfen zeigten kurzfristig positive Effekte und lösten die dringlichsten Haushaltsprobleme. Eine dauerhafte Haushaltskonsolidierung bewirkten sie nicht (Diemert, 2005, S. 80).

1991 änderte das Land seine Politik, strich diese Bedarfszuweisung ersatzlos und fokussierte auf eine strengere Kommunalaufsicht, da dies zur Lösung der Haushaltsprobleme besser geeignet erschien (Geißler, 2009, S. 19 ff.). Allerdings gewährte das Land einer Gruppe kleinerer Gemeinden eine einmalige Tilgungshilfe, um diesen die Einhaltung des neuen Regulierungsrahmens zu erleichtern (Held, 1995, S. 61).

## 4.2 1991-2009: Ausbreitung kommunaler Haushaltskrisen in Deutschland

Im Nachklang der Deutschen Einheit erlebte (West-)Deutschland eine räumliche Ausweitung kommunaler Haushaltsprobleme. Die westdeutschen Gemeinden waren direkt von den Kosten der Wiedervereinigung betroffen. Seit 1991 mussten sie über eine einheitsbedingte Erhöhung der Gewerbesteuerumlage zusätzliche Anteile ihrer Gewerbesteuererinnahmen an Bund und Länder abführen. Im Zuge der Integration der ostdeutschen Länder in das Finanzausgleichssystem wurde die Gewerbesteuerumlage 1995 nochmals erhöht (Schwarting, 2019, S. 95-96). Diese Eingriffe in die Gemeindesteuern in Kombination mit wirtschaftlicher Stagnation und steigender Langzeitarbeitslosigkeit führten in vielen Kommunen zu wachsendem Haushaltsdruck.

In dieser Periode reagierten alle Länder (außer Nordrhein-Westfalen) auf lokale Haushaltskrisen primär durch die Vergabe von Bedarfszuweisungen. Allerdings zeigten sich zunehmend Schwächen dieser Herangehensweise (Geißler, 2009, S. 67 ff.): Das Volumen der Finanzhilfen war zu gering, um eine wirksame Unterstützung für die wachsende Zahl betroffener Gemeinden zu generieren. Die Länder reagierten mit immer härteren Empfangsaufgaben und stets kleineren Zuteilungen. Infolgedessen schwan-

den ihre Anreizwirkungen. Die Bedarfszuweisungen erwiesen sich zunehmend als ungeeignet zur Lösung kommunaler Haushaltsprobleme. Dennoch hielten die meisten Länder an ihnen fest (Geißler, 2009, S. 51 ff.).

Neben den Bedarfszuweisungen gab es kaum offizielle Diskussionen über Schuldenhilfen in diesem Zeitraum, wie eine Auswertung der Parlamentsdokumentationen ergab. Die einzige Ausnahme war Bayern, was angesichts der soliden Finanzlage der bayerischen Gemeinden überrascht (Landtag Bayern, 2003, S. 125 ff.). Im Jahr 2006 startete Bayern ein Pilotprojekt zur kommunalen Haushaltskonsolidierung mit einem Volumen von 56 Mio. € (Landtag Bayern, 2018). Über vier Jahre erhielten 32 strukturschwache Gemeinden mit großen Haushaltsproblemen jährliche Finanzhilfen, um ihre Schuldenlast zu reduzieren. Im Gegenzug mussten sie ein HSK entwickeln, Sparmaßnahmen ergreifen und wurden einer strengeren Finanzaufsicht unterworfen. Im Falle einer Nichteinhaltung konnten Zuschüsse zurückgefordert werden. Die Teilnahme war freiwillig. Insgesamt war das Programm hochselektiv, da nur 32 von 2056 Gemeinden teilnahmen, die weniger als 1% der Bevölkerung abdeckten (Landtag Bayern, 2011). Aufgrund des Programmerfolgs wurden die Stabilisierungshilfen 2012 in den KFA integriert.

Zudem gab es in Ostdeutschland Diskussionen über Schuldenhilfen für Gemeinden, welche durch die Errichtung überdimensionierter Abwasser- und Abfallentsorgungseinrichtungen in Schwierigkeiten gerieten (Landtag Brandenburg, 1997; Landtag Mecklenburg-Vorpommern, 1998; Landtag Sachsen, 1997; Landtag Sachsen-Anhalt, 1999; Landtag Thüringen, 1995). Diese Infrastruktureinrichtungen, im Zuge der Wiedervereinigung gebaut, führten aufgrund von Planungsfehlern, Bevölkerungsrückgang und ökonomischen Fehlentscheidungen zu massiven Verlusten, die einzelne Kommunalhaushalte erheblich belasteten. Daher stellten die ostdeutschen Länder Finanzhilfen bereit, um die Probleme kommunaler Auslagerungen zu adressieren.

Abgesehen hiervon standen Schuldenhilfen vor der globalen Finanzkrise in keinem Bundesland auf der Agenda. Stattdessen reagierten die Länder mit einer Verschärfung des haushaltsrechtlichen Rahmens und einer Stärkung der Kommunalaufsicht (Geißler, 2009, S. 27 ff.). Das HSK wurde zu einem Massenphänomen. In der Folge entstanden komplexe, informelle Verhandlungssysteme zwischen Kommunen und Aufsichtsbehörden. Allerdings provozierten die immer stärkeren Interventionen der Länder zunehmenden Widerstand der Lokalpolitik (Holtkamp, 2010). Zudem war der Trend zu stärkerer Regulierung nicht in allen Ländern einheitlich. Beispielsweise zeichnete sich Nordrhein-Westfalen durch eine volatile Gesetzgebung aus und lockerte die Haushaltsregeln mehrfach angesichts zunehmender Haushaltsdefizite und lokalpolitischer Widerstände (Person & Geißler, 2020). Ein weiterer Fall politisch getriebener Lockerung von Fiskalregeln ist Rheinland-Pfalz, welches 2005 das erst neun Jahre zuvor eingeführte HSK wieder abschaffte (Geißler, 2009, S. 49 f.). Trotz der Intensivierung haushaltsrechtlicher Regulierungen und der Vergabe von Bedarfszuweisungen wuchs die Haushaltskrise weiter an.

### 4.3 2009-2013: Neue Hilfsprogramme im Zuge der globalen Finanzkrise

Die kommunale Haushaltskrise eskalierte im Kontext der internationalen Finanzkrise. Die Haushaltsdefizite erreichten Rekordwerte (2010: neun Mrd. €), ebenso die Kassenkredite (2012: 47 Mrd. €). Im Gegensatz zur Vorperiode reagierte eine Mehrheit von acht Ländern mit speziellen Entschuldungsprogrammen. Einige Länder wählten markante Namen wie „Stärkungspakt Stadtfinanzen“ (Nordrhein-Westfalen), „Zukunftsvertrag“ (Niedersachsen) oder „Schutzschirm“ (Hessen). Die Gestaltung dieser Programme unterschied sich stark zwischen den Ländern wie *Tabelle 1* verdeutlicht (siehe hierzu auch Arnold, Boettcher, Freier, Geißler & Holler, 2015; Stolzenberg & Heinelt, 2013). Lediglich fünf Bundesländer sahen keine Notwendigkeit für Entschuldungsprogramme.

*Tabelle 1:* Kommunale Entschuldungsprogramme im Kontext der globalen Finanzkrise

| Land | Teilnahme                   | Betr. Bevölke-<br>rung in % | Dauer     | Ziel                    | Art der Schuldenhilfe                                                                                                     | Volumen<br>(Mrd. €) | Kassenkre-<br>dite (Mrd. €) |
|------|-----------------------------|-----------------------------|-----------|-------------------------|---------------------------------------------------------------------------------------------------------------------------|---------------------|-----------------------------|
| HE   | freiwillig                  | 26                          | 2013-2043 | Schulden-<br>abbau      | Schuldenübernahme:<br>Teil-Umschuldung von<br>Kassen- und Investitions-<br>krediten (max. 46%)<br>Zinshilfen              | 3,200               | 7,488 (2012)                |
| MV   | freiwillig                  | keine Daten                 | 2012-2020 | Haushalts-<br>ausgleich | Konsolidierungshilfen:<br>Jährliche Zuweisungen zum<br>Haushaltsausgleich                                                 | 0,100               | 0,549 (2011)                |
| NI   | freiwillig                  | 20                          | 2010-2016 | Schulden-<br>abbau      | Schuldenübernahme:<br>Teil-Umschuldung von<br>Kassenkrediten (75%)<br>Zinshilfen                                          | 2,048               | 4,534 (2009)                |
| NW   | freiwillig und<br>pflichtig | 29                          | 2011-2020 | Haushalts-<br>ausgleich | Konsolidierungshilfen:<br>Jährliche Zuweisungen zum<br>Haushaltsausgleich                                                 | 5,850               | 18,703 (2010)               |
| RP   | freiwillig                  | 59                          | 2012-2026 | Schulden-<br>abbau      | Tilgungshilfen:<br>Jährliche Zuweisungen zum<br>Abbau von Kassenkrediten                                                  | 3,825               | 5,598 (2011)                |
| SH   | freiwillig                  | 26                          | 2012-2018 | Haushalts-<br>ausgleich | Konsolidierungshilfen:<br>Jährliche Zuweisungen<br>zum Haushaltsausgleich                                                 | 0,420               | 0,766 (2011)                |
| SL   | freiwillig                  | 64                          | 2013-2019 | Schulden-<br>abbau      | Tilgungshilfen:<br>Jährliche Zuweisungen<br>zum Schuldenabbau                                                             | 0,120               | 1,872 (2012)                |
| ST   | freiwillig                  | 92                          | 2013-2018 | Schulden-<br>abbau      | Schuldenübernahme:<br>Langfristige Umschuldung<br>und teilweise Schulden-<br>übername (30%) von Inves-<br>titionskrediten | 0,513               | 3,317 (2012)                |

*Anmerkungen:* Dauer umfasst bei Niedersachsen und Sachsen-Anhalt die Antragsphase; für Sachsen-Anhalt ist die Gesamtverschuldung statt der Kassenkredite ausgewiesen.

*Quelle:* Eigene Darstellung, Kassenkredite nach DESTATIS, Kern- und Extrahaushalte beim nicht-öffentlichen Bereich.

Die Teilnahme an den Hilfsprogrammen war grundsätzlich freiwillig. Lediglich in Nordrhein-Westfalen wurden einige Städte zur Teilnahme verpflichtet. Der Kreis potenzieller Teilnehmer wurde durch Indikatoren wie Kassenkredite, Steuereinnahmen, usw. eingeschränkt. Dieses Verfahren hatte aus Sicht der Länder Vorteile: Über Indikatoren lässt sich die Auswahl der Teilnehmer transparent begründen. Die freiwillige Teilnahme erzeugt eine höhere lokale Motivation und begrenzt den politischen Widerstand. Faktisch war die Teilnahme jedoch nicht freiwillig, sondern lokalpolitisch geboten.

Der Anteil der Landesbevölkerung, der in Programm-Kommunen lebt, spiegelt nicht nur das Ausmaß der Haushaltskrise wider, sondern auch die räumliche Konzentration der Finanzhilfen, was auf eine breitere oder engere Ausrichtung der Programme hinweist. Vor dem Hintergrund der finanziellen und politischen Auswirkungen der Programmteilnahme sind diese Anteilswerte beachtlich. In jedem Land leben mindestens 20% der Bevölkerung in finanziell angeschlagenen Kommunen, in manchen sogar mehr als die Hälfte.

Allerdings litten nicht alle Kommunaltypen in allen Bundesländern in gleicher Weise unter der Haushaltskrise. Die Hälfte der Länder (Hessen, Niedersachsen, Sachsen-Anhalt, Rheinland-Pfalz) öffnete die Programme für alle Kommunen, auch die Landkreise. Der Rest beschränkte die Teilnahme auf die Gemeinden. Diese Entscheidung spiegelt die Struktur der Haushaltskrise und die verfügbaren Mittel wider.

Hinsichtlich der Ziele und Maßnahmen zeigt *Tabelle 1* deutliche Unterschiede. Einige Länder konzentrierten sich auf den Haushaltsausgleich und stellten jährliche Konsolidierungshilfen bereit (z. B. Mecklenburg-Vorpommern, Nordrhein-Westfalen, Schleswig-Holstein). Andere fokussierten auf den Schuldenabbau, entweder in Bezug auf Kassenkredite (Niedersachsen), Investitionskredite (Sachsen-Anhalt) oder beides (Hessen). Grundsätzlich waren die Hilfen an Auflagen geknüpft: Kommunen mussten Konsolidierungsmaßnahmen beschließen und der Kommunalaufsicht über deren Umsetzung berichten. Dadurch sollten der im vorherigen Kapitel diskutierten Anreizproblematik, die mit Schuldenhilfen verbunden ist, begegnet werden. Aufgrund des kommunalen Selbstverwaltungsrechts konnten die Länder keine Maßnahmen vorschreiben. Sie gaben allerdings Empfehlungen und setzten Mindeststeuersätze fest. Nach Felix Arnold, Florian Boettcher, Ronny Freier, René Geißler und Benjamin Holler (2015) verteilten sich die Konsolidierungsmaßnahmen hälftig auf Ausgabekürzungen und Einnahmesteigerungen. Interessant ist das Fallbeispiel Niedersachsen. Das Land verfolgte zwei Ziele simultan: Die Bewältigung der Haushaltskrise und die Überwindung fragmentierter Gemeindefstrukturen. Schuldenhilfen machten daher Gemeindefusionen zur Bedingung.

Die Programmlaufzeit ist aufgrund der unterschiedlichen Maßnahmen schwierig zu fassen und kann Verschiedenes beinhalten: Den Antragszeitraum, den Zeitraum für die Zielerreichung, den Zeitraum für den Erhalt der Schuldenhilfen oder den Zeitraum der Schuldentilgung. Die Laufzeit des nordrhein-westfälischen Programms beträgt beispielsweise zehn Jahre. Innerhalb von fünf Jahren mussten die Teilnehmer den Haushaltsausgleich mit Hilfe der Landeszuweisungen erreichen. In den nächsten fünf Jahren mussten sie den Haushaltsausgleich bei sinkenden Zuschüssen bewahren. In Hessen verpflichteten sich die Programmkommunen für 30 Jahre zur Zahlung von Annuitäten, um die umgeschuldeten Kassenkredite abzutragen.

Bezüglich der Verantwortlichkeit für die Programme zeigen sich ebenfalls Divergenzen. In sechs Ländern entschied das Innenministerium über die Anträge und die

Angemessenheit lokaler Konsolidierungspläne. Im Gegensatz zur regulären Struktur der Kommunalaufsicht wiesen zwei Länder dem Finanzministerium die Programmverantwortung zu. Hierfür gibt es zwei Erklärungen: Erstens waren auch frühere Schuldenhilfen (Bedarfszuweisungen) Angelegenheit der Finanzministerien. Zweitens wurde in Hessen die Kompetenzverlagerung auch auf Steuerungsdefizite des Innenministeriums zurückgeführt.

Ein weiteres Merkmal von höchster politischer Relevanz ist die Finanzierung der Programme. In einigen Ländern trugen die Landeshaushalte diese Hilfsprogramme selbst (z. B. Hessen, Mecklenburg-Vorpommern, Saarland, Sachsen-Anhalt). Anderorts wurden die Kommunen an der Refinanzierung beteiligt. In Nordrhein-Westfalen stammen 40% der Mittel aus der kommunalen Ebene. Das niedersächsische Entschuldungsprogramm wird gleichmäßig von Land und Kommunen finanziert. Noch umstrittener war der Fall Rheinland-Pfalz: Die Schuldenhilfen wurden zu gleichen Teilen vom Land, der kommunalen Ebene und den beteiligten Städten finanziert. Auch das Programm Schleswig-Holsteins war kontrovers, da weitgehend durch KFA-Mittel finanziert.

#### 4.4 Seit 2017: „Große Lösungen“ dank guter Einnahmeentwicklung der Länder

Im Laufe der Zeit liefen die Diskussionen in manchen Ländern aus (Sachsen-Anhalt, Niedersachsen), sei es durch reale Verbesserungen oder durch lange Programmlaufzeiten. In sechs Bundesländern (Hessen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Schleswig-Holstein) erhielt die Diskussion neue Impulse und führte zu erneuten landespolitischen Aktivitäten. Mit Brandenburg ist ein zusätzliches Land aktiv geworden. Um diese neuesten Programme von den früheren abzugrenzen, werden sie als Entschuldungsprogramme der „zweiten Generation“ benannt (siehe *Tabelle 2*).

Den größten Schritt ging erneut Hessen. Nachdem das Land knapp die Hälfte der Kassenkredite im ersten Hilfsprogramm umgeschuldet hatte, legte es nun ein zweites Programm (Hessenkasse) für die restlichen Kassenkredite auf. Zum Ablauf des Jahres 2018 waren die hessischen Kommunen statistisch frei von Kassenkrediten. Praktisch gründete das Land ein Sondervermögen, das sich im notwendigen Volumen verschuldete und die kommunalen Kassenkredite ablöste. Land und Kommunen tragen die Tilgung der umgeschuldeten Kredite über Jahrzehnte gemeinsam. Außerdem änderte das Land das Haushaltsrecht, um die Entstehung neuer Kassenkredite zu verhindern. Parallel zur Schuldenhilfe beschloss Hessen ein Investitionsförderprogramm.

Mecklenburg-Vorpommern legte 2018 ebenfalls ein neues Entschuldungsprogramm auf (Kommunaler Entschuldungsfonds). Das Programm zielt auf den Abbau von Haushaltsdefiziten und ab 2020 auch auf den Abbau von Altschulden kommunaler Wohnungsbauunternehmen. Die Finanzierung erfolgt mittelfristig über Mehreinnahmen aus dem bundesstaatlichen Finanzausgleich. Im Gegensatz zu anderen Programmen existiert keine zeitliche Begrenzung. Da das vorhergehende Programm bis 2020 läuft und die Fehlbedarfzuweisungen des KFA ohnehin existieren, führt Mecklenburg-Vorpommern für einige Jahre drei Hilfsprogramme mit Unterschieden in Zielen, Ausgestaltung und Umfang parallel.



Tabelle 2: Entschuldungsprogramme der „zweiten Generation“

| Land | Teilnahme  | Empfänger                           | Dauer                   | Ziel               | Art der Schuldenhilfe                                                                                                                                                                                                                                                                                                  | Zuständigkeit                         | Volumen (Mrd. €)   | Kassenkredite (Mrd. €)       |
|------|------------|-------------------------------------|-------------------------|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|--------------------|------------------------------|
| BB   | freiwillig | drei kreisfreie Städte              | 2019-2023 (Förderung)   | Schuldenabbau      | Tilgungshilfen: Jährliche Zuweisungen an drei kreisfreie Städte zum Abbau von 40% der Kassenkredite; weitere 10% sind durch Eigenleistungen zu erbringen                                                                                                                                                               | Innenministerium                      | 0,211              | 0,651                        |
| HE   | freiwillig | offen für alle Gemeinden und Kreise | 2018-2048 (Tilgung)     | Schuldenabbau      | Schuldenübernahme: Vollständige Übertragung der verbliebenen Kassenkredite auf Sondervermögen des Landes; Annuitäten über max. 30 Jahre, zu 2/3 vom Land und zu 1/3 von Kommune finanziert<br>Parallel Auflage eines Investitionsprogramms                                                                             | Finanzministerium                     | ca. 5,000          | 5,642 (2017)<br>0,415 (2018) |
| MV   | freiwillig | offen für alle Gemeinden            | Seit 2018 (unbefristet) | Schuldenabbau      | Konsolidierungshilfen und Tilgungshilfen: Jährliche Zuweisungen zum Abbau von Haushaltsdefiziten und von Altschulden der Wohnungsbauunternehmen; finanziert über zusätzliche Einnahmen im bundesstaatlichen Finanzausgleich                                                                                            | Innenministerium                      | 0,070 (2018/2019)  | 0,394                        |
| RP   | freiwillig | 94 Kommunen<br><br>52 Kommunen      | 2019-2028 (Förderung)   | Schuldenabbau      | Zinshilfen: Zuweisungen für langfristige Umschuldungen von Kassenkrediten und Wertpapierschulden (Zinssicherungsschirm)<br>Tilgungshilfen: Zuweisungen zum Schuldenabbau (Stabilisierungs- und Abbaubonus) als Belohnung für eigenständigen Schuldenabbau oder Stabilisierung des Schuldenstands                       | Finanzministerium                     | 0,180<br><br>0,120 | 5,320                        |
| SH   | freiwillig | vier kreisfreie Städte              | 2019-2023 (Förderung)   | Haushaltsausgleich | Konsolidierungshilfen: Jährliche Zuweisungen an kreisfreie Städte zum Abbau von Haushaltsdefiziten                                                                                                                                                                                                                     | Innenministerium                      | 0,225              | 0,375                        |
| SL   | freiwillig | offen für alle Gemeinden            | 2020-2065 (Tilgung)     | Schuldenabbau      | Schuldenübernahme: Übernahme von 50% der Kassenkredite in ein Sondervermögen des Landes (Tilgung durch das Land); Finanzierung über Sonderzuweisungen des Bundes für finanzschwache Gemeinden;<br>Gemeinden verpflichten sich zur Tilgung der restlichen Kassenkredite<br>Parallel Auflage eines Investitionsprogramms | Finanzministerium<br>Innenministerium | 1,000              | 1,937                        |

Quelle: Eigene Darstellung, Kassenkredite 2018 nach DESTATIS Schuldenstatistik (Kern- und Extrahaushalte beim nicht-öffentlichen Bereich).

Im Jahr 2019 beschloss das Saarland eine massive Umschuldung der Kassenkredite (Saarlandpakt). Das Land löst auf Basis einzelgemeindlicher Verträge jeweils die Hälfte der Kassenkredite ab, überträgt sie auf ein Sondervermögen und tilgt sie über 45 Jahre. Die Kommunen verpflichten sich zur Tilgung der restlichen Kassenkredite. Zur Überwachung wird der bereits aus dem Vorgängerprogramm bestehende Sanierungsrat fortgeführt. Die Programmfinanzierung erfolgt über neue Zuweisungen im föderalen Finanzausgleich ab 2020. Parallel zur Schuldenhilfe wird ein Investitionsförderprogramm aufgelegt.

Rheinland-Pfalz reagierte auf die wachsenden kommunalen Forderungen mit einer Ergänzung des bestehenden Entschuldungsprogramms. Im Zeitraum von 2019 bis 2028 bietet das Land Zinshilfen für die langfristige Umschuldung und Anreize zur Schuldentilgung. Dieses Programm ist zwar minimal (30 Mio. € jährlich), wird aber dennoch zu gleichen Teilen vom Land und dem KFA finanziert. Die Teilnahme ist freiwillig und auf einen kleinen Teil der Kommunen beschränkt.

Schleswig-Holstein konzentriert sich in seinem neuen Programm exklusiv auf die vier kreisfreien Städte. Diese leiden unter dem höchsten Haushaltsdruck. Zwischen 2019 und 2023 können diese Städte jährliche Konsolidierungshilfen zum Abbau von Haushaltsdefiziten und aufgelaufenen Jahresfehlbeträgen beantragen. Brandenburg legte 2019 ebenfalls ein Hilfsprogramm für drei von vier kreisfreien Städten auf. Das Land finanziert die Tilgung von 40% der Kassenkredite über jährliche Zuweisungen binnen vier Jahren.

Auch in Nordrhein-Westfalen begannen 2018 ernsthafte Planungen für eine Umschuldung kommunaler Kassenkredite. Aufgrund des hohen Volumens (ca. 23 Mrd. €) verschob die Landesregierung jedoch die Entscheidung, um das Agieren der Bundesregierung abzuwarten.

Zusammenfassend lässt sich feststellen, dass auch die sechs Entschuldungsprogramme der „zweiten Generation“ deutliche Unterschiede aufweisen. Drei der sechs Programme stehen allen Kommunen offen, die anderen konzentrieren sich auf einen Ausschnitt der kommunalen Familie. Die Zuständigkeit zwischen Finanz- und Innenministerium ist hälftig verteilt. Die Programme Hessens und des Saarlands zielen in Volumen und Ausrichtung darauf ab, das Kassenkreditproblem endgültig zu lösen. Die kleineren Lösungen der übrigen Länder adressieren landesindividuell spezifische Problemlagen.

## 5 Analyse

### 5.1 Ausgestaltungsformen kommunaler Schuldenhilfen

Die Darstellung der historischen Entwicklung kommunaler Schuldenhilfen verdeutlicht, dass ein weiter Spielraum bezüglich ihrer Ausgestaltung existiert. Basierend auf diesen Informationen wurde eine Typologie von Schuldenhilfen abgeleitet (siehe *Tabelle 3*). Schuldenhilfen variieren nach Zielsetzung, Instrumentarium und Umfang. Sie haben jedoch eine Gemeinsamkeit: Sie sind stets an Auflagen geknüpft und erfordern Konsolidierungsmaßnahmen der Empfänger. Zwar sind auch Schuldenhilfen ohne Auflagen theoretisch denkbar. Allerdings würde dies die in Kapitel 3 aufgezeigten massiven Fehlanreize auf Seiten der kommunalen Schuldner verstärken. Die Folgen einer riskanten Haushaltspolitik ließen sich dann auf die Länder externalisieren. Die effekti-

ven Kosten der Verschuldung würden aus Sicht der Kommunen sinken und diese besäßen keine Anreize, eine nachhaltige Haushaltswirtschaft zu betreiben. Um entsprechende *moral-hazard*-Probleme zu vermeiden, wurden daher Schuldenhilfen in Deutschland stets mit Konsolidierungsaufgaben versehen. Dadurch wird den Kommunen signalisiert, dass sie sich im Falle einer Haushaltskrise zwar auf die Unterstützung durch die Länder verlassen können, dies jedoch mit schmerzhaften Haushaltskonsolidierungsprozessen einhergeht.

Tabelle 3: Arten von Schuldenhilfen

| Ziel               | Art der Schuldenhilfe                     | Umfang      | Fälle                                                                           |
|--------------------|-------------------------------------------|-------------|---------------------------------------------------------------------------------|
| Haushaltsausgleich | Konsolidierungs-/<br>Stabilisierungshilfe | anteilig    | MV (1G)<br>MV (2G)<br>NW (1980er)<br>NW (1G)<br>SH (1G)<br>SH (2G)              |
|                    |                                           | vollständig |                                                                                 |
|                    | Zinshilfe                                 | anteilig    | HE (1G)<br>NI (1G)<br>RP (2G)                                                   |
|                    |                                           | vollständig |                                                                                 |
| Schuldenabbau      | Tilgungshilfe                             | anteilig    | BB (2G)<br>NW (1980er)<br>BY (2006)<br>MV (2G)<br>RP (1G)<br>RP (2G)<br>SL (1G) |
|                    |                                           | vollständig |                                                                                 |
|                    | Schuldenübernahme                         | anteilig    | HE (1G)<br>NI (1G)<br>ST (1G)<br>SL (2G)                                        |
|                    |                                           | vollständig | HE (2G)                                                                         |

Quelle: Eigene Darstellung.

Hinsichtlich ihrer Zielsetzungen lassen sich zwei Gruppen von Schuldenhilfen unterscheiden. Die erste Gruppe zielt auf den Haushaltsausgleich, während die zweite Gruppe einen Schuldenabbau bewirken möchte. Innerhalb der ersten Kategorie finden sich zwei Instrumente: Haushaltskonsolidierungs-/ Stabilisierungshilfen und Zinshilfen. In Bezug auf ihren Umfang können Konsolidierungshilfen entweder das gesamte Haushaltsdefizit oder einen Anteil davon abdecken. Im letzteren Fall obliegt es der Kommune, die verbleibende Haushaltslücke zu schließen. Während Konsolidierungshilfen den Gesamthaushalt adressieren, fokussieren Zinszuschüsse auf eine bestimmte Ausgabenkategorie: Zinsen. Durch die Vergabe von Zinsdiensthilfen soll der Haushalt von hohen Zinsausgaben entlastet und dadurch der Haushaltsausgleich erleichtert werden, wobei auch hier eine anteilige oder vollständige Deckung der Zinsausgaben möglich ist.

Die zweite Kategorie von Schuldenhilfen zielt auf den Abbau bestehender Schulden. Auch diesbezüglich existieren zwei Instrumente. Die Empfänger können jährliche

Zuweisungen zur Schuldentilgung erhalten, wobei bestimmte Schuldenarten (z. B. Kassenkredite) priorisiert werden können. Die Zahl der jährlichen Zuweisungsraten ist ebenso offen wie der kumulierte Umfang der Zuweisungen. Solche Tilgungshilfen entlasten den Haushalt der Empfänger, da sich der Umfang der zu zahlenden Annuitäten reduziert. Die Schulden verbleiben jedoch bei den Empfängern. Eine weitere Form ist die prozentual anteilige oder vollständige Übernahme kommunaler Schulden durch das Land<sup>6</sup>, wobei diese Schuldenhilfe auf bestimmte Schuldenarten begrenzt sein kann. Damit geht eine Übertragung von Tilgung, Zins sowie Zahlungsausfall- und Refinanzierungsrisiken einher. Die Schuldenübernahme durch das Land kann jedoch Auflagen vorsehen: Zum Beispiel können Kommunen verpflichtet sein, Annuitäten an das Land zu entrichten, um sie an der Abtragung der Schulden zu beteiligen. Die Kreditübernahme dient auch dem Ziel des Haushaltsausgleichs, da sie durch verbesserte Refinanzierungskonditionen die laufende Last aus Zins und Tilgung mindert.

In den meisten Fällen nutzten die untersuchten Programme ausschließlich eines der Instrumente. Allerdings ist auch eine Kombination möglich. Beispielsweise verknüpfte Nordrhein-Westfalen 1991 die Vergabe von Haushaltskonsolidierungshilfen vereinzelt mit einer einmaligen Tilgungshilfe. Das hessische Entschuldungsprogramm von 2011 ergänzte die partielle Übernahme kommunaler Altschulden mit der Gewährung von Zinsdiensthilfen.

Obwohl die meisten Länder nach der Finanzkrise mit eskalierenden kommunalen Haushaltskrisen kämpften, fielen die Reaktionen unterschiedlich aus. Einige konzentrierten sich auf die Stabilisierung der Kommunalhaushalte und gewährten Zuweisungen zum Haushaltsausgleich (Mecklenburg-Vorpommern, Nordrhein-Westfalen, Schleswig-Holstein). Andere fokussierten auf den Schuldenabbau und vergaben Tilgungshilfen oder übernahmen einen Teil der Kredite (Hessen, Niedersachsen, Rheinland-Pfalz, Saarland, Sachsen-Anhalt). Allen gemein war die Sorge, auch in Zeiten einer kommunalen Haushaltskrise eine kontinuierliche Leistungserbringung durch die Kommunen sicherzustellen und Zweifel an der Solvenz der Kommunen zu zerstreuen.

Die unterschiedliche Ausgestaltung der Hilfskonzepte lässt sich im Wesentlichen auf eine Kombination dreier Erklärungsfaktoren zurückführen: Fiskalischer Problemdruck in Form persistenter Haushaltsdefizite und ansteigender Schuldenstände, wahrgenommener Handlungsbedarf durch das Land und Handlungsspielräume des Landeshaushalts. Anhand der Höhe kommunaler Kassenkredite (Stand 2010) als Krisenindikator wird deutlich, dass viele Bundesländer hohe, stark ansteigende Kassenkredite aufwiesen. Kommunen sahen sich ernsthaften Zinsänderungsrisiken, Risiken der Anschlussfinanzierung und vereinzelt sogar der Zahlungsunfähigkeit ausgesetzt, woraus die Länder in unterschiedlichem Maße Handlungsbedarf ableiteten.

In Mecklenburg-Vorpommern, Sachsen-Anhalt und Schleswig-Holstein lagen die Pro-Kopf-Kassenkredite unter dem Bundesdurchschnitt (Boettcher & Junkernheinrich, 2011, S. 281). Der kurzfristige Problemdruck war weniger ausgeprägt. Daher konzentrierten sich diese Länder auf den strukturellen Haushaltsausgleich, um zu verhindern, dass Kassenkredite zu einem ernsthaften Problem werden (Mecklenburg-Vorpommern, Schleswig-Holstein), oder auf die Reduzierung langfristiger Verbindlichkeiten (Sachsen-Anhalt). Die anderen fünf Länder sahen sich mit erheblichen Kassenkrediten konfrontiert. Daher adressierten sie (ausgenommen Nordrhein-Westfalen) die Kassenkreditproblematik und die damit verbundenen Risiken direkt durch Maßnahmen zum Schuldenabbau. Diese Strategie war tragfähig, da das Gesamtvolumen der Kas-

senkredite in Relation zu den Kapazitäten des Landeshaushaltes noch überschaubar war. In Nordrhein-Westfalen war die Situation anders. Die Kommunen wiesen sowohl hohe Kassenkreditbestände als auch massive strukturelle Haushaltsdefizite auf. In dieser Situation war das Ziel Schuldenabbau zwecklos. Darüber hinaus hätte eine Übernahme der Kassenkredite den Landeshaushalt überfordert. Folglich konzentrierte man sich primär auf die strukturelle Haushaltskonsolidierung als ersten Schritt, um den weiteren Anstieg der Kassenkredite zu stoppen. Nordrhein-Westfalen illustriert somit einen zentralen, von Kritikern häufig vorgebrachten Kritikpunkt an Schuldenhilfen (siehe Kapitel 3): Bei einer bestimmten Schwelle der kommunalen Finanzprobleme können die Kosten einer umfassenden Entschuldung prohibitiv hoch werden und sich die Maßnahme als kaum durchsetzbar erweisen. Zwar wäre grundsätzlich anstelle einer vollständigen auch eine anteilige Entschuldung denkbar gewesen. Allerdings hätte dies nur einen „Tropfen auf den heißen Stein“ bedeutet und wäre wenig effektiv gewesen.

Des Weiteren lässt sich beobachten, dass im Regelfall in keinem der Programme die Schuldenhilfen das gesamte Haushaltsdefizit oder die gesamte Verschuldung abdeckten. Unabhängig von der Leistungsfähigkeit des Landeshaushaltes wären solch umfassende Schritte politisch kaum vertretbar und nicht anreizadäquat gewesen. Die formal einzige Ausnahme ist die „Hessenkasse“. Diese löste die bestehenden Kassenkredite vollständig ab, was der Praxis aller übrigen Programme widerspricht. Gleichwohl bedeutet diese Umschuldung materiell nur eine Teilentlastung, denn die Kommunen tragen über die Zahlung fester Annuitäten einen Teil ihrer Entschuldung selbst.

Darüber hinaus wurden Schuldenhilfen stets unter Auflagen, z. B. eigene Konsolidierungsmaßnahmen, vergeben und die Länder legten großen Wert auf deren Durchsetzung. Dabei müssen Auflagen nicht nur fiskalischer Natur sein. Beispielsweise nutzte Niedersachsen die Schuldenhilfen, um das Problem fragmentierter Kommunalstrukturen anzugehen und machte Gemeindegebietsreformen zur Bedingung. Brandenburg beabsichtige dies 2018 ebenfalls, scheiterte allerdings an deren Umsetzung. Sowohl die lediglich anteilige Übernahme kommunaler Verbindlichkeiten als auch die Kopplung der Schuldenhilfen an strikte Konsolidierungserfordernisse deutet an, dass sich die landespolitischen Entscheidungsträger bei der Ausgestaltung der Hilfsmaßnahmen der in der einschlägigen Literatur diskutierten Risiken und Anreizproblematiken bewusst waren und diese zu vermeiden suchten.

## 5.2 Implementationszeitpunkte der Schuldenhilfen

Die Historie kommunaler Entschuldungsmaßnahmen zeigt, dass divergierende Formen von Schuldenhilfen zu unterschiedlichen Zeitpunkten implementiert wurden. Dies wirft die Frage auf, warum manche Länder zu diesen Zeitpunkten in dieser Form reagierten, andere jedoch nicht. Wie Kapitel 4 verdeutlicht, waren Schuldenhilfen zwischen 1980 und 2009 seltene, lokalspezifische Ereignisse. In den 1980ern waren kommunale Haushaltskrisen kein bundesweites Massenphänomen, sondern konzentrierten sich auf strukturschwache Industriestädte in Nordrhein-Westfalen. Folglich entwickelte nur dieses Land ein spezifisches Hilfsprogramm, dessen Dauer, Umfang und Teilnehmerzahl begrenzt war. Nachdem die dringendsten Probleme behoben waren, wurde das Programm beendet.

Zwischen 1991 und 2009 verschärften sich kommunale Haushaltsprobleme bundesweit. Viele Länder verzeichneten steigende Zahlen von Kommunen mit Haushalts-

schieflagen. Trotzdem wurde nicht über Umschuldungen diskutiert. Anstatt mit neuen Arten von Schuldenhilfen zu experimentieren, verfolgten die Länder alternative Strategien. Einerseits adressierten sie lokale Haushaltskrisen durch die Vergabe spezieller Bedarfszuweisungen innerhalb des KFA. Andererseits änderten sie den haushalts- und aufsichtsrechtlichen Rahmen, um Haushaltsdisziplin zu fördern (Geißler, 2009).

Lediglich Bayern und einige ostdeutsche Länder legten in diesem Zeitraum begrenzte Hilfsprogramme auf. Während die ostdeutschen Länder primär Finanzprobleme kommunaler Auslagerungen lösen und eine Ansteckung der Kommunalhaushalte verhindern wollten, zielte das bayerische Hilfsprogramm auf eine Fortentwicklung des KFA. Es wurde als Pilotprojekt zur Entwicklung einer neuen Art von Bedarfszuweisungen konzipiert. Diese sollten gezielt strukturschwache Gemeinden unterstützen, eine offensichtliche Anomalie angesichts der insgesamt soliden Haushaltslage der bayerischen Kommunen.

Während Entschuldungsmaßnahmen bis 2009 eine Seltenheit darstellten, kann man nach der globalen Finanzkrise eine flächendeckende Verbreitung kommunaler Entschuldungsprogramme in den meisten Bundesländern beobachten. Dahinter steckte die Sorge, dass Kommunen aufgrund ihrer erheblichen Haushaltsprobleme bald nicht mehr in der Lage sein würden, die ihnen zugewiesenen Aufgaben zu erfüllen und eine kontinuierliche Bereitstellung kommunaler Leistungen zu gewährleisten. Dieser Paradigmenwechsel, der sich in den Entschuldungsprogrammen der „ersten Generation“ manifestiert, lässt sich auf mehrere Erklärungsfaktoren zurückführen. Erstens änderten sich der gesellschaftliche Diskurs und normative Grundpositionen gegenüber der Staatsverschuldung. Wurde die Kreditaufnahme durch die öffentliche Hand jahrzehntelang als unproblematisch wahrgenommen, wuchs zunehmend das Bewusstsein über deren Schattenseiten. Die wachsende Verschuldung aller föderalen Ebenen und die Häufung von Haushaltskrisen verdeutlichten die Notwendigkeit der Haushaltskonsolidierung (Schuppert & Rossi, 2006, S. 5). Diskussionen über Vor- und Nachteile von Staatsverschuldung intensivierten sich während der Weltfinanzkrise (Holtfrerich et al., 2015). Über die Parteigrenzen hinweg entstand ein Konsens über die Notwendigkeit einer Begrenzung der Staatsverschuldung (Deutscher Bundestag, 2006, S. 7393 ff.). Die kritischere Haltung gegenüber öffentlicher Verschuldung beschränkte sich nicht auf Deutschland, sondern war ein paneuropäisches Phänomen.

Zweitens wuchsen im Bankensektor Zweifel an der Kreditwürdigkeit der Kommunen. In der Vergangenheit sahen Banken Kommunalkredite aufgrund des impliziten Haftungsverbunds zwischen Ländern und Kommunen als nicht ausfallgefährdet an (Schwaring, 2014, S. 150). Der starke Anstieg der Kassenkredite und spektakuläre Kommunalinsolvenzen im Ausland führten jedoch zu schlechteren Refinanzierungskonditionen (Brand, 2013, S. 88). Viele Banken entwickelten interne Ratingverfahren für kommunale Schuldner. Einige Banken begrenzten die Kreditvergabe, bildeten Risikokriterien oder stellten das margenarme Geschäft ein (Osman, 2011; Elbers, 2011). Folglich befürchteten Landespolitiker eine Kreditklemme (Landtag Nordrhein-Westfalen, 2011).

Drittens nahm der fiskalische Problemdruck angesichts permanenter Haushaltsdefizite und eskalierender Kassenkredite in den 2000ern erheblich zu. Weder die Gewährung von Bedarfszuweisungen noch die Intensivierung der Kommunalaufsicht stellten eine tragfähige Problemlösung dar. Die Strategie des *Muddling through* war offensichtlich gescheitert. Daher beschloss eine bemerkenswerte Anzahl von Ländern neue, in diesem Volumen bisher unbekannte Entschuldungsmaßnahmen.

Interessanterweise lassen sich weder in Bezug auf Zeitpunkt noch Umfang und Ausgestaltung der Hilfsprogramme parteipolitische Effekte identifizieren. Sowohl linke als auch konservative oder lagerübergreifende Koalitionsregierungen beschlossen Entschuldungsprogramme, was einen allgemeinen, parteiübergreifenden Konsens widerspiegelt. Darüber hinaus deutet die Auswertung der Plenarprotokolle auf eine begrenzte Politikdiffusion hin, da die Programme nicht simultan, sondern mit zeitlicher Verzögerung entwickelt wurden und in den landespolitischen Diskussionen auf Entwicklungen in anderen Bundesländern verwiesen wurde.

Während in manchen Ländern die Debatten über Entschuldungsmaßnahmen ausliefen, gewannen sie in anderen eine neue Dynamik und führten zur Entwicklung einer „zweiten Generation“ kommunaler Entschuldungsprogramme. Hierfür existieren verschiedene Erklärungen. Erstens erkannten einige Länder, dass die früheren Programme nur ein erster Schritt sein konnten. Der Umfang dieser Programme war zwangsläufig begrenzt, da die Länder selbst von der Weltfinanzkrise betroffen waren und durch die nahende Schuldenbremse unter erheblichem Konsolidierungsdruck standen. Aufgrund der günstigen makroökonomischen Rahmenbedingungen der 2010er Jahre verbesserten sich die Länderhaushalte jedoch deutlich und generierten hohe Überschüsse. Zweitens verspricht die aktuelle Reform des bundesstaatlichen Finanzausgleichs allen Ländern zusätzliche Einnahmen. Beide Entwicklungen erweitern den Handlungsspielraum der Landeshaushalte, den einige Länder für neue Entschuldungsprogramme nutzten (z. B. Hessen, Saarland, Mecklenburg-Vorpommern).

Drittens beschäftigten sich die Länder zunehmend mit den Zinsänderungsrisiken kommunaler Kassenkredite. Angesichts der Niedrigzinspolitik der Europäischen Zentralbank (EZB) schien es angebracht, kurzfristig laufende Kassenkredite in langfristige Kredite umzuwandeln, um Zinsrisiken auszuschalten. Hierzu schuldete Hessen die verbleibenden Kassenkredite vollständig in langfristige Kredite um, verpflichtete die Kommunen jedoch zur Zahlung von Annuitäten an das Land. Auch das Saarland übernimmt ab 2020 50% der Kassenkredite, um Zinsänderungsrisiken zu eliminieren. Im Gegenzug müssen die Kommunen die verbleibenden Kassenkredite ebenfalls abbauen. Der finanzielle Spielraum und damit auch der Implementationszeitpunkt des saarländischen Programms ergibt sich aus den neuen, ab 2020 fließenden Bundeszuweisungen für finanzschwache Kommunen.

Rheinland-Pfalz ergänzte sein bestehendes Hilfsprogramm über Zinszuschüsse und als Prämie für eigene Konsolidierungsanstrengungen konzipierte Zuweisungen zum Schuldenabbau. Allerdings zeichnen sich diese neuerlichen Schuldenhilfen durch ihren begrenzten Umfang aus, insbesondere im Vergleich zum Saarland, welches sich offiziell sogar in einer Haushaltskrise befindet und Konsolidierungshilfen des Bundes erhält. Rheinland-Pfalz „leidet“ seit Jahren unter einem umstrittenen KFA, der mehrfach vor dem Verfassungsgerichtshof des Landes verhandelt wurde und ein ständiges „Damoklesschwert“ für den Landeshaushalt darstellt. Daher ist nicht planbar, welche Finanzmittel langfristig für Entschuldungsmaßnahmen zur Verfügung stehen.

Eine besondere Situation findet sich in Schleswig-Holstein. Aus Sicht des Landes erreichte das erste Hilfsprogramm seine Ziele, mit Ausnahme der vier kreisfreien Städte. Daher konzentriert sich das Nachfolgeprogramm ausschließlich auf diese Städte und fällt folgerichtig klein aus. Auch der Fall Brandenburg hat einen speziellen Hintergrund. Eigentlich wollte das Land drei kreisfreie Städte zu einer Fusion mit den benachbarten Kreisen bewegen. Als Anreiz bot es Tilgungshilfen für Kassenkredite. Der

Fusionsplan scheiterte, aber das Angebot der Schuldenhilfen war nicht mehr zurückzunehmen.

Schließlich begannen auch im Ursprungsland der kommunalen Haushaltskrise, Nordrhein-Westfalen, Diskussionen über ein weiteres Hilfsprogramm. Dieses soll die Kassenkreditbestände spürbar reduzieren. Einer großen Lösung steht allerdings das hohe Niveau kommunaler Kassenkredite (23 Mrd. €) entgegen, welches den Landeshaushalt überlasten würde. Aus diesem und aus polittaktischen Gründen (Warten auf das 2019 angekündigte Hilfsprogramm des Bundes) wurde bis dato kein neues Hilfsprogramm aufgelegt.

### 5.3 Auswirkungen kommunaler Schuldenhilfen

Vor dem Hintergrund der Ausbreitung kommunaler Entschuldungsprogramme im Nachgang der globalen Finanzkrise entspann sich eine lebhaft Diskussions über deren Auswirkungen. Allerdings ist die Evaluation der Programmeffekte schwierig, da nicht nur die Programme, sondern auch die allgemeine Konjunkturlage und zusätzliche Bundeshilfen die kommunale Haushaltslage beeinflussten. Außerdem muss man bei der Beurteilung den Blick über fiskalische Kernindikatoren hinaus weiten und auch (un)beabsichtigte Nebeneffekte berücksichtigen (Boettcher, Feier, Geißler & Niemann, 2018).

Grundsätzlich führten die Programme zu einer strengen Sparpolitik. Der Zwang zu einer rigorosen Austeritätspolitik, dem Programmkommunen unterworfen waren, sollte auch einen Abschreckungseffekt entfalten, damit diese zukünftig eine sorgsamere Haushaltspolitik betreiben und sich nicht leichtfertig erneut in eine haushälterische Bredouille bringen. Dadurch wollte man die im theoretischen Diskurs befürchteten negativen Anzeizeffekte von Schuldenhilfen vermeiden. Im Hinblick auf die unmittelbar nach der Finanzkrise implementierten Programme erreichten die lokalen Konsolidierungsmaßnahmen ein Volumen von mehreren Milliarden Euro, die sich gleichmäßig auf einnahme- und ausgabeseitige Maßnahmen aufteilen. Am stärksten fielen die Konsolidierungsbeiträge in Nordrhein-Westfalen aus, gefolgt von Hessen (Arnold, Boettcher, Freier, Geißler & Holler, 2015, S. 132-133).

Ein Vergleich der finanziellen Entwicklung zwischen Teilnehmern und Nicht-Teilnehmern (2008-2015) zeigt gemischte Befunde. Zwar konnten die Teilnehmer den weiteren Anstieg der Kassenkredite stoppen und Haushaltsdefizite abbauen (Boettcher, Freier, Geißler & Niemann, 2018, S. 595-596). Die Schuldenhilfen hatten jedoch auch Nebenwirkungen. Sie konnten die Entwicklungsdynamik der Sozialausgaben nicht bremsen, welche eine zentrale Ursache kommunaler Haushaltsdefizite darstellt. Zudem betrafen die Sparmaßnahmen überproportional die Investitionsausgaben. Schon vor den Hilfsprogrammen wiesen die Programmkommunen einen erheblichen Rückstand bei den Investitionen auf, der sich im Zuge der Programme vergrößerte. Ähnliche Diskrepanzen sind für die Steuerpolitik zu beobachten. Vor den Hilfsprogrammen lagen die Hebesätze der Grundsteuer B in Teilnehmer- und Nichtteilnehmerkommunen auf ähnlichem Niveau. Im Verlauf der Hilfsprogramme entwickelten sie sich zunehmend auseinander (Rappen, 2017, S. 27 f.). Ein analoger Trend ist bei der Gewerbesteuer nicht zu beobachten, da die Programmkommunen offenbar versuchen, weitere Verluste an Wettbewerbsfähigkeit zu vermeiden. Dennoch beeinflussen die unterschiedlichen Trends



bei Investitionsausgaben und Grundsteuerbelastung die langfristigen Entwicklungsperspektiven der Teilnehmerkommunen.

Obwohl die Schuldenhilfen kurzfristig die Haushaltskonsolidierung erleichtern, ist die Einhaltung der langfristigen Verpflichtungen fraglich. Es bleibt offen, ob die Kommunen den Konsolidierungskurs auch bei veränderten makroökonomischen Rahmenbedingungen strikt umsetzen werden. Daher argumentieren Kritiker, dass diese Schuldenhilfen echte Lösungen nur verschieben. Erforderlich sei vielmehr eine umfassende Reform der Kommunalfinanzen, welche die Finanzierung der Sozialausgaben, eine Neuzuteilung der Steuerertragskompetenzen sowie das kommunale Altschuldenproblem adressiert (Stolzenberg & Heinelt, 2013, S. 474; Bogumil, Holtkamp, Junkernheinrich & Wagschal, 2014).

Über die fiskalischen Effekte hinaus zeitigten die Entschuldungsprogramme auch Auswirkungen auf die Kommunalaufsicht. Deren ineffektive Umsetzung wurde wiederholt als ein Erklärungsfaktor für lokale Haushaltskrisen identifiziert (Herrmann, 2011). Einige Länder waren sich dieser Problematik bewusst und nutzten die Hilfsprogramme, um ihre Aufsichtssysteme in Bezug auf Aufsichtsprozesse, Regulierungsrahmen und Instrumentarium zu ändern. In der Regel einigten sich Länder und Kommunen vertraglich auf Umfang und Auflagen der Schuldenhilfen, wobei die Kommunalaufsicht beratend involviert war. Die Überwachung der Gemeindehaushalte wurde intensiviert, die Durchsetzungsinstrumente verstärkt. Eine grundlegende Änderung war die Übertragung der Finanzaufsicht über kreisangehörige Gemeinden von der Kreis- auf die Bezirksebene, um die Homogenität der Rechtsanwendung zu gewährleisten. Alles in allem wurden die Änderungen der Aufsichtsprozesse von den Kämmereien der Programmkommunen größtenteils begrüßt (Ebinger, Geißler, Niemann, Person & Zabler, 2017; Person & Zabler, 2017).

Neben den fiskalischen und regulatorischen Effekten wirkten sich die Programme auch auf die Kommunalpolitik aus. Da kommunale Haushaltsdefizite in beachtlichem Umfang durch exogene Einflussfaktoren bedingt sind, fühlten sich Lokalpolitiker lange Zeit hilflos (Bogumil, Holtkamp, Junkernheinrich & Wagschal, 2014, S. 615). Angesichts eines Teufelskreises aus strukturellen Budgetdefiziten, hohen Schuldenständen und wirtschaftlicher Stagnation verpufften viele Konsolidierungsmaßnahmen weitgehend (Holtkamp, 2010). Die Schuldenhilfen der Länder halfen, diese „Vergeblichkeitsfalle“ zu überwinden, und ließen einen Haushaltsausgleich erstmals seit Jahren wieder realistisch erscheinen. Der damit verbundene Motivationsschub erbrachte bisher undenkbbare Konsolidierungsbemühungen (Arnold, Boettcher, Freier, Geißler & Holler, 2015, S. 124).

Allerdings haben die Schuldenhilfen ihren Preis: Sie gehen mit starken Einschränkungen lokaler Autonomie einher. Die Kommunen verpflichteten sich zu einer langfristigen Haushaltskonsolidierung, die den zukünftigen Handlungsspielraum einschränkt. Dies stellt erhebliche Belastungen für nachfolgende Generationen lokaler Akteure dar, die weder für die Haushaltskrise verantwortlich waren noch die Konsolidierungsmaßnahmen beeinflussen konnten (Stolzenberg & Heinelt, 2013, S. 473).

Des Weiteren resultierten die Konsolidierungsmaßnahmen oft aus intransparenten Verhandlungen zwischen Kommunalaufsicht und Kommunen. Die Aufsichtsbehörden formulierten Hinweise, welche Maßnahmen sie erwarteten. Aus Sicht der Wähler war unklar, wer letztlich die inhaltliche Verantwortung trägt (Stolzenberg & Heinelt, 2013, S. 473), was die demokratische Verantwortlichkeit und Legitimation untergräbt.

Außerdem wurde die Überwachung der Gemeindehaushalte intensiviert. Bei Nichteinhaltung der Konsolidierungspläne drohen drakonische Strafen. Im Extremfall kann die Landesregierung einen Staatskommissar ernennen, der einzelne Gemeindeorgane temporär ersetzt und die vom Land geforderten Konsolidierungsmaßnahmen durchführt, auch gegen lokale Widerstände. Kritiker argumentieren, dass dieses Vorgehen einen dramatischen Eingriff in die Selbstverwaltung darstellt, der den normativen Mindeststandards lokaler Demokratie widerspricht, da Staatsbeauftragte der örtlichen Bevölkerung nicht rechenschaftspflichtig sind und ihr Verhalten nicht sanktioniert werden kann (Holtkamp & Fuhrmann, 2014).

## 6 Fazit

Seit Jahrzehnten sind viele Kommunen Deutschlands mit erheblichen Haushaltsproblemen konfrontiert, welche die Qualität kommunaler Leistungserbringung erheblich beeinträchtigen. Zur Überwindung dieser Probleme nutzten alle Bundesländer diverse Formen von Schuldenhilfen, die sich jedoch in Zielsetzung, Instrumentarium und Reichweite unterscheiden. Die Erfahrungen der Bundesländer zeigen, dass Bedarfszuweisungen zur Unterstützung des Haushaltsausgleichs hilfreich sind, um kleine, auf wenige Gemeinden beschränkte Haushaltsprobleme anzugehen. Sie sind ein notwendiger Bestandteil des KFA, da stets Gemeinden mit besonderen fiskalischen Herausforderungen existieren. Im Falle flächendeckender Haushaltskrisen stoßen sie jedoch an ihre Grenzen. Außergewöhnliche Krisen erfordern spezifische, temporäre und umfangreiche Entschuldungsprogramme.

Die Mehrzahl der Bundesländer hat im Nachgang der Finanzkrise solche Entschuldungsprogramme aufgelegt. Allerdings differiert deren Ausgestaltung erheblich. Vor diesem Hintergrund untersucht der Artikel erstmals die Geschichte der kommunalen Schuldenhilfen in Deutschland. Um die empirische Vielfalt zu ordnen und den oftmals diffus genutzten Begriff der Schuldenhilfe stringenter in seinen verschiedenen Facetten zu fassen, wurde daher im Rahmen dieses Beitrages induktiv eine Kategorisierung kommunaler Schuldenhilfen erarbeitet. Hinsichtlich ihrer grundlegenden Ziele konzentrieren sich diese Hilfsmaßnahmen entweder auf die Stabilisierung der Haushalte oder den Schuldenabbau. Um diese Zielsetzungen zu erreichen, kommen unterschiedlichste Instrumente zum Einsatz, z. B. Konsolidierungshilfen, Zinszuschüsse, Tilgungshilfen oder die Übertragung von Schulden auf das Land. Die Instrumente können dabei den zu Grunde liegenden Krisenindikator vollständig oder anteilig umfassen.

Eine vollständige Deckung von Budgetdefiziten oder vollständige Übernahme von Schulden ist jedoch nie vorgekommen. Zu groß wäre das Risiko von *moral hazard* und von Verstimmungen innerhalb der kommunalen Familie. Aus demselben Grund fordern Schuldenhilfen stets harte Sparmaßnahmen und gehen üblicherweise mit einer Intensivierung der Finanzaufsicht einher.

Für den Implementationszeitpunkt und die Ausgestaltung von Schuldenhilfen sind mehrere Einflussfaktoren relevant: Die normative Grundposition gegenüber öffentlicher Verschuldung und damit die Problemwahrnehmung, der fiskalische Problemdruck, der die Länder zu Maßnahmen zwingt, sowie die Spielräume der Länderhaushalte. Typische Krisenindikatoren sind ausufernde Haushaltsdefizite, eskalierende Schuldenstände, erhebliche Zinsänderungsrisiken und drohende Kreditklemmen. Ab

welchem Zeitpunkt und Ausmaß Budgetdefizite und Schuldenstände aus Sicht der Entscheidungsträger als problematisch wahrgenommen werden und sich entsprechender Handlungsbedarf ergibt, ist jedoch in hohem Maße kontextabhängig und stark subjektiv. Eine fiskalische Situation, die in einem Bundesland schon als dringend handlungsbedürftig angesehen wird, kann in einem anderen Land noch als akzeptabel betrachtet werden. Auch die Diffusion politischer Ideen zwischen den Ländern hat einen Effekt. Frühe Aktivitäten einzelner Vorreiter setzten andere Länder unter Handlungsdruck, da sich das Spektrum potentieller Reaktionsmöglichkeiten weitete und Erwartungshaltungen bei den Kommunen geweckt wurden.

Die Auswahl eines konkreten Instruments hängt primär von der Art der vorliegenden kommunalen Haushaltskrise und dem abgeleiteten Handlungsbedarf ab. Wenn das Problem in Haushaltsdefiziten, nicht jedoch in gefährlich hohen Kassenkrediten, besteht, ist die Vergabe von Konsolidierungshilfen sinnvoll. Dadurch werden Kommunen darin unterstützt, den Haushaltsausgleich wieder zu erreichen und eigenständig ihren Kassenkreditbestand zu stabilisieren bzw. abzubauen. Haben die Kassenkreditbestände hingegen ein erhebliches Volumen erreicht, so dass schon geringe Zinserhöhungen erhebliche Mehrbelastungen hervorrufen, kann es erforderlich werden, die Kassenkreditproblematik direkt durch die Vergabe von Tilgungshilfen oder eine Schuldenübernahme zu mindern. Die Umsetzbarkeit entsprechender Maßnahmen zum Schuldenabbau wird jedoch durch die fiskalischen Kapazitäten des Landeshaushaltes determiniert. Die Vergabe von Zinshilfen erfolgt häufig als flankierendes Instrument. Dadurch will das Land die Kommunen bei den im Rahmen von Tilgungshilfen oder Teilentschuldungen geforderten Haushaltskonsolidierungsprozessen unterstützen.

Hinsichtlich der Wirkungen von Schuldenhilfen zeigt die Analyse eine gemischte Bilanz. Obwohl die Entschuldungsprogramme zu spürbaren Verbesserungen bei wichtigen Haushaltskennziffern führten, zeigten sich auch negative fiskalische Effekte wie sinkende Investitionen und steigende Steuerbelastungen, welche die Wettbewerbsfähigkeit der betroffenen Kommunen erheblich einschränken. Außerdem konnte die Dynamik der Sozialausgaben als fundamentaler Treiber kommunaler Haushaltsprobleme nicht ausgeschaltet werden. Die mit den Schuldenhilfen einhergehende Intensivierung der Kommunalaufsicht veränderte zudem die Beziehungen zwischen Ländern und Kommunen, da die Vergabe von Schuldenhilfen mit erheblichen Eingriffen in die kommunale Autonomie einherging.

Alles in allem bieten sowohl die aus der Empirie induktiv abgeleitete Kategorisierung des oft uneinheitlich verwendeten Terminus Schuldenhilfe als auch die Darstellung zentraler Erklärungsfaktoren für die divergierende Ausgestaltung und Implementationszeitpunkte dieser Maßnahmen einen Mehrwert für die wissenschaftliche Diskussion. Aus Sicht der Praxis, insbesondere der Länder, erlaubt der historische Überblick die Identifikation unterschiedlicher Reaktionsmuster und gibt Hinweise auf erfolgversprechende Varianten in der Ausgestaltung von Schuldenhilfen je nach Problemlage.

## Anmerkungen

- 1 Am 10. Juli 2019 beschloss die Bundesregierung im Rahmen der Kommission zur Schaffung gleichwertiger Lebensverhältnisse eine anteilige Übernahme kommunaler Kassenkredite.
- 2 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Die weibliche Form ist selbstverständlich jeweils eingeschlossen.

- 3 Kommunalkredite lassen sich in Kassenkredite und Investitionskredite unterteilen. Während Investitionskredite langfristiger Natur sind und der Ausgabendeckung (im investiven Bereich) dienen, handelt es sich bei Kassenkrediten um kurzfristige Verbindlichkeiten. Diese waren ursprünglich zur Überbrückung temporärer Liquiditätsengpässe und zur Aufrechterhaltung einer ordnungsgemäßen Kassenwirtschaft gedacht. Allerdings stellen sie vielerorts längst dauerhafte Finanzierungsquellen dar (Bogumil, Holtkamp, Junkernheinrich & Wagschal, 2014, S. 620; siehe Schwarting, 2019 für eine weitergehende Ausdifferenzierung). Der Fokus der nachfolgenden Ausführungen liegt auf der Kassenkreditproblematik, da diese in weiten Teilen der Wissenschaft und der Praxis als zentraler Problemindikator angesehen wird. Auch die analysierten Entschuldungsprogramme konzentrierten sich fast ausschließlich auf diesen Typus des Kommunalkredits, weshalb ein entsprechender Fokus als sinnvoll erachtet wird.
- 4 Eine Diskussion der Determinanten kommunaler Kassenkredite bieten Junkernheinrich & Wagschal, 2014.
- 5 Diese Definition umfasst nicht Finanzhilfen zur Adressierung fiskalischer Ungleichgewichte, die auf exogenen Schocks beruhen und somit vom Empfänger nicht zu verantworten sind.
- 6 Im Hinblick auf den Umfang der Schuldenübernahme bedeutet „anteilig“, dass ein bestimmter Prozentsatz einer Schuldenart (Kassenkredite, Investitionskredite, Gesamtschulden) auf das Land übertragen wird (nicht die Begrenzung der Schuldenübernahme auf eine bestimmte Kreditart).

## Literatur

- Arnold, Felix, Boettcher, Florian, Freier, Ronny, Geißler, René & Holler, Benjamin (2015). *Kommunaler Finanzreport 2015*. Gütersloh: Bertelsmann-Stiftung.
- Allers, Maarten (2015). The Dutch local government bailout puzzle. *Public Administration*, 93 (2), 451-470. <https://doi.org/10.1111/padm.12123>
- Baskaran, Thushyanthan (2017). Local fiscal Policy after a Bailout: Austerity or soft Budget Constraints? *Economics of Governance*, 18 (3), 209-238. <https://doi.org/10.1007/s10101-017-0194-8>.
- Bogumil, Jörg, Holtkamp, Lars, Junkernheinrich, Martin & Wagschal, Uwe (2014). Ursachen kommunaler Haushaltsdefizite. *Politische Vierteljahresschrift*, 55 (4), 614-647.
- Boettcher, Florian & Junkernheinrich, Martin (2011). Kommunalfinanzen im Jahr 2010: Krisenverschärfung und Disparitätenzunahme. In Martin Junkernheinrich, Stefan Koriath, Thomas Lenk, Henrik Scheller & Matthias Woisin (Hrsg.). *Jahrbuch für öffentliche Finanzen 2011* (271-292). Berlin: BWV.
- Boettcher, Florian & Geißler, René (2017). Die Nettobelastung der Kommunen aus Sozialausgaben. *Analysen und Konzepte*, 2/2017.
- Boettcher, Florian, Freier, Ronny, Geißler, René & Niemann, Friederike-Sophie (2018). Konsolidierungsprogramme der Länder für finanzschwache Kommunen. *Wirtschaftsdienst*, 98 (7), 592-599.
- Boettcher, Florian, Freier, Ronny, Geißler, René, Schubert, Moritz & Stollhoff, Rainer (2019). *Kommunaler Finanzreport 2019*. Gütersloh: Bertelsmann-Stiftung
- Brand, Stephan (2013). Die Kommunalinsolvenz als Folge der kommunalen Finanznot? *Wirtschaftsdienst*, 93 (2), 87-93.
- Canuto, Otaviano & Liu, Lili (2013). An Overview. In Dies. (Ed.). *Until debt do us part* (pp. 1-30). Washington: World Bank.
- DESTATIS (o. J.). Schuldenstatistik.
- Deutscher Bundestag (2006). *Plenarprotokoll 16/74*, vom 15.12.2006.
- Diemert, Dörte (2005). Das Haushaltssicherungskonzept. Verfassungs- und haushaltsrechtliche Grundlagen in NRW. Stuttgart: Kohlhammer.
- Ebinger, Falk, Geißler, René, Niemann, Friederike-Sophie, Person, Christian & Zabler, Steffen (2017). Die kommunale Finanzaufsicht. Strukturen, Rationalitäten und Umsetzung im Ländervergleich. *Analysen und Konzepte*, 1/2017.
- Ebinger, Falk, Zabler, Steffen, Person, Christian, Niemann, Friederike-Sophie & Geißler, René (2018). Wirksame Finanzaufsicht über die Kommunen. Dimensionen und Determinanten der Umsetzungspraxis. *dms – der moderne Staat*, 11 (1), 165-187.
- Elbers, Martin (2011). Wir erleben einen Paradigmenwechsel. *Der Neue Kämmerer*, 05/2011.

- Geißler, René (2009). *Kommunalrechtliche Steuerungsansätze der Haushaltskonsolidierung*. Potsdam: Universitätsverlag Potsdam.
- Geißler, René (2019). Germany. In René Geißler, Gerhard Hammerschmid & Christian Raffer (Eds.), *Local Public Finance in Europe. Country Reports* (pp. 102-110). Gütersloh: Bertelsmann-Stiftung.
- Held, Friedrich (1995). Zur Rolle der Länder bei der Bewältigung der kommunalen Finanzprobleme. In Birgit Frischmuth (Hrsg.), *Sparstrategien* (S. 58-68). Berlin: DifU.
- Held, Friedrich (1996). Steuerung kommunaler Aufgabenerfüllung durch Haushaltssicherungskonzepte und staatliche Genehmigungsvorbehalte bei Umlageerhebungen. In Hans-Günter Henneke (Hrsg.), *Steuerung der kommunalen Aufgabenerfüllung durch Finanz- und Haushaltsrecht* (S. 63-79). Stuttgart: Boorberg.
- Herrmann, Karoline (2011). Der Missbrauch der kommunalen Kassenkredite. *Wirtschaftsdienst*, 91 (10), 686-693.
- Holtfrerich, Carl, Feld, Lars, Heun, Werner, Illing, Gerhard, Kirchgässner, Gebhard, Kocka, Jürgen, Schularick, Moritz, Streeck, Wolfgang, Wagschal, Uwe, Walter, Stefanie & von Weizsäcker, Christian (2015). *Staatsschulden: Ursachen, Wirkungen und Grenzen*. Berlin: o. V.
- Holtkamp, Lars (2010). *Kommunale Haushaltspolitik bei leeren Kassen*. Berlin: Edition Sigma.
- Holtkamp, Lars & Fuhrmann, Thomas (2014). Kommunale Selbstverwaltung zwischen Steuerschraube und Staatskommissar. *Der Gemeindehaushalt*, 115 (7), 145-148.
- Inman, Robert (2003). Transfers and Bailouts. In Jonathan Rodden, Gunnar Eskeland & Jennie Litvack (Eds.), *Fiscal Decentralization and the Challenge of Hard Budget Constraints* (pp. 35-83). Cambridge, Massachusetts: MIT Press.
- Junkernheinrich, Martin & Wagschal, Uwe (2014). Kommunale Kassenkredite: Bedeutung, Verursachung und Lösungsoptionen. *Zeitschrift für Staats- und Europawissenschaften*, 12 (2-3), 300-334. <https://www.jstor.org/stable/24237355>.
- Kornai, Janos (1986). The Soft Budget Constraint. *Kyklos*, 39 (1), 3-30. <https://doi.org/10.1111/j.1467-6435.1986.tb01252.x>
- Landtag Bayern (2003). *Plenarprotokoll 15/6*, vom 27.11.2003.
- Landtag Bayern (2011). *Drucksache 16/7966*, vom 26.04.2011.
- Landtag Bayern (2018). *Drucksache 17/19457*, vom 05.02.2018.
- Landtag Brandenburg (1997). *Plenarprotokoll 2/51*, vom 22.01.1997.
- Landtag Mecklenburg-Vorpommern (1998). *Plenarprotokoll 3/7*, vom 16.12.1998.
- Landtag Nordrhein-Westfalen (2011). *Plenarprotokoll 15/44*, vom 19.10.2011.
- Landtag Sachsen (1997). *Plenarprotokoll 2/58*, vom 12.06.1997.
- Landtag Sachsen-Anhalt (1999) *Drucksache 3/2006*, vom 12.08.1999.
- Landtag Thüringen (1995). *Plenarprotokoll 2/16*, vom 16.06.1995.
- Osman, Yasmin (2011). KfW will Kredite für Kommunen begrenzen. *Handelsblatt vom 17.04.2011*.
- Person, Christian & Zabler, Steffen (2017). Veränderter Rahmen – veränderte Praxis? Auswirkungen kommunaler Entschuldungsfonds auf die Praxis der Finanzaufsicht. *Zeitschrift für Kommunal-finanzen*, 67 (1), 1-6.
- Person, Christian & Geißler, René (2020). State Supervision of Local Budgets: from Forbearance to No Concession. In Tomas Bergström, Jochen Franzke, Sabine Kuhlmann & Ellen Wayenberg (Eds.), *The Future of Local Self-Government*. London: palgrave macmillan. [im Erscheinen]
- Pettersson-Lidbom, Per (2010). Dynamic Commitment and the Soft Budget Constraint. *American Economic Journal: Economic Policy*, 2 (3), 154-179. DOI: 10.1257/pol.2.3.154.
- Rappen, Herrmann (2017). Stärkungspakt Stadtfinanzen. Weg aus der Schuldenfalle oder gekaufte Zeit? *RWI Materialien*, Heft 120. <http://dx.doi.org/10.4419/86788829>
- Rodden, Jonathan (2002). The Dilemma of Fiscal Federalism: Grants and Fiscal Performance around the World. *American Journal of Political Science*, 46 (3), 670-687.
- Rodden, Jonathan, Eskeland, Gunnar & Litvack, Jennie (2003). Introduction and Overview. In Dies. (Eds.), *Fiscal Decentralization and the Challenge of Hard Budget Constraints* (pp. 3-31). Cambridge, Massachusetts: MIT Press.
- Schuppert, Gunnar & Rossi, Matthias (2006). Bausteine eines bundesstaatlichen Haushaltsnotlagen-regimes. *Hertie School of Governance – working papers*, 3/2006.

- Schwarting, Gunnar (2014). *Kommunales Kreditwesen* (4. Aufl.). Berlin: Erich Schmidt Verlag.
- Schwarting, Gunnar (2019). *Der kommunale Haushalt* (5. Aufl.). Berlin: Erich Schmidt Verlag.
- Seitz, Helmut (1999). Subnational government bailouts in Germany. *ZEI Working Paper B, 20/1999*.
- Stolzenberg, Philipp & Heinelt, Hubert (2013). Die Griechen von NRW. *Kommunale Rettungsschirme der Bundesländer. dms – der moderne Staat*, 6 (2), 463-484.
- Wegweiser Kommune (o. J.). *Wegweiser Kommune*. Online verfügbar unter: <https://www.wegweiser-kommune.de/>.
- Wildasin, David (1997). *Externalities and Bailouts: Hard and Soft Budget Constraints in Intergovernmental Fiscal Relations*. Unpublished manuscript. Vanderbilt University.
- Wildasin, David (2004). The Institutions of Federalism: Toward an Analytical Framework. *National Tax Journal*, 57 (2), 247-72.

### *Anschriften der Autoren:*

Dr. Christian Person, Arbeitsbereich Öffentliche Verwaltung, Public Policy, Institut für Politikwissenschaft, Technische Universität Darmstadt, Dolivostraße 15, D-64293 Darmstadt,

E-Mail: [christian.person@tu-darmstadt.de](mailto:christian.person@tu-darmstadt.de).

Dr. René Geißler, Programm LebensWerte Kommune, Bertelsmann Stiftung, Carl-Bertelsmann-Straße 256, D-33311 Gütersloh,

E-Mail: [rene.geissler@bertelsmann-stiftung.de](mailto:rene.geissler@bertelsmann-stiftung.de).

*Sandra Stötzer, Michaela Stangl, Christian Brödl*

# Auswirkungen von Gemeindestrukturreformen – Eine Einzelfallstudie einer Gemeindefusion in der österreichischen Steiermark<sup>1</sup>

## **Zusammenfassung**

Gemeindefusionen sind ein ungebrochen aktuelles Thema, das sowohl in der Praxis als auch in der Wissenschaft rege diskutiert wird. Der Stand der Forschung zeigt, dass die Befunde bezüglich der Auswirkungen von Gebietsreformen heterogen sind und die Reformaktivitäten in Österreich bislang kaum beleuchtet wurden. Dieser Beitrag bietet einen Überblick zu den jüngsten Strukturreformen in Österreich und spezifische Einblicke in eine steirische Fusionsgemeinde als Einzelfallstudie. Hierfür wurden einerseits vor der Fusion mittels Leitfadeninterviews die Erwartungen von Vertretern<sup>2</sup> aus Gemeindepolitik und -verwaltung eruiert. Andererseits wurden drei Jahre nach der Fusion (vollzogen per 1.1.2015) erneut Interviews mit Repräsentanten und Mitarbeitern dieser Gemeinde geführt, sodass den ex-ante artikulierten Erwartungen die ex-post wahrgenommenen Auswirkungen gegenübergestellt werden können.

*Schlagnworte:* Auswirkungen, Fallstudie, Gebietsreform, Gemeindefusion, Kommunalverwaltung

## **Abstract**

*Effects of structural reforms – A single case study of a municipal merger in Styria*

Municipal mergers continue to be a current topic actively discussed by both practitioners and scholars. Recent research shows heterogeneous findings concerning the effects of municipal amalgamations and that reform activities in Austria are only sparsely examined. This paper outlines recent structural reforms in Austria and provides an insight into a Styrian case study. Based on guided interviews, we first investigated the expectations of representatives of Styrian municipalities before their merger, which was implemented on 1.1.2015. Second, three years after the reform, we again conducted interviews with politicians and public servants of the merged municipality. Thus, expectations articulated ex-ante are contrasted with the ex-post perceived effects.

*Keywords:* case study, effects, municipal merger, municipal reform, local government

## **1 Einleitung**

Gebietsreformen als territoriale Umstrukturierung der (Kommunal-)Verwaltung sind kein neues, jedoch in Theorie und Praxis nach wie vor aktuelles Thema; sie zählen zu den zentralen verwaltungspolitischen Themen. Die Beweggründe für solche Strukturveränderungen inkludieren i. d. R. einen bestehenden oder erwarteten Problemdruck, etwa durch zunehmende und komplexer werdende Aufgaben, steigende Ansprüche der Bürger, Rekrutierungsprobleme (hinsichtlich qualifizierter Fachkräfte oder Bewerber für politische Mandate), demografische Veränderungen, Entsiedelung strukturschwacher Regionen und „Spardruck“ angesichts oft angespannter öffentlicher Haushalte und eingeschränkter fi-

nanzieller Spielräume für Investitionen (Kuhlmann, Seyfried & Siegel, 2018b, S. 13 f.; Pitlik, Wirth & Lehner, 2010, S. 1 f.; Seuberlich, 2017, S. 26 ff.; Tavares, 2018, p. 5). Ebenso vielfältig sind die damit verfolgten Ziele bzw. daran geknüpften Erwartungen, bspw. verbesserte Leistungsfähigkeit, Servicequalität und Standortattraktivität, erhöhte Handlungsspielräume der Kommunen sowie insbesondere Effizienzsteigerungen (Seuberlich, 2017, S. 29; Steiner, Kaiser & Eythórsson 2016, p. 25 ff.; Swianiewicz, 2018, p. 3).

Strukturreformen ziehen intendierte, aber auch unerwartete Effekte auf die Effizienz und Effektivität der kommunalen Leistungserbringung sowie die demokratische Teilhabe nach sich (Tavares, 2018, pp. 8 ff.). Der mittlerweile umfangreiche *State-of-the-Art* bezüglich der Auswirkungen von Gebietsreformen zeigt – sowohl im Ländervergleich als auch hinsichtlich einzelner Wirkungsdimensionen – nicht eindeutige bzw. teils widersprüchliche Befunde (Allers & Geertsema, 2016, pp. 661 ff.; Kuhlmann, Seyfried & Siegel, 2018a, S. 120 ff.; 2018b, S. 17 ff.; Swianiewicz, 2018, pp. 5 ff.), speziell in Bezug auf Einsparungen und Wirtschaftlichkeitsverbesserungen. Vergleichsweise positiv werden Einflüsse auf die Leistungs- und Handlungsfähigkeit von Kommunen eingeschätzt. Auffallend wenig beleuchtet wurden bislang Gebietsreformaktivitäten in Österreich.

Nach langem „Stillstand“ wurde in jüngerer Vergangenheit in Österreich – mit seiner vergleichsweise kleinstrukturierten Gemeindelandschaft – der Ruf nach Strukturreformen lauter. Seit 2011 intensivierte sich die Reformdiskussion und man setzte Gemeindefusionsprozesse um, primär in der Steiermark und fallweise in Oberösterreich (Heinisch, Lehner, Mühlböck & Schimpf, 2018, pp. 472 f.; Stangl & Stötzer, 2015, S. 89 ff.). Wenngleich die mit diesen Reformen verknüpften Erwartungen (aber auch Vorbehalte) hoch waren und sind, stehen begleitende sowie ex-post durchgeführte Reformevaluationen bisher aus. Die vorliegende Arbeit untersucht anhand einer Einzelfallstudie von Gemeinden, die im Zuge der Gemeindestrukturreform Steiermark 2015 freiwillig fusionierten, inwieweit sich die vor der Fusion geäußerten, subjektiven Erwartungen aus Sicht kommunaler Akteure erfüllt haben. Konkret waren folgende Fragen forschungsleitend:

- Welche Auswirkungen erwarten ausgewählte Gemeindevertreter und -mitarbeiter durch die Fusion der steirischen Gemeinden Anger, Baierdorf, Feistritz und Naintsch?
- Sind die erwarteten Auswirkungen dieser Fusion aus Sicht ausgewählter Akteure der (neuen) Gemeinde Anger eingetreten? Welche Schlussfolgerungen lassen sich auf Basis dieser vergleichenden ex-ante- und ex-post-Studie ableiten?

Der Beitrag dieser Arbeit liegt darin, dass die (sowohl vor als auch nach der Fusion) analysierte Einzelfallstudie detaillierte Reformeinblicke in die vergleichsweise selten betrachtete Mikroebene (Handlungen, Einstellungen und Wahrnehmungen von Mandatsträgern und Verwaltungspersonal) und teils die Mesoebene (Effekte auf die Organisationen bzw. Fusionsgemeinden) bietet, während vorhandene Untersuchungen meist auf der abstrakteren, aggregierten Makroebene ansetzen. Zudem existieren kaum Forschungsarbeiten, welche ex-ante- und ex-post-Perspektiven verbinden (Kuhlmann, Seyfried & Siegel, 2018b, S. 76 ff.). Der vorliegende Erfahrungsbericht illustriert insbesondere die subjektiv wahrgenommenen Herausforderungen bei der freiwilligen Fusion kleiner, ländlich geprägter Gemeinden und kann auch für Entscheidungsträger in Politik und Verwaltung eine Orientierungshilfe bieten.

Kapitel 2 stellt den Wissensstand zu den Auswirkungen von Gebietsreformen im Überblick dar und skizziert die Gegebenheiten in Österreich. In Kapitel 3 wird das Un-



tersuchungsdesign erörtert. Die gewonnenen Ergebnisse werden in Kapitel 4 präsentiert und in Kapitel 5 diskutiert. Abschließend folgen ein Fazit sowie ein Blick auf Limitationen und weitere Forschungsdesiderata.

## 2 Auswirkungen von Gemeindefusionen

Die Darstellung des Stands der Forschung wird durch aktuelle Literaturreviews zu den Wirkungen von kommunalen Gebietsreformen erleichtert (Kuhlmann, Seyfried & Siegel, 2018a; 2018b; Tavares, 2018). Anhand dieser Arbeiten wird der Kenntnisstand zum Thema zusammengefasst und ein Überblick über die Entwicklungen in Österreich und deren wissenschaftliche Rezeption geboten.

### 2.1 Stand der Forschung

Die Auswirkungen von Gebietsreformen hängen u. a. von der Reformstrategie (*bottom-up* vs. *top-down* initiierte und flächendeckend vs. schrittweise vollzogene Reformen), etwaigen Konflikten während des Implementierungsprozesses sowie von exogenen Faktoren (z. B. Abgabenrückgang durch Konjunkturerinbruch) ab (Steiner, Kaiser & Eythórsson, 2016, pp. 26 ff.). Obwohl die Zahl an Studien zu den Effekten von Strukturreformen in den letzten zehn Jahren stark gestiegen und deren inhaltliche Bandbreite groß ist, steht häufig die ökonomische Dimension im Vordergrund (Swianiewicz, 2018, pp. 5 ff.). Der derzeitige Kenntnisstand wird zu drei Dimensionen gebündelt: Auswirkungen auf die Leistungsfähigkeit, auf Einsparungen und Wirtschaftlichkeit sowie auf politische Teilhabe und Partizipation.

Die Wirkungsdimension der Leistungsfähigkeit subsumiert neben dem Umfang und der Qualität des Leistungsangebots sowie der Korrektheit und Effektivität der Verwaltungsarbeit auch den Professionalisierungsgrad der Verwaltung, ihre (fiskalische) Leistungskraft, Innovations- und Problemlösungsfähigkeit, kommunale Autonomie, Bürgerzufriedenheit sowie strategische Handlungs- und Planungsfähigkeit (Kuhlmann, Seyfried & Siegel, 2018b, S. 21). Antonio F. Tavares (2018, pp. 11 f.) spricht von „managerial implications“ und resümiert, dass mehrheitlich von keinen oder positiven Effekten auf die Qualität der lokalen Leistungserbringung berichtet wird. Demnach können fusionierte bzw. größere Gemeinden oft qualitativ bessere und diversere Leistungen anbieten. Wenngleich die Literatur sowohl positive als auch kritische Befunde zu tatsächlich realisierten Service- und Effektivitätsverbesserungen zeige, konstatieren auch Sabine Kuhlmann, Markus Seyfried und John Siegel (2018a, S. 126 ff.; 2018b, S. 33 ff.), dass international weitgehender Konsens darüber bestehe, dass Fusionen die institutionelle Kapazität von Kommunen stärken.

Beim Wirkungsfeld Einsparungen und Wirtschaftlichkeit orientiert man sich i. d. R. an Effizienzkriterien und Skaleneffekten („economies of scale and scope“); man geht davon aus, dass größere Gebietskörperschaften in Relation weniger Ressourcen benötigen und eine größere Zahl an Verwaltungsmitarbeitern eine Spezialisierung und professionellere fachliche Aufgabenerfüllung ermöglicht (Kuhlmann, Seyfried & Siegel, 2018b, S. 21). Tavares (2018, pp. 9 ff.) hält fest, dass signifikante Skaleneffekte eher unwahrscheinlich sind (außer in stark fragmentierten Systemen) und Einsparungen

im Bereich der allgemeinen Verwaltung häufig durch „diseconomies of scale“ und zusätzliche Ausgaben aufgehoben würden. Die Mehrzahl an Studien ließe die Erwartungen an Ausgabenreduktionen in Folge von Fusionen weitgehend unbefriedigt. Verbreitet sei aber ein opportunistisches Ausgabenverhalten im Vorfeld vieler (v. a. erzwungener) Gemeindefusionen. Auch Kuhlmann, Seyfried und Siegel (2018a, S. 128 ff.; 2018b, S. 49 ff.) erachten dieses sog. „Common-Pool“-Problem als empirisch bestätigt. Dabei tätigen Kommunen vor einer (meist schon beschlossenen) Fusion hohe Ausgaben bzw. verschulden sich zusätzlich, da dies der fusionierten Gemeinde angelastet werde. Dieser Mitnahmeeffekt beschneide mögliche Effizienzgewinne im Vorfeld. Insgesamt würden die Befunde zu postulierten Einsparwirkungen, Skaleneffekten und Effizienzsteigerungen zwar ein mehrdeutiges, unklares Bild zeichnen, dennoch seien entsprechende Wirkungen weder ausgeschlossen noch unwahrscheinlich.

Bei der politischen Partizipation stehen potenzielle Zusammenhänge zwischen der Gebietsgröße und demokratischer Mitwirkung, Integrationsfähigkeit, lokaler Identitätsbildung und sozialem Zusammenhalt (inkl. zivilgesellschaftlichem Engagement) im Blick (Kuhlmann, Seyfried & Siegel, 2018b, S. 23 f.). Tavares (2018, pp. 11 f.) konstatiert, dass das Gros an Studien, die Effekte von Gemeindefusionen auf die „quality of local democracy“ beleuchten, auf keine oder mehrheitlich negative Folgen hindeute (bspw. sinkende Wahlbeteiligungen, Zugewinne für populistische Parteien, negative Auswirkungen auf das Gemeinschaftsgefühl der Bewohner). Laut Kuhlmann, Seyfried und Siegel (2018a, S. 133 ff.; 2018b, S. 61 ff.) dürfe angesichts eher widersprüchlicher Befunde der Zusammenhang zwischen Gebietsreformen und den Dimensionen politischer Beteiligung nicht überschätzt werden. Trotz teils negativer Effekte auf die Partizipation und Integrationsfähigkeit bestünden vielfach ambivalente Befunde, etwa bei der Bürgerorientierung durch Zielkonflikte zwischen höherer Professionalisierung der Verwaltungsarbeit und längeren Anfahrtswegen. Negative Effekte könnten durch ergänzende digitale Angebote teils abgeschwächt werden. Generell sei die konkrete Ausgestaltung des Reformumsetzungsprozesses maßgeblich. Insgesamt folgern sie, dass die Gebietsgröße nur einen moderaten Effekt auf die lokale Demokratie habe. Etwaige „demokratische Kosten“ seien gering bis marginal und die teils befürchtete generelle Bedrohung des Ehrenamts sei nicht festzustellen. Zugleich ließen sich aber positive Effekte (z. B. auf das Politikinteresse) empirisch nur bedingt bis nicht nachweisen.

Kurzum: Fusionen scheinen keine bis mehrheitlich positive Effekte auf die Leistungsfähigkeit von Kommunen zu haben. Die ökonomischen Wirkungen bieten ein heterogenes Bild, welches Tavares (2018) kritischer zeichnet als Kuhlmann, Seyfried und Siegel (2018a; 2018b). Einig ist man sich hinsichtlich des *Common-Pool*-Problems. Relativ weit differieren die Einschätzungen zur politischen Teilhabe. Insgesamt bleiben also einige Fragen offen, obwohl die wissenschaftliche Auseinandersetzung mit Gebietsreformen deutlich zugenommen hat (Swianiewicz, 2018, p. 7). Die oft hohen Erwartungen seitens mancher Akteure in Politik, Verwaltung und Beratung dürften aber teilweise überhöht sein.

## 2.2 Befunde zur Situation in Österreich

Österreich gliedert sich in neun selbstständige Länder, die weitgehend frei bei der Gliederung ihres Gebietes in Gemeinden sind. Die Bundeshauptstadt Wien ist zugleich

Bundesland und Gemeinde mit eigenem Statut (Fallend, Mühlböck & Wolfgruber, 2001, S. 45). Gemeindefusionen, im Zuge derer zwei oder mehr Gemeinden zusammengelegt werden und ihre rechtliche Selbstständigkeit vollständig aufgeben (Steiner, 1999, S. 31), wurden in Österreich lange tabuisiert. Erst die 2011 initiierten und 2015 vollzogenen Strukturvereinbarungen in der Steiermark brachten Dynamik in die Reformdiskussion und -umsetzung (Wirth, 2011, S. 721 ff.; Stangl & Stötzer, 2015, S. 89 f.). Dennoch stellt sich die österreichische Gemeindestruktur im internationalen Vergleich noch immer kleinteilig dar (Bouckaert & Kuhlmann, 2016, pp. 11 f.; Prettenhaler, Kernitzkyi & Kurzmann, 2011, S. 2). *Tabelle 1* bietet einen Überblick zur Zahl der Gemeinden nach Bundesländern und Größenklassen und führt die durchschnittlichen Einwohnerzahlen pro Gemeinde an. Dabei wird deutlich, dass 20% aller Gemeinden weniger als 1.000 Einwohner (Ew) aufweisen; nur in etwa 4% aller Gemeinden leben mehr als 10.000 Menschen. In Österreich dominieren also kleine Gemeinden des ländlichen Raumes (Österreichischer Gemeindebund, 2019).

*Tabelle 1:* Gemeinden in Österreich nach Bundesländern<sup>1</sup> und Gemeindegrößenklassen (Gebietsstand 1.1.2019)

| Größenklasse              | B            | K            | NÖ           | OÖ           | S            | ST           | T            | V            | W                | Ö                                                     | % <sup>2</sup> | % <sup>3</sup> | Einwohner        |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|------------------|-------------------------------------------------------|----------------|----------------|------------------|
| 0-500                     | 17           | –            | 21           | 10           | 7            | 3            | 35           | 15           | –                | 108                                                   | 5,2            | 0,4            | 36.478           |
| 501-1.000                 | 39           | 12           | 84           | 70           | 14           | 15           | 59           | 18           | –                | 311                                                   | 14,8           | 2,7            | 237.810          |
| 1.001-1.500               | 44           | 26           | 123          | 80           | 15           | 48           | 55           | 10           | –                | 401                                                   | 19,1           | 5,7            | 501.430          |
| 1.501-2.000               | 27           | 29           | 104          | 67           | 10           | 48           | 32           | 8            | –                | 325                                                   | 15,5           | 6,3            | 559.886          |
| 2.001-2.500               | 16           | 14           | 59           | 55           | 8            | 42           | 22           | 10           | –                | 226                                                   | 10,8           | 5,7            | 500.568          |
| 2.501-3.000               | 9            | 11           | 42           | 37           | 10           | 26           | 16           | 4            | –                | 155                                                   | 7,4            | 4,8            | 424.815          |
| 3.001-5.000               | 14           | 21           | 73           | 67           | 34           | 58           | 36           | 13           | –                | 316                                                   | 15,1           | 13,7           | 1.209.394        |
| 5.001-10.000              | 4            | 12           | 41           | 39           | 14           | 34           | 16           | 8            | –                | 168                                                   | 8,0            | 12,9           | 1.146.332        |
| 10.001-20.000             | 1            | 4            | 19           | 8            | 5            | 10           | 7            | 6            | –                | 60                                                    | 2,9            | 9,0            | 794.915          |
| 20.001-30.000             | –            | 1            | 5            | 2            | 1            | 2            | –            | 2            | –                | 13                                                    | 0,6            | 3,6            | 322.077          |
| 30.001-50.000             | –            | –            | 1            | 1            | –            | –            | –            | 2            | –                | 4                                                     | 0,2            | 1,9            | 166.895          |
| 50.001-100.000            | –            | 1            | 1            | 1            | –            | –            | –            | –            | –                | 3                                                     | 0,1            | 2,0            | 179.014          |
| 100.001-200.000           | –            | 1            | –            | –            | 1            | –            | 1            | –            | –                | 3                                                     | 0,1            | 4,4            | 387.138          |
| 200.001-500.000           | –            | –            | –            | 1            | –            | 1            | –            | –            | –                | 2                                                     | 0,1            | 5,6            | 494.532          |
| > 1.000.000               | –            | –            | –            | –            | –            | –            | –            | –            | 1                | 1                                                     | 0,05           | 21,4           | 1.897.491        |
| <b>Summe</b>              | <b>171</b>   | <b>132</b>   | <b>573</b>   | <b>438</b>   | <b>119</b>   | <b>287</b>   | <b>279</b>   | <b>96</b>    | <b>1</b>         | <b>2.096</b>                                          | <b>100</b>     | <b>100</b>     | <b>8.858.775</b> |
| <b>Ø Ew/<br/>Gemeinde</b> | <b>1.712</b> | <b>4.249</b> | <b>2.916</b> | <b>3.349</b> | <b>4.644</b> | <b>4.321</b> | <b>2.692</b> | <b>4.080</b> | <b>1.897.491</b> | <b>3.306</b><br>(ohne W) /<br><b>4.205</b> (mit<br>W) |                |                |                  |

*Anmerkungen*

- Die Abkürzungen bezeichnen die Bundesländer Österreichs (Ö): Burgenland (B), Kärnten (K), Niederösterreich (NÖ), Oberösterreich (OÖ), Salzburg (S), Steiermark (ST), Tirol (T), Vorarlberg (V) und Wien (W).
- Diese Prozentangabe bezieht sich auf den Anteil der Gemeindeanzahl in der jeweiligen Größenklasse an der Gesamtanzahl der Gemeinden in Ö.
- Diese Prozentangabe bezieht sich auf den Anteil der Einwohnerzahl in der jeweiligen Größenklasse an der Gesamtbevölkerung in Ö.

*Quelle:* Eigene Darstellung; Datenbasis: Österreichischer Gemeindebund, 2019; Statistik Austria, 2019a.

Die Fusionsaktivitäten seit dem Zweiten Weltkrieg stellen sich in Österreichs Bundesländern unterschiedlich dar. *Tabelle 2* illustriert die Entwicklung der jeweiligen Gemeindeanzahl.

*Tabelle 2:* Gemeindeanzahl nach Bundesländern seit 1945

| Bundesland       | Gemeindeanzahl 1945 | Gemeindeanzahl 2012 | Gemeindeanzahl 2019 |
|------------------|---------------------|---------------------|---------------------|
| Vorarlberg       | 96                  | 96                  | 96                  |
| Tirol            | 287                 | 279                 | 279                 |
| Salzburg         | 119                 | 119                 | 119                 |
| Oberösterreich   | 445                 | 444                 | 438                 |
| Niederösterreich | 1.653               | 573                 | 573                 |
| Burgenland       | 320                 | 171                 | 171                 |
| Kärnten          | 241                 | 132                 | 132                 |
| Steiermark       | 1.004               | 542                 | 287                 |
| Summe (mit Wien) | 4.165               | 2.356               | 2.096               |

*Quelle:* Eigene Darstellung; Datenbasis: Gödl, 2013, S. 18; Statistik Austria, 2019a.

Die Zahl der Gemeinden wurde primär in den 1960er und 1970er Jahren durch Strukturbereinigungen in den Bundesländern Niederösterreich (NÖ), Burgenland (B), Kärnten (K) und Steiermark (ST) reduziert. Damit erhoffte man sich Verbesserungen der Leistungskraft und Effizienz. So fanden etwa in NÖ in den 1960ern freiwillige und erzwungene Fusionen statt; vergleichbar gestaltete sich die Entwicklung in der ST (Gödl, 2013, S. 18 ff.), welche mit der dortigen Gemeindestrukturreform 2015 einen neuen Höhepunkt erreichte.

Die Steiermark wurde von 2010 bis 2015 von einer Koalition der Sozialdemokratischen Partei Österreichs (SPÖ) und der Österreichischen Volkspartei (ÖVP) regiert. Diese sog. „Reformpartnerschaft“ initiierte u. a. die Gemeindestrukturreform, um die Durchschnittseinwohnerzahl der 542 Gemeinden (1.754 Ew pro Gemeinde; ohne Graz) zu erhöhen. Davor lebten in 200 steirischen Gemeinden nicht mehr als 1.000 Ew und die demografische Entwicklung ließ eine Zunahme der Zahl an Kleingemeinden erwarten. Die Kleinstrukturiertheit brachte diverse Probleme mit sich (z. B. ökonomische Probleme in vielen sog. Abgangsgemeinden<sup>3</sup>, finanzielle Belastungen durch Investitionen und den Erhalt der Infrastruktur, ineffizienter Wettbewerb um Unternehmensansiedelungen, Abwanderung usw.). Zugleich hätten sich die Aufgabenkomplexität sowie das Anspruchsniveau der Bürger und deren Mobilität erhöht. Auch sei der Koordinations- und Verwaltungsaufwand angesichts vieler Gemeinden hoch (Gödl, 2013, S. 21; Hörmann, 2016, S. 19 f.; Land Steiermark, 2011, S. 1; 2012, S. 4 ff.; Stangl & Stötzer, 2015, S. 90). Damit entspricht die skizzierte Ausgangslage gängigen Fusionsmotiven.

Als Ziele der Gemeindestrukturreform wurden gemäß dem Leitbild „Stärkere Gemeinden – Größere Chancen“ die Schaffung von wirtschaftlichen, leistungsfähigen und professionellen Gemeinden, abgestimmte raumordnungs- und verkehrspolitische Maßnahmen sowie eine Stärkung des Ehrenamtes bzw. Engagements für das Gemeinwohl propagiert. Durch die Zusammenschlüsse sollten alle Gemeinden grundlegende administrative Aufgaben erfüllen können. Ihre Finanzkraft sollte gestärkt, eine Spezialisierung der Mitarbeiter ermöglicht und Kosten eingespart werden. Zudem beabsichtigte man, die Leistungsqualität zu verbessern, verfügbare Infrastruktur effizienter und Flä-

chen besser zu nutzen sowie das vielfältige Vereinsleben zu erhalten bzw. zu stärken (Amt der Steiermärkischen Landesregierung, 2015a, S. 5; Land Steiermark, 2012, S. 22 ff.). Eine von der Steiermärkischen Landesregierung in Auftrag gegebene Studie zur Abschätzung möglicher Einsparpotentiale der Gemeindestrukturreform bezifferte diese auf ca. 40 Mio. Euro (Prettenthaler, Kernitzky & Kurzmann, 2011, S. 19).

Nach einer Vorschlagsphase (09/2011-01/2012), in der Gemeinden freiwillig Fusionsinitiativen einbringen konnten, folgten eine Verhandlungsphase (für Finanz- und Organisationsanalysen sowie Verhandlungen mit dem Land; 02-09/2012), eine Entscheidungsphase (10/2012-01/2013) und die Umsetzung per Jänner 2015, wodurch sich die Gemeindezahl von 542 auf 287 reduzierte (Amt der Steiermärkischen Landesregierung, 2015a, S. 34).

Obwohl sich die Landesregierung bemühte, die Gemeinden einzubinden, und man freiwillige Fusionen durch finanzielle Anreize belohnte (Fusions- bzw. „Hochzeitsprämien“; Seuberlich, 2017, S. 27; Pitlik & Wirth, 2012, S. 961), wurden jene Fusionskandidaten, die sich weigerten zu fusionieren, schließlich durch das Land dazu gezwungen. Insgesamt waren 385 der 542 Gemeinden unmittelbar von der Reform betroffen; 306 fassten einen Gemeinderatsbeschluss zur freiwilligen Vereinigung (Hörmann, 2016, S. 19). Gegen die übrigen Zwangsfusionen legten viele Gemeinden Beschwerden beim Verfassungsgerichtshof (VfGH) ein, der diese allesamt zurückwies (Amt der Steiermärkischen Landesregierung, 2015a, S. 34).<sup>4</sup>

Bislang liegt zu den Auswirkungen der Gemeindestrukturreform keine offizielle Evaluation durch das Land Steiermark vor. Jedoch veröffentlichte die Recherche-Plattform „Addendum“ in Kooperation mit der „Kleinen Zeitung“ im Jänner 2019 eine Analyse,<sup>5</sup> wonach die Verwaltungsausgaben pro Kopf in ca. 90% aller Fusionsgemeinden höher seien als vor der Reform (ebenso in 84% aller nicht fusionierten Gemeinden (ohne Graz)). Die Verwaltungsausgaben seien in den meisten Fusionsgemeinden bereits 2013 und 2014 (also vor der Fusion) und auch danach gestiegen. Man folgert: „Von Einsparungen ist also in den ersten drei Jahren nach der Gemeindestrukturreform nichts zu sehen (...) aber für die Bürger kann sich aus verschiedenen Blickwinkeln – Öffnungszeiten, Professionalisierung, Infrastruktur – dennoch einen [sic!] Mehrwert bieten“ (Gartner, Hellmann, Hametner & Müller, 2019). Seitens der Gemeindeabteilung des Landes Steiermark wurde entgegnet, dass sich Pensionierungen und Immobilienverwertungen erst in einigen Jahren auswirken würden und Einsparungen nicht das zentrale Motiv der Reform gewesen seien; im Vordergrund habe die Steigerung der Leistungsfähigkeit, Qualität und Wirtschaftlichkeit der Gemeinden gestanden (Hecke & Rossacher, 2019).

Abseits der Steiermark intensivierte sich die Fusionsdiskussion auch in Oberösterreich. *Tabelle 3* bietet einen Überblick zu dortigen (umgesetzten, geplanten sowie gescheiterten) Fusionsinitiativen.

Tabelle 3: Gemeindereforminitiativen in Oberösterreich

| Fusionsgemeinden/<br>-kandidaten                      | Status                                                 | neue Gemeinde                       |
|-------------------------------------------------------|--------------------------------------------------------|-------------------------------------|
| Weyer-Land, Weyer-Markt                               | umgesetzte Fusion (1.1.2007)                           | Weyer<br>(ca. 4.200 Ew)             |
| Aigen, Schlägl                                        | umgesetzte Fusion (1.5.2015)                           | Aigen-Schlägl<br>(ca. 3.200 Ew)     |
| Berg, Rohrbach                                        | umgesetzte Fusion (1.5.2015)                           | Rohrbach-Berg<br>(ca. 5.200 Ew)     |
| Bruck-Waasen,<br>Peuerbach (Steegen stieg aus)        | teilweise umgesetzte Fusion (1.1.2018)                 | Peuerbach<br>(ca. 4.500 Ew)         |
| Schönegg, Vorderweißenbach                            | umgesetzte Fusion (1.1.2018)                           | Vorderweißenbach (ca. 2.700 Ew)     |
| Ahorn, Helfenberg                                     | umgesetzte Fusion (1.1.2019)                           | Helfenberg<br>(ca. 1.600 Ew)        |
| Afiesl, St. Stefan am Walde                           | umgesetzte Fusion (1.1.2019)                           | St. Stefan-Afiesl<br>(ca. 1.200 Ew) |
| Hagenberg, Pregarten, Unterweisers-<br>dorf, Wartberg | gescheiterte Fusion                                    | (Stadt Aist)<br>(ca. 12.000 Ew)     |
| Innerschwand,<br>Mondsee, St. Lorenz, Tiefgraben      | ergebnisoffener<br>Fusionsprozess<br>(seit 2014)       | (Mondseeland)<br>(> 10.000 Ew)      |
| Eferding, Fraham, Hinzenbach, Popping                 | angedachte Fusion (Volksbefragung bis<br>2021 geplant) | (Eferding)<br>(> 10.000 Ew)         |

*Quelle:* Eigene Darstellung; Datenbasis: Auer, 2017; Dunst, 2014; Krenn-Aichinger, 2018; OÖNachrichten, 2017; Stangl & Stötzer, 2015; Statistik Austria, 2019a.

In Oberösterreich fanden bislang *bottom-up* Reformen statt, wobei die Fusionsinitiative von den Gemeinden selbst ausging und ohne Druck durch übergeordnete Ebenen stattfand. In der Steiermark hingegen verfolgte man eine „carrot and stick“-Strategie: Zuerst sammelte man – unterstützt durch finanzielle Anreize („carrots“) – freiwillige Fusionsvorschläge und anschließend wurden *top-down* Zusammenschlüsse gesetzlich („stick“) erzwungen (Kaiser, 2015, pp. 115 f.; Steiner, Kaiser & Eythórsson, 2016, pp. 26 ff.). Da sich in Oberösterreich (OÖ) bzw. Österreich (Ö) insgesamt immer noch viele Gemeinden mit einer Einwohnerzahl bis zu 1.000 finden und solche kleinteiligen Strukturen weithin als nachteilig erachtet werden (Wirth, 2011, S. 719), sind zukünftig weitere Gemeindefusionen wahrscheinlich.

International verbreitete Publikationen über Gebietsreformen in Österreich sind rar. Auch die in 2.1 angeführten Literaturüberblicke sowie der Ländervergleich von Reto Steiner, Claire Kaiser und Grétar Thór Eythórsson (2016) enthalten keine Bezüge zu Österreich. Pawel Swianiewicz (2018) erwähnt die Alpenrepublik am Rande als eines von vielen europäischen Ländern, wo zwischen 2008 und 2017 Gebietsreformen durchgeführt wurden (Steiermark); ähnlich stellt sich dies bei Jostein Askim, Jan Erling Klausen, Signy Irene Vabo und Karl Bjurstrøm (2016; 2017), Jens Blom-Hansen, Kurt Houlberg, Søren Serritzlew und Daniel Treisman (2016) sowie Claire Kaiser (2015) dar, wobei hier die Reformen zwischen 1950 und 2010 Erwähnung finden. Eine der spärlichen Veröffentlichungen mit österreichischem Fokus behandelt die Auswirkungen der steirischen Gemeindestrukturreform auf die Beteiligung bei Gemeinderatswahlen. Dabei kommen Reinhard Heinisch, Thomas Lehner, Armin Mühlböck und Christian Schimpf (2018) auf Basis aggregierter Daten der steirischen Kommunalwahl-

len 2010 und 2015 zu dem Schluss, dass Fusionen einen negativen Effekt auf die Wahlbeteiligung hätten. Zu ergänzen ist, dass die Reformdiskussion auch die (flächen- und bevölkerungsmäßig heterogenen) Bezirksverwaltungsbehörden erfasst.<sup>6</sup> Diesbezüglich analysieren Sebastian Blesse und Felix Rösel (2018) fiskalische und politische Effekte von Fusionen von Landkreisen in Sachsen-Anhalt und Bezirken in der Steiermark. Sie folgern, dass sich weder Kosten- noch Personalreduktionen fänden, jedoch „politische Kosten“ wie sinkende Wahlbeteiligungen und ein Erstarken rechtspopulistischer Parteien. Die Validität solcher Befunde wird jedoch von anderen Autoren stark angezweifelt. Bisweilen würden Korrelationen mit Kausalitäten verwechselt und unzulässig simple und pauschale Schlussfolgerungen gezogen (Kuhlmann, Seyfried & Siegel, 2018b, S. 28 und 62 f.).

### 3 Forschungsdesign

Um einen detaillierten Einblick in die erwarteten Auswirkungen von steirischen Gemeindefusionen und in die subjektiv wahrgenommenen Veränderungen zu gewinnen, wurden ein Fallstudiendesign und qualitative Forschungsmethoden gewählt. Ende 2012 wurden die oststeirischen Gemeinden Anger, Baierdorf, Feistritz und Naintsch als Einzelfallstudie ausgewählt. Diese Gemeinden entschlossen sich freiwillig zur Fusion und waren mit für die Region typischen Problemen konfrontiert, weshalb sie als „typischer Fall“ (Gläser & Laudel, 2010, S. 98) angesehen werden können. Zugleich erschien dieser Fall auch interessant, weil intendiert war, dass die Nachbargemeinde Floing mitfusionieren sollte; sie stieg jedoch 2012 aus dem Fusionsprozess aus. Alle Fusionskandidaten waren kleine Gemeinden, mit Bevölkerungszahlen zwischen 614 und 1.669 Ew (Statistik Austria, 2013a; 2013b; 2013c; 2013d), denen eine negative Bevölkerungsentwicklung prognostiziert wurde.

Im Jänner 2013 wurden neun Leitfadeninterviews (Gläser & Laudel, 2010, S. 42) mit Akteuren aller genannten Gemeinden persönlich geführt. Bei der Auswahl der Interviewpartner war intendiert, sowohl politische Repräsentanten (möglichst aller fünf involvierten Gemeinden) als auch Verwaltungsmitarbeiter als Gesprächspartner zu gewinnen. Die Mitwirkungsbereitschaft war gegeben, sodass die damaligen fünf Bürgermeister, ein Vizebürgermeister, eine Gemeinderätin sowie zwei Gemeindebedienstete interviewt werden konnten. Zusätzlich wurde im Mai 2013 ein Experteninterview geführt, mit der Intention, die gewonnenen Erkenntnisse und ihre Relevanz für andere Gemeinden zu diskutieren.

Es erschien sinnvoll, nach der Fusion (vollzogen per 1.1.2015) etwas Zeit verstreichen zu lassen, bevor ergründet wurde, ob sich die 2013 geäußerten Erwartungen erfüllt hätten. Im Jänner und Februar 2018 wurden auf Basis der Ergebnisse fünf Jahre zuvor erneut leitfadengestützte Interviews mit Vertretern der neuen Gemeinde Anger geführt. Erneut war beabsichtigt, sowohl politische Mandatäre (möglichst aller im aktuellen Gemeinderat vertretenen Parteien) als auch Gemeindebedienstete als Gesprächspartner zu gewinnen, was auch gelang. Von den zehn interviewten Personen waren acht bereits vor der Fusion in einer der Gemeinden (politisch oder in der Verwaltung) tätig; vier Personen sind identisch mit den Interviewpartnern im Jahr 2013. Aufgrund einzelner Anonymisierungswünsche werden sämtliche Interviewpartner (IP) nicht namentlich genannt.

Die insgesamt 20 Interviews dauerten im Durchschnitt 32 min (Bandbreite: 12-75 min), wurden alle aufgezeichnet und transkribiert. Das Material wurde mittels einer qualitativen Inhaltsanalyse (Mayring, 2010; Gläser & Laudel, 2010, S. 197 ff.) ausgewertet. Das zu Projektbeginn literaturgestützt entwickelte Kategoriensystem diente sowohl der Strukturierung beider Interviewleitfäden als auch der Kodierung und Materialanalyse sowie der Gliederung der Ergebnisdarstellung. Im Zuge der Extraktion bzw. Auswertung durch jeweils zwei Autoren wurden teilweise induktiv generierte Subkategorien ergänzt. Die Erkenntnisse beider Interviewwellen werden nachfolgend präsentiert, wobei lediglich wörtliche Stellungnahmen (in anonymisierter Form) zitiert werden.

## 4 Die neue Gemeinde Anger – Vergleich von ex-ante artikulierten Erwartungen und ex-post wahrgenommenen Auswirkungen

### 4.1 Gesamtschau

Als Beweggründe für die Reform führten die Interviewten die kleinstrukturierte Gemeindelandschaft, sich wandelnde kommunale Aufgabenstellungen, die von Kleingemeinden kaum noch zu bewältigen seien, und die vielfach prekäre finanzielle Lage an. Auch angesichts der negativen Bevölkerungsentwicklungsprognosen habe man sich freiwillig zur Fusion entschlossen. Die folgende Aussage fasst die Fusionsmotivation zusammen:

„Es ist in den letzten Jahren immer schwieriger geworden (...) da ist man eben gezwungen gewesen, dass man sich irgendetwas überlegt, wo wir auch in Zukunft ein bisschen gestalten können; dass man zusammengeht und mit Finanzstärke (...) wieder ein bisschen etwas für die Leute machen kann“ (IP1, 17.01.2013).

Die im Vorfeld der Fusion kommunizierten Erwartungen waren vielfältig. *Tabelle 4* stellt die (positiven wie negativen) Erwartungen der Interviewpartner im Überblick dar und stellt sie den im Jahr 2018 subjektiv wahrgenommenen Vor- und Nachteilen gegenüber. Nach der Fusion äußerten die meisten Interviewpartner, dass diese eine positive Entwicklung befördert habe. Hervorgehoben wurden die Mitarbeiterspezialisierung, dass es keine Personalengpässe mehr gäbe, und die Fusionsprämie, welche Investitionen ermöglicht habe. Ohne die vormaligen Gemeindegrenzen als Hindernis könne man Projekte einfacher und rascher umsetzen. Kritik gab es vereinzelt bezüglich der Gebührenharmonisierung, der nicht mehr so ausgeprägten persönlichen Kontakte und einer schwierig(er)en Mobilisierung der Bevölkerung für Veranstaltungen. Eine Diskrepanz zwischen den ex-ante artikulierten Erwartungen und den ex-post wahrgenommenen Effekten zeigte sich speziell bei den (bislang) nicht realisierten Einsparungen. Die schon im Vorfeld positive Einstellung der meisten Gesprächspartner gegenüber der Fusion sei dennoch unverändert. Nur eine Person ist weiterhin skeptisch.



Tabelle 4: Einschätzungen der IP zu Vor- und Nachteilen der Gemeindefusion

|                  | Erwartungen vor der Fusion                                                                                                                                                                                                   | Einschätzungen nach der Fusion                                                                                                                                                                                                       |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Vorteile</b>  | höherer finanzieller Spielraum, Einsparungen<br>effiziente Infrastrukturnutzung<br>mehr Professionalität (v.a. durch Mitarbeiter-spezialisierung),<br>mehr Qualität & Services für Bürger, Flexibilität,<br>Rechtssicherheit | finanzielle Vorteile (v.a. Fusionsprämie)<br>verbesserte Infrastruktur & Ressourcenbündelung<br>(v.a. im Bauhof)<br>zentrale Gemeindeverwaltung: Spezialisierung,<br>zeitgemäße EDV, Vertretungssystem<br>=> schneller & kompetenter |
| <b>Nachteile</b> | größere Anonymität (Rückgang der „Volksnähe“/<br>persönlicher Kontakte)<br>Vernachlässigung zentrumsferner Gebiete inkl.<br>beschleunigter Abwanderung, längere Wege, ggf.<br>Konflikte<br>z.T. Gebührenerhöhungen           | persönliche Kontakte nicht mehr so ausgeprägt<br>eher rückläufiges (freiwilliges) Engagement<br>einzelne kritische Bürger (v.a. wegen Gebührener-<br>höhungen)<br>keine Einsparungen                                                 |

Quelle: Eigene Darstellung.

Die Interviewergebnisse werden nachfolgend in sieben Kategorien gegliedert: Erwartete bzw. wahrgenommene Auswirkungen auf die Bürger, Demokratiequalität, Gemeindepolitik, Gemeindeverwaltung, Gemeindefinanzen und -investitionen, lokale Wirtschaft, innere Integration und Eigenständigkeit. Abschließend werden die Beweggründe der Gemeinde Floing, aus den Fusionsverhandlungen auszusteigen, beleuchtet.

## 4.2 Bürger

Neben (partiellen) Umstellungen bei Ortstafeln, Wappen und Adressen wurden Veränderungen bei öffentlichen Einrichtungen thematisiert. Seit der Fusion gibt es nur noch ein Gemeindeamt in Anger, das barrierefrei zugänglich ist und längere Öffnungszeiten hat. Wenngleich es Bevölkerungsgruppen gäbe, für die etwaige längere Anfahrtswege nachteilig sein könnten, seien diese angesichts der gestiegenen Mobilität vernachlässigbar: „Wenn jemand da oben (...) wohnt und er muss einkaufen gehen, dann muss er so und so runterfahren in den Ort. (...) wir sind nicht mehr mit Ochsen unterwegs (...) das ist heutzutage überhaupt kein Problem mehr“ (IP5, 17.1.2013). Man war sich einig, dass das Amt zentral liege, gut erreichbar sei und sich die Wege nicht merklich verändert hätten. Zudem war man überzeugt, dass sich der Bürgerservice verbessert habe.

Wenn durch Fusionen Wege teils länger werden und man über mehr Ressourcen verfügt, könnte dies den Ausbau von E-Government-Angeboten befördern. Diesbezüglich hielten sich aber zu beiden Befragungszeitpunkten sowohl das Angebot als auch die Nachfrage in Grenzen. Nur gewisse Leistungen (z. B. Online-Beantragung von Stimmkarten für Volksbefragungen) würden intensiver beansprucht. Alle Befragten äußerten, dass die Bevölkerung weiterhin persönliche Kontakte bzw. konventionelle Verwaltungswege präferiere. Insgesamt seien die persönlichen Kontakte (im Sinne von «jeder kennt jeden») in der größeren Gemeinde aber nicht mehr so ausgeprägt wie früher.

Man erwartete ferner durch eine erweiterte Wohnraumschaffung positive Auswirkungen auf die Wohnortattraktivität, wofür auch die Infrastruktur und Arbeitsplätze vor Ort wichtig seien. Die Bebauungsmöglichkeiten waren in Anger alt überwiegend erschöpft. Durch die gemeinsame Raumplanung könne man das kleinräumige Denken überwinden und Standortnachteilen begegnen. 2018 arbeitete man aber noch an einem

neuen Flächenwidmungsplan bzw. Raumordnungskonzept (geplante Fertigstellung: Mitte 2019). Bei der Bevölkerungsentwicklung zeigen sich leichte Rückgänge (Einwohnerzahl 2015: 4.141; 2018: 4.098 Ew; KDZ – Zentrum für Verwaltungsforschung, 2019).

### 4.3 Demokratiequalität

Die Gesprächspartner interpretierten Demokratiequalität als hohe Wahlbeteiligung, gute Informationsflüsse für die Bürger und deren Möglichkeit, ihre Anliegen einzubringen. Man war sich einig, dass die Demokratiequalität hoch sei, was man auf die persönlichen Beziehungen mit den Bürgern, kurze Wege und eine rasche Beseitigung von Missständen zurückführte. Bürgernähe wurde mit persönlichen Kontakten gleichgesetzt. Damit verbunden sei Beziehungsarbeit, die erneut durch persönliche Kontakte dominiert sei, neben formellen Informationskanälen wie Gemeindezeitung und -homepage. Angesichts der Reduktion der Mandatare durch die Fusion erwartete man, dass die Beziehungen loser würden, sich die (davor kurzen) Problemlösungswege verlängern und persönliche Naheverhältnisse abschwächen könnten; dies könne aber auch zu einer Objektivierung beitragen. Man rechnete damit, dass die Bürgernähe dennoch bestehen und informelle Kommunikation wichtig bliebe. Zudem werde der neue Bürgermeister künftig von Verwaltungsaufgaben entlastet und hätte mehr Zeit für Beziehungsarbeit.

Zu erwähnen ist, dass die Gemeinden von einer Volksbefragung über die Fusion absahen und diese durch vier gleichlautende Gemeinderatsbeschlüsse verabschiedeten. Dies wurde so begründet: „Weil wir einfach sagen, diese Entscheidung wäre nur dann sinnvoll, wenn wirklich alle Bürger mit Sicherheit wissen, welche Vor- und welche Nachteile das haben wird. Hat sich jeder gut informiert? Und dass er es nicht aus dem Bauchgefühl herausmacht“ (IP8, 18.1.2013).

Auch eine aktive Bürgerbeteiligung wurde als Aspekt der Demokratiequalität erachtet. Jedoch sei das politische Engagement der Einwohner gering, der Zulauf zu Informationsveranstaltungen schwach und angesichts einer allgemeinen Politikverdrossenheit werde es immer schwerer, Menschen zu einem politischen Engagement zu motivieren. Die Mehrzahl erwartete diesbezüglich keine Veränderungen. Zugleich sei es nach der Fusion wichtig, dass sich Menschen aller Gemeindegebiete, Alters-, Geschlechts- und Berufsgruppen beteiligen, damit die Interessen möglichst vieler Bevölkerungsgruppen aus allen Ortsteilen artikuliert und berücksichtigt würden. Es werde an den Menschen liegen, ob sie sich engagieren: „(...) sie müssen halt, um sich Gehör zu verschaffen und ihre Interessen zu vertreten, (...) lauter melden wie bisher“ (IP2, 17.1.2013).

Rückblickend habe sich an der politischen Partizipation der Bürger und am Vereinsleben nichts geändert. Vereinzelt wurden aber sinkende Besucherzahlen und Mithilfebereitschaft bei Veranstaltungen erwähnt. Die Verwaltung unterhalte unveränderte (primär persönliche) Kontakte zur Bevölkerung; die Gemeinderäte pflegten Kontakte v. a. auf Vereinsebene, bei Veranstaltungen und in Gasthäusern. Die Zahl der Termine der (Vize-)Bürgermeister sei angesichts vieler Veranstaltungen im größeren Gemeindegebiet gestiegen. Bei den Gemeinderatswahlen 2015 betrug die Beteiligung 78,5%. Bei den Gemeinderatswahlen 2010 war diese in allen früheren Gemeinden höher (Anger: 83,8%; Baierdorf: 80,8%; Feistritz: 83,6%; Naintsch: 92,2%). Trotz dieser Rück-

gänge ist die Beteiligung im Vergleich zu den Gesamtergebnissen im Land Steiermark (2015: 73,4%; 2010: 77,2%) vergleichsweise hoch. Auch in (der nicht mitfusionierten Gemeinde) Floing ging die Wahlbeteiligung (2015: 76,9%; 2010: 87,3%) zurück (Amt der Steiermärkischen Landesregierung, 2015b).

#### 4.4 Gemeindepolitik

Alle vier Kommunen waren klar von der ÖVP dominiert. Zweitstärkste Fraktion war in drei Gemeinden die SPÖ und in einer die Freiheitliche Partei Österreichs (FPÖ). Für kleinere Parteien war es aufgrund der kleinen Wahlsprengleinheiten bei Gemeinderatswahlen kaum möglich, dass sie die für ein Mandat erforderliche Stärke erreichten. Man erwartete, dass sich die Parteienvielfalt nach der Fusion erhöhen könnte (v. a. zugunsten der FPÖ), was sich bestätigte. Hinsichtlich der Zusammensetzung des neuen Gemeinderats blieb die ÖVP die stärkste Fraktion (13 Sitze; 62,2%), neben der SPÖ (4 Sitze; 19,6%) und FPÖ (4 Sitze; 18,2%), die zuvor nur in Feistritz vertreten war (Amt der Steiermärkischen Landesregierung, 2015b). Da neu vertretene Parteien versuchen würden, Aufmerksamkeit zu erregen, erwartete man ein erhöhtes Konfliktpotential. 2018 verwiesen alle Befragten darauf, dass es mehr Diskussionen im Gemeinderat gäbe (aufgrund der größeren Parteienvielfalt und der verschiedenen Ortsteilzugehörigkeit). Die Fraktionszusammenarbeit funktioniere jedoch, da man darauf achte, dass kein Ort bevorzugt werde. Auch sei man sich fraktionsübergreifend einig (gewesen), intern alle Informationen weiter zu geben und nach außen geschlossen mit einer Sprache zu sprechen. Es sei positiv (v. a. bei ortsteilübergreifenden Projekten), dass Entscheidungen nur noch in einem Gemeinderat getroffen werden und Entscheidungen somit einfacher möglich wären.

Die Zahl der Gemeinderäte reduzierte sich von 48 auf 21. Man erwartete bezüglich der Personenauswahl und -reihung auf den Fraktionslisten intensive Verhandlungen und ggf. Konflikte sowie dass manche Mandatäre mehr Verantwortung und Entscheidungsbefugnisse bekommen würden. In der Rückschau berichtete man, dass es bei der Zusammenlegung der Gemeinderatsklubs<sup>7</sup> keine Konflikte gegeben habe. Vorteilhaft wäre gewesen, dass drei der vier ÖVP-Bürgermeister bereits während der Fusionsverhandlungen angekündigt hatten, ihr Amt niederzulegen. Der Bürgermeister von Anger (seit 2006 Hubert Höfler) wurde im Zuge der Fusion vom Land als Regierungskommissär beauftragt und bei den Gemeinderatswahlen im März 2015 Bürgermeister der neuen Gemeinde. Er arbeite nun nicht mehr so oft operativ mit, jedoch habe sich seine Arbeit intensiviert (mehr Veranstaltungen, Besprechungen etc.).

#### 4.5 Gemeindeverwaltung

Ab 2015 sollten die Gemeindeämter in Baierdorf, Feistritz und Naintsch geschlossen und anderweitig genutzt werden. Beim Verwaltungspersonal erwartete man erst Einsparungen, wenn Stellen nach Pensionierungen nicht nachbesetzt würden; bis dahin würden alle acht Vollzeit- und zwei Teilzeitangestellten bleiben. Dementsprechend wurden alle Mitarbeiter übernommen und in das Amt in Anger übersiedelt, wo eine Bürgerservicestelle mit längeren Öffnungszeiten und besser geschultem Personal geschaffen wurde. Eine Stelle wurde nach einer Pensionierung nicht nachbesetzt, da man

diesen Abgang durch interne Umstrukturierungen kompensieren konnte. Jedoch wurde der Mitarbeiterstand in den Bereichen Bauhof und Kinderbetreuung aufgestockt. Durch die Bündelung habe sich das Vertretungssystem wesentlich verbessert, sodass es keine gravierenden Engpässe wie früher mehr gäbe: „Wenn man zwei Mitarbeiter hat, einer ist krank, der andere ist in Urlaub, dann steht der Betrieb (...)“ (IP17, 29.1.2018).

Erwartungsgemäß kam es zu Abteilungsbildungen und Spezialisierungen. Zuvor konnte man angesichts begrenzter Ressourcen keine ausdifferenzierte Arbeitsteilung organisieren. Die Mitarbeiter deckten meist alle Fach- und Aufgabenbereiche als Generalisten ab. Auch die Bürgermeister übernahmen oft Verwaltungsaufgaben, was nach der Fusion nicht mehr möglich (oder nötig) sei. Durch die Spezialisierung versprach man sich eine höhere Professionalität, Rechtssicherheit und Qualität von Bescheiden sowie eine raschere Aufgabenerledigung. Zwei Aussagen veranschaulichen das diesbezüglich durchgängig positive Fazit:

„Es ist sicherlich zeitgemäß so (...) das ist mehr ein Modell der Zukunft, weil sich so (...) Abteilungen im Gemeindeamt bilden können, sodass die einen nur für das Baurecht zuständig sind und die anderen nur für die Finanzgebarung. (...) Das ist die Zukunft“ (IP12, 23.1.2018).

„Ich bin überzeugt davon, dass wir schneller sind (...) und kompetenter, weil einfach durch diese Spezialisierung (...) die Auskünfte rascher, fundierter und sicherer gegeben werden können. Weil wenn du alles können musst (...), wirst du einfach nicht in der Lage sein, trotz Bemühen, in manchen Bereichen das auch 100%ig sicher zu sagen. (...)“ (IP17, 29.1.2018).

Man ging davon aus, dass in der Umstellungsphase und bei der Ablaufoptimierung auf die Gemeindebediensteten viel Arbeit zukomme. Eine Verwaltungsmitarbeiterin erwartete zudem „Reibereien“, da die Umstellung von einem kleinen, eingespielten Team auf ein größeres mit neuen Abläufen und neuer Arbeitsteilung zu Konflikten führen könnte. Auch der interviewte Experte erwartete solche Herausforderungen: „Es wird sicher eines längeren Zeitraums bedürfen, bis sich das alles eingespielt hat, die ganzen Verwaltungsabläufe, da ist sicher ein höherer Koordinationsaufwand gegeben. (...) und natürlich prallen bei einer Fusion mehrere Verwaltungskulturen aufeinander (...) natürlich kann es da zu Konflikten kommen oder zu (...) Spannungen im Team“ (IP10, 2.5.2013). Im Rückblick äußerte man oft, dass die Fusion für die Mitarbeiter schwieriger gewesen sei als für die Bürger, da es anfangs Abstimmungsprobleme gegeben habe. Mehrheitlich beurteilte man die Zusammenarbeit im Verwaltungsbereich jedoch als harmonisch. Nur eine Person äußerte, dass es in der Verwaltung und im Bauhof Konflikte (v. a. um Leitungsfunktionen) gäbe.

#### 4.6 Gemeindefinanzen und -investitionen

Österreichische Gemeinden finanzieren sich primär aus Finanzausweisungen von Bund und Ländern, eigenen Steuereinnahmen sowie Gebühren und Beiträgen (Fallend, Mühlböck & Wolfgruber, 2001, S. 50). Im Rahmen des Finanzausgleichs erhöhten sich die Gemeindeertragsanteile insgesamt nicht, da die neue Gemeinde die Stufe von 10.000 Ew nicht erreicht.<sup>8</sup> Die Bedarfszuweisungen des Landes hätten sich gesteigert. Das Kommunalsteueraufkommen differierte stark zwischen Naintsch als finanzschwächster Gemeinde mit wenigen Betrieben und Baierdorf als finanzstärkster Gemeinde mit den höchsten Kommunalsteuereinnahmen. Durch die Fusion teilen sich die Mittel auf die gesamte neue Gemeinde auf.

Die Höhe der Gebühren für Infrastruktureinrichtungen bestimmt sich nach dem individuellen Kostendeckungsprinzip; die Bürger bezahlen diese nach der tatsächlichen Nutzung (Schauer, 2019, S. 128). Hier gab es Anpassungen und (teilweise) Verschlechterungen für die Bürger, da die heterogenen Gebühren (für Kanal, Wasser, Müll) in der neuen Gemeinde vereinheitlicht wurden. Eine Person äußerte sich hierzu sehr kritisch:

„Billiger geworden ist nichts. Beim Wasser hat man sogar zwei Jahre gewartet, die Erhöhung kommt jetzt erst, obwohl man damals schon gewusst hat, dass Wasser nicht kostendeckend ist. Kanal hat man im ersten Jahr nach der Fusionierung ein bisschen angehoben (...) jetzt müssen wir gewaltig anheben, weil wir auch hier nicht kostendeckend sind“ (IP19, 29.1.2018).

Generell erwartete man eine bessere Finanzsituation sowie Infrastrukturverbesserungen in baulicher und maschineller Hinsicht, da Fusionsprämien gewährt wurden, die pro wegfällender Gemeinde 200.000 Euro über Bundesmittel vorsahen und nochmal dieselbe Summe über Landesmittel aus dem Bedarfszuweisungstopf für Gemeinden, die in der ersten Reformphase einer freiwilligen Fusion zustimmten. Somit erhielt die neue Gemeinde 1,2 Mio. Euro. Dieses Geld wurde investiert u. a. in die Schulsanierung, den Musikverein, Sport- und Kultursaal, ein Abfallsammelzentrum, Kindergärten, Straßen und die Beleuchtung. Einige dieser Investitionen wären ohnehin nötig gewesen, waren aber durch die Fusion leichter zu finanzieren. Zwar ist der Schuldenstand durch diese Investitionen gestiegen, aber es gibt keine Abgangsgemeinde mehr (früher zwei). Man achte bei Investitionen darauf, dass alle Ortsteile profitieren, damit „da einfach keiner zu kurz kommt, und dass ja nicht das entsteht, dass sie sagen: naja, ist eh logisch, wir sind ganz im letzten Eck und wir kriegen nichts“ (IP17, 29.1.2018).

Angesichts der Zusammenlegung auf ein Gemeindeamt rechnete man mit Einsparungspotentialen durch verringerte Betriebs- und Instandhaltungskosten. Welchem Zweck die aufzulassenden Gemeindeämter zugeführt würden, hatte man 2013 noch nicht festgelegt. Beim Verwaltungspersonal benötigte man kurzfristig alle Mitarbeiter für den Umstrukturierungsprozess; man sei ohnehin schlank aufgestellt. Einsparungen erwartete man ferner durch die gemeinsame Beschaffung und Nutzung von Anlagen, im EDV-Bereich und durch die Reduktion der Mandatare, Bürgermeister und Vizebürgermeister (von vier auf zwei). Anzumerken ist, dass man schon vielfach kooperierte. Insgesamt erwarteten die Interviewpartner positive Auswirkungen auf die Gemeindefinanzen; monetäre Aspekte waren ein primäres Motiv für den Zusammenschluss. Man hätte berechnet, dass man durch die Fusion in der Verwaltung ca. 90.000 Euro jährlich einsparen könne. Durch die Reduktion der politischen Mandatare ergäbe sich eine jährliche Ausgabenreduktion von 100.000 Euro. Das jährliche Gesamteinsparungspotential wurde auf 250.000 bis 300.000 Euro geschätzt. Die diesbezüglich großteils hohen Erwartungen blieben drei Jahre nach der Fusion unerfüllt.

Das Gemeindeamt in Anger wurde renoviert und adaptiert, während die aufgelassenen Ämter für die Feuerwehr und kulturelle Veranstaltungen (Feistritz) sowie als Kinderkrippe und -garten (Baierdorf) genutzt werden. Das ehemalige Amtsgebäude in Naintsch steht leer und soll verkauft werden: „Naintsch wird jetzt im Moment gar nicht genutzt, dort hoffen wir, dass wir es irgendwann verkaufen können“ (IP18, 29.1.2018). Insgesamt ergäben sich Einsparungen bei den Betriebskosten, wobei ein Gesprächspartner solche nicht sieht. Laut einer anderen Person würde die Feuerwehr in Feistritz zwar die Betriebskosten tragen, aber keine Miete bezahlen.

## 4.7 Lokale Wirtschaft

In Baierdorf befindet sich der größte Industriebetrieb der Region, in dem Bewohner aller fünf Gemeinden tätig seien. Jedoch gab es keine Bebauungsflächen mehr für gewerbliche oder industrielle Entwicklungen (dasselbe Problem hatten Anger und Naintsch). Durch die Fusion könne man neue (Vorbehalts-)Flächen für Betriebsansiedlungen zur Verfügung stellen und die Einnahmen aus Gebühren und Kommunalsteuer gerechter verteilen. Wie geplant gestaltete man gemeinsame Entwicklungskonzepte und harmonisierte die Gebührenpolitik und Wirtschaftsförderung mit dem Ziel, Betriebe in der Region zu halten und neue anzulocken. Allerdings kritisierte eine Person, dass Kleinbetriebe die Förderkriterien nicht erreichen und somit nicht ansuchen würden. Zudem wurde nach der Fusion durchweg betont, dass sich das Kommunalsteueraufkommen und die Arbeitsplatzsituation fusionsunabhängig gemäß der allgemeinen Wirtschaftslage entwickelten.

## 4.8 Innere Integration und Eigenständigkeit

Wesentlich für das Gelingen der Fusion sei ein Ausbalancieren des Spannungsfeldes zwischen der Integration in der neuen Gemeinde und der Bewahrung der Eigenständigkeit der Orte. Ein Maß für Eigenständigkeit wurde darin gesehen, inwieweit eine Gemeinde ausreichend Arbeitsplätze für die Bevölkerung bereitstellen kann. Die Arbeitsplatzsituation in der Oststeiermark sei durch eine Zentralisierung hin zu Weiz und Graz gekennzeichnet; die Auspendlerquoten seien vielfach hoch. Der große Möbelhersteller in Baierdorf ist ein wichtiger Arbeitgeber in der Region, dennoch werde Anger auch künftig eine Auspendlergemeinde sein. Der Erhalt der lokalen Arbeitsplätze durch wirtschaftsfördernde Maßnahmen sei eine der größten Herausforderungen.

Alle Befragten waren sich einig, dass Anger als zentraler Ort anerkannt und ein historisch gewachsenes Zentrum im Tal sei. Das Zusammengehörigkeitsgefühl zwischen den Gemeinden sei – auch durch die gemeinsame Pfarrei – schon früher vorhanden gewesen. Wichtige Einrichtungen konzentrieren sich in Anger (Ärzte, Polizei, Kindergarten, Schulen, Standesamt, Einkaufszentrum), dessen zentralörtliche Funktion durch Kooperationen vor der Fusion ausgebaut wurde: Es entstanden Wohnungen, eine Bücherei sowie der Staatsbürgerschafts- und Standesamtsverband. Somit hätte man sich in Feistritz und Baierdorf bereits vor dem Zusammenschluss stark nach Anger orientiert. Nur Heilbrunn in Naintsch (ein auf einem Berg gelegener Wallfahrtsort) wurde als Ort mit hoher Eigenständigkeit angesehen, da man über ein reges Vereinsleben, eine eigene Pfarrei, Bank, Poststelle, einen Kindergarten, Gasthäuser und andere Betriebe verfüge. Die dortige Volksschule wurde 2012 aufgrund niedriger Schülerzahlen geschlossen; auch dadurch seien „die Weichen“ für die Fusion gestellt worden. Generell dürften Schulzusammenlegungen (Zuständigkeit des Landes) angesichts der Schülerzahlentwicklung in Zukunft ein Thema werden.

Die Ansichten, ob es seit der Fusion zu einer Zentralisierung und Vernachlässigung zentrumsferner Orte (und Abwanderung von Menschen) komme, waren geteilt. Großteils war man der Ansicht, dass es nicht sinnvoll sei, andere Zentren zu erhalten oder auszubauen, da dies das Einsparungspotential beeinträchtige. Jedenfalls habe man nach 2015 keine neuen Zentren geschaffen, jedoch den Ort Heilbrunn und die dortigen Ver-

eine unterstützt und besonders in entlegeneren Gebieten einiges unternommen, um den Bewohnern ein positives Gefühl gegenüber der neuen Gemeinde zu vermitteln und Ängste auszuräumen. Es sei wichtig, die Interessen aller Ortsteile zu berücksichtigen, um Identitätsverlusten vorzubeugen. Identitätsstiftend sei primär das örtliche Vereinsleben. Insofern war man darauf bedacht, dieses unangetastet zu lassen und keine Förderungen zu reduzieren. Speziell die freiwilligen Feuerwehren seien lokal verwurzelte Institutionen, mit denen sich Mitglieder und Förderer stark identifizierten. Deshalb komme eine Fusion dieser Organisationen nicht in Frage, zumal man das dortige Engagement nicht beeinträchtigen möchte: „Das war das große Schreckgespenst vieler Argumentationen: die Vereine werden aufgelöst“ (IP8, 18.1.2013). Sinnvoll wären aber eine gemeinsame Geräteanschaffung und stärkere Kooperation der Feuerwehren. Die Fusion habe keine wesentlichen Auswirkungen auf das Vereinsleben, allerdings werde es generell schwieriger, Menschen zu einem (v. a. langfristigen) Engagement zu bewegen. Dies sei aber im urbanen Bereich noch ausgeprägter, da man sich am Land persönlich kennt und stärkere Bindungen bestehen.

#### 4.9 Der „Fusions-Aussteiger“ Floing

Die Gemeinde Floing entschied 2012, aus den Fusionsverhandlungen auszusteigen. Dies begründete man mit einer guten Finanzsituation, Infrastrukturausstattung und positiven Bevölkerungsentwicklung. Man verfüge über ein Zentrum mit Kindergarten, Volksschule und Geschäften sowie ein reges Vereinsleben. Erwartungsgemäß stand man in Floing Gemeindefusionen ablehnend gegenüber, speziell da man keine Einsparungspotentiale sehe. Folgende Aussage fasst diese Ansicht pointiert zusammen: „Wir sind ja keine Abgangsgemeinde, wir stehen relativ gut da. Also warum sollen wir fusionieren? Wir haben ziemlich viel, wir können uns noch einiges leisten. (...) Und darum wollen wir auch alleine bleiben“ (IP9, 18.1.2013).

Durch eine Fusion würden sich die Bürgernähe und Lebendigkeit des Zentrums verschlechtern. Das dann in Anger situierte Gemeindeamt hätte eine zunehmende Orientierung dorthin und längere Anfahrtszeiten impliziert, was die Attraktivität von Floing negativ beeinflusst hätte. Die Umstellungen hätten nachteilige Effekte auf die Leistungsqualität ergeben (können). Zudem erwartete man keine Erhöhung der Abgabeneinnahmen und des Investitionsspielraums; vielmehr müsste Floing einiges an Infrastruktur aufgeben. Auch hätte man durch die geografische Randlage ggf. Nachteile, da man in einer „Großgemeinde“ nicht adäquat auf die vielfältigen Interessen eingehen könne. Potenzielle Identitätsverluste seien eine generationsspezifische Thematik: Jungen Bürgern sei es relativ gleichgültig, ob sie Floing oder Anger angehören, aber ältere Einwohner hätten mit solchen Veränderungen oft Schwierigkeiten. Man war ferner überzeugt, dass die Bereitschaft zur Übernahme von Ehrenämtern abnehmen würde, was zu Vereinszusammenlegungen führen könnte. Hinsichtlich der Arbeitsplatzsituation würde eine Fusion keine Unternehmen dazu animieren, sich in Floing anzusiedeln. Eine Fusion brächte also primär Nachteile für die Gemeinde, die ihre Eigenständigkeit verloren hätte. Die anderen Interviewpartner bedauerten diese Haltung durchwegs. Floings Ausstieg sei „ein kleiner Wehrmutstropfen. (...) Das hätte gut dazu gepasst. Leider Gottes gibt es dort halt andere Überlegungen, ob die jetzt gut und richtig sind, ist eine andere Frage; das wird sich herausstellen“ (IP8, 18.1.2013).

## 5 Diskussion

Die gewonnenen Erkenntnisse deuten darauf hin, dass die Fusion der vier Gemeinden für die Bürger Verbesserungen bezüglich des Umfangs und der Qualität kommunaler Leistungen mit sich brachte; speziell durch erweiterte Öffnungszeiten und Leistungsangebote, die Mitarbeiterspezialisierung, deren Vertretungssystem und den Infrastrukturausbau. Diese wurden jedoch mit teils gestiegenen Gebühren, einem erhöhten Schuldenstand, nicht ausgeschöpften Einsparungspotentialen und längeren Wegen „erkauft“. Klaus Wirth (2011) ist beizupflichten, dass „Bürgernähe (...) nicht nur [bedeutet], das Gemeindeamt fußläufig erreichen zu können (...) [und dass] mobile Angebote (...) einen Rückzug der Verwaltung aus der Fläche kompensieren können“ (S. 724). Zugleich können sich für Bevölkerungsgruppen, deren Mobilität und Präferenzen für E-Government-Angebote eingeschränkt sind, Nachteile ergeben. Man war sich einig, dass die Bürger persönliche Kontakte (mit Verwaltung und Politik) bevorzugten und diese in der größeren Gemeinde nicht mehr so ausgeprägt möglich seien wie zuvor. Ein Rückgang persönlicher Bindungen bzw. Naheverhältnisse könnte mittel- bis langfristig die Mobilisierung von zivilgesellschaftlichem und politischem Engagement (weiter) erschweren.

Die neuen politischen Kräfteverhältnisse in Anger (ÖVP, SPÖ, FPÖ) und die rückläufige Wahlbeteiligung lassen sich anhand dieser Fallstudie nicht als fusionsbedingte Veränderungen einstufen. Auch wenn – wie zuvor geschildert – Blesse und Rösel (2018) sowie Heinisch, Lehner, Mühlböck und Schimpf (2018) konstatieren, dass (Gemeinde- bzw. Bezirks-)Fusionen zu sinkenden Wahlbeteiligungen und einem Erstarken rechtspopulistischer Parteien führen würden, ist aus Sicht der Autoren hier keine einfache Kausalität anzunehmen, zumal es sich um multifaktoriell beeinflusste Entwicklungen handelt.

Die Einblicke in die Gemeindeverwaltung illustrieren, dass Fusionen herausfordernde Transformationsprozesse für die Mitarbeiter darstellen. Auch ohne Stellenabbau bringen Strukturanpassungen Verunsicherung sowie Umstellungs- bzw. Transformationskosten inkl. Abstimmungsprobleme und Konflikte mit sich. Darüber hinaus könnte es mittelfristig zu einer Arbeitsverdichtung und einer „Überalterung“ der Belegschaft kommen (siehe auch Kuhlmann, Seyfried & Siegel, 2018b, S. 82), wenn nach Pensionierungen vakante Stellen nicht nachbesetzt werden. Dies könnte sowohl für die Mitarbeiterzufriedenheit als auch die Leistungsfähigkeit (tendenziell negative) Folgen haben.

Hinsichtlich der Finanzen ist positiv festzuhalten, dass die neue Gemeinde keine Abgangsgemeinde ist. Da die Gemeinden Anger und Floing ihre Finanzdaten nicht öffentlich zugänglich machen,<sup>9</sup> können keine konkreten Zahlen zu ihrer Finanzsituation dargestellt werden. Jedoch erlaubt die erwähnte Addendum-Analyse einen kompakten Einblick: Demnach lagen in Anger die durchschnittlichen Verwaltungsausgaben pro Kopf in den drei Jahren nach der Fusion (2015-2017) um 12,8% über jenen davor (2012-2014); der deutlichste Anstieg zeigte sich dabei von 2013 auf 2014. Auch in Floing seien die Verwaltungsausgaben seit 2015 gestiegen (Gartner, Hellmann, Hametner & Müller, 2019). Es geht nicht hervor, ob inflationsbedingte Preis- bzw. Ausgabenanstiege berücksichtigt wurden. Nichtsdestotrotz deuten die Analyseergebnisse auf ein verstärktes Ausgabenverhalten vor der Fusion hin (*Common-Pool-Problematik*). Auch hinsichtlich der im Vorfeld angenommenen Effizienzsteigerungen und Einsparungen (bis zu 300.000 Euro jährlich) wird deutlich, dass sich die diesbezüglichen



chen Erwartungen bislang nicht erfüllt haben (außer bei den Ausgaben für Mandatare). Insofern ist Kuhlmann, Seyfried und Siegel (2018b) beizupflichten, dass „die Schaffung von Einsparungsmöglichkeiten nicht mit deren tatsächlicher Nutzung zu verwechseln ist“ (S. 52). Denkbar wäre, dass Einsparungen realisiert und direkt zur Leistungserweiterung oder -professionalisierung (z. B. im Bauhof) verwendet wurden. Für eine fundierte Wirtschaftlichkeitsanalyse bräuchte es nicht nur detailliertes Datenmaterial zum Ressourceneinsatz, sondern auch zur Quantität und Qualität der erstellten Leistungen.

Durch die Fusion(-sprämie) konnten zahlreiche Investitionen in die kommunale Infrastruktur getätigt werden, welche gemäß den Interviewpartnern teilweise dringend vonnöten waren. Dies deutet auf einen „Investitionsstau“ hin. Es ist zu vermuten, dass viele andere (v. a. kleine) steirische Gemeinden unter einem vergleichbaren Investitionsdruck standen. Dies könnte ein möglicher Erklärungsansatz dafür sein, dass die Addendum-Analyse keine Einsparungen feststellen konnte. Relevant dürfte hierbei auch die – speziell in ländlichen Gebieten – schwierige Immobilienverwertung sein. Von den drei aufgelassenen Amtsgebäuden werden zwei gemeindeintern genutzt, eines steht leer; bis zum Erhebungszeitpunkt konnte kein einziges verkauft oder vermietet werden, wodurch weiterhin Kosten anfallen. Angesichts der begrenzten Bebauungsmöglichkeiten erscheint es sinnvoll, nicht mehr benötigte Gebäude bzw. Grundstücke bspw. zur Wohnraumschaffung zu nutzen, um (weiteren) Bevölkerungsrückgängen und Standortnachteilen (wie Schulschließungen) entgegenzusteuern. Es bleibt abzuwarten, inwiefern und inwieweit die politischen Akteure die geschilderten Problembereiche adressieren bzw. lösen (können). Insgesamt wird deutlich, dass die proklamierten Vorteile von Gebietsreformen (insbesondere Einsparungen) „keineswegs ein Selbstläufer [sind]“ (Kuhlmann, Seyfried & Siegel, 2018b, S. 94). Auch wenn die geschilderte Gemeindefusion sicherlich Chancen eröffnet hat, um die Zukunftsfähigkeit der Gemeinde abzusichern, ist offen, ob diese mittel- bis langfristig umfassend genutzt werden können.

## 6 Fazit und Ausblick

Die vor der Fusion geäußerten Erwartungen haben sich weitgehend erfüllt, sowohl bezüglich befürchteter nachteiliger als auch erwarteter vorteilhafter Effekte; jedoch konnten die im Vorfeld erwarteten und vielfach proklamierten Einsparungspotentiale nicht gehoben werden. Problemfelder zeigen sich ferner bei der Gebäudeverwertung, Aufrechterhaltung des zivilgesellschaftlichen Engagements und Gebührenharmonisierung. Trotz des Ausstiegs der Gemeinde Floing und der nicht realisierten Einsparungen zeichnen die gewonnenen Einschätzungen ein überwiegend positives Bild des Fusionsfalles Anger. Wenngleich die Effekte erst mittel- bis langfristig umfassend beurteilt werden können, weisen die kurzfristig vorgenommenen Serviceerweiterungen, Infrastrukturinvestitionen und die Spezialisierung in der Verwaltung auf Verbesserungen zugunsten der Bürger hin. Die Befunde der Literaturanalyse, dass sich mehrheitlich positive Gebietsreformwirkungen auf die Leistungsfähigkeit der Kommunalverwaltung, jedoch heterogene Befunde zu Einsparungen und teils negative Effekte auf die Partizipation und politische Teilhabe zeigen, bestätigen sich tendenziell in Anger. Es bleibt abzuwarten, ob sich künftig noch Einsparungen realisieren lassen. Die in Anger und vielen anderen steirischen Gemeinden zuvor sehr kleinteilige Struktur hätte Einsparun-

gen bzw. Skaleneffekte durchaus erwarten lassen. Es ist zu vermuten, dass die getätigten Investitionen und der dezidiert ausgeschlossene Personalabbau diese (zumindest kurzfristig) konterkariert haben.

Dass der Fusionsprozess in Anger überwiegend als geglückt wahrgenommen wird, dürfte primär auf dessen Freiwilligkeit, individuelle Kontextbedingungen (wie die vorhandene Orientierung hin zum Zentralort sowie zahlreiche Gemeindekooperationen) und die weitgehende Übereinstimmung kommunaler Promotoren der Fusion zurückzuführen sein (v. a. der Bürgermeister, die derselben Partei angehörten und nicht um das Amt konkurrierten). Auch der interviewte Experte betonte, dass ein breiter Konsens auf politischer Ebene zentral für das Gelingen von Fusionen sei. Ergänzt werden diese möglichen Erfolgsfaktoren durch einen gewissen „ökonomischen (Leidens-)Druck“; offenbar sah man durch die Fusion eine Chance gekommen, diesen zu lindern und die Zukunftsfähigkeit der Gemeinde(n) als Lebensmittelpunkt ihrer Bewohner zu bewahren.

Bedenkt man, dass sich die österreichische Gemeindelandschaft trotz der jüngsten Reformen sehr kleinstrukturiert darstellt, so sind in Zukunft weitere Gemeindefusionen in Österreich wahrscheinlich. Praxisorientierte Empfehlungen betreffen die zuvor genannten (potentiellen) Erfolgsfaktoren und Evaluationen: für Entscheidungsträger in den Kommunen sowie auf Landesebene wäre zu empfehlen, die Ziele bzw. Erfolgskriterien weiterer Gebietsreformen im Vorhinein klar und operationalisiert zu definieren und deren Effekte sowohl prozessbegleitend als auch ex-post zu evaluieren. Dabei sollten aus Sicht der Verfasser in Zukunft die (Landes-)Rechnungshöfe eine essenzielle Rolle übernehmen.<sup>10</sup> Bei der Festlegung von Reformzielen erscheint es zentral, dass Zielkonflikte reflektiert und offen diskutiert werden. So deuten bspw. die gewonnenen Einblicke auf einen Zielkonflikt zwischen wirtschaftlichen Desiderata und dem Erhalt des ländlichen Raums sowie des bürgerschaftlichen Engagements hin. Sofern politische Entscheidungsträger den ländlichen Raum fördern und Zentralisierungs- bzw. Urbanisierungstendenzen im Sinne einer „Landflucht“ entgegenwirken wollen, wird dies mit Kosten und – über Fusionen hinausgehenden – Reformanstrengungen verbunden sein.

Hinsichtlich der Limitationen dieser Untersuchung ist zum einen der regionale Fokus auf die Steiermark und die betrachtete(n) Fusionsgemeinde(n) zu nennen. Die Erkenntnisse sind nur mit Vorsicht auf andere Gemeinden übertragbar, da diese spezifische Kontext- und Reformdurchführungsbedingungen aufweisen (können). Zum anderen ist ein Antwortverhalten im Sinne sozialer Erwünschtheit anzunehmen; die subjektiven Erwartungen und Einschätzungen der Gesprächspartner dürften tendenziell positive Effekte stärker betonen als problembehaftete Fusionsauswirkungen. Die dennoch thematisierten Herausforderungen verweisen umso mehr auf kritische Aspekte, welche auch für andere Gemeinden relevant sein könnten und denen man in künftigen Forschungsarbeiten besonderes Augenmerk schenken sollte. Weiterführend könnte man durch eine Befragung der Bürger oder von ortsansässigen Unternehmen vertiefte Einblicke in die Auswirkungen der Fusion in Anger gewinnen. Studien zur Wahrnehmung von Fusionseffekten aus Bürgersicht sind bislang selten (Tavares, 2018, p. 11). Zudem erscheinen umfassende Wirkungsanalysen zu den (mittel- bis langfristigen) Effekten der Gemeindestrukturereform Steiermark sowie vergleichende Untersuchungen mit den oberösterreichischen Gemeindefusionen sinnvoll, um über Tendenzaussagen hinausreichende Erkenntnisse zu den Effekten von Strukturreformen in Österreich ableiten zu

können. Dabei sollte man speziell potenziellen Problemfeldern und Erfolgsfaktoren in der Vorbereitung und Umsetzung von (freiwilligen sowie erzwungenen) Fusionen Aufmerksamkeit widmen. Lohnend erscheinen auch begleitende Untersuchungen mit Blick auf die Veränderungs- bzw. Umstellungskosten von Fusionen sowie Dokumentenanalysen von Rechnungsabschlüssen, um die ökonomischen Auswirkungen besser beurteilen zu können. Denkbar wäre für die Steiermark auch eine quasi-experimentelle Studie auf Basis der (mittlerweile häufig angewandten) *difference-in-difference*-Methodik. Um den Effekt einer Maßnahme (Gemeindestrukturreform) zu schätzen, wird dabei die Veränderung einer Variable (z. B. Pro-Kopf-Verwaltungsausgaben oder Wahlbeteiligung) für eine Versuchsgruppe (z. B. alle 2015 fusionierten Gemeinden) und eine Kontrollgruppe (nicht fusionierte Kommunen) zwischen zwei Zeitpunkten (vor und nach der Maßnahme) beobachtet (Lüchinger & Stutzer, 2002, S. 36; Swianiewicz, 2018, p. 5). Angesichts der intensivierten Reformdiskussionen und -initiativen gibt es jedenfalls in Österreich zunehmend Gelegenheit und Bedarf, empirisch fundiertes Wissen über die Wirkungen und Nebenwirkungen von Gemeindefusionen zu gewinnen, abseits pauschaler Annahmen und teils ideologisch gefärbter Behauptungen.

## Anmerkungen

- 1 Die Verfasser danken den anonymen Gutachter/innen für ihre umfassenden und hilfreichen Überarbeitungshinweise.
- 2 Zugunsten einer besseren Lesbarkeit wird auf eine gleichzeitige Angabe weiblicher und männlicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für alle Geschlechter.
- 3 Abgangsgemeinden sind Gemeinden, die ihren ordentlichen Haushalt nicht mit den eigenen Einnahmen ausgleichen können; d. h. das Ergebnis ihres ordentlichen Haushalts ist negativ (o. V., 2017, S. 42).
- 4 Die Erkenntnisse und Entscheidungstexte des VfGH können unter <https://www.ris.bka.gv.at/Vfgh/> abgerufen werden. Der VfGH hielt fest, dass keine grundsätzlichen verfassungsrechtlichen Bedenken gegen die Fusionen sprechen, der Gesetzgeber weitreichenden rechtspolitischen Gestaltungsspielraum bei der Änderung der Gemeindegebiete habe und einzelne Gemeinden kein Recht auf „ungestörte Existenz“ hätten.
- 5 Dabei führte man Berechnungen für die Jahre 2012 bis 2017 auf Basis von Rechnungsabschlussdaten der Statistik Austria durch. Eine Einzelpfung von Rechnungsabschlüssen der Gemeinden war nicht möglich. Der Fokus lag auf den Ausgaben im ordentlichen Haushalt bzw. den „Verwaltungskosten“, worunter Personalkosten (in der Verwaltung, aber nicht in Kindergärten usw.) und Sachkosten (Miete, Energie, Verbrauchsmaterial etc. für allgemeine Verwaltungseinrichtungen wie Gemeindeämter oder Bauhöfe (aber ohne Kindergärten, Schulen usw.)) subsumiert wurden. Damit nicht schrumpfende Einwohnerzahlen zu steigenden Pro-Kopf-Kosten führen, wurde die Einwohnerzahl am Stand von 2010 fixiert. Bei den Berechnungen wurde simuliert, dass die Strukturreform bereits 2010 erfolgte; man summierte die Ausgaben der jeweiligen Fusionsgemeinden vor deren Zusammenschluss und verglich sie ab 2015 mit den Ausgaben der neuen Gemeinde (Gartner, Hellmann, Hametner & Müller, 2019).
- 6 Österreich gliedert sich in 94 politische Bezirke (15 Statutarstädte, 79 Landbezirke). Bezirkshauptmannschaften sind als Verwaltungsbehörden für alle Gemeinden in ihrem Bereich zuständig (Statistik Austria, 2019b).
- 7 In österreichischen Gemeinderäten bezeichnet man die jeweiligen Gemeinderatsfraktionen als Gemeinderatsklubs. Diese Gruppen von Abgeordneten werden von Klubobleuten (d. h. Fraktionsvorsitzenden) geführt.
- 8 Aufgrund des abgestuften Bevölkerungsschlüssels erhalten größere Gemeinden mehr Mittel aus dem Finanzausgleich, da größere Kommunen höheren Infrastrukturanforderungen nachkommen müssen (Schauer, 2019, S. 137).
- 9 Anger sowie die Mehrzahl aller steirischen Gemeinden geben ihre Finanzdaten bislang nicht für den öffentlich zugänglichen Bereich auf der Plattform „Offener Haushalt“ (KDZ – Zentrum für Verwaltungsforschung, 2019) frei. Ziel dieser Plattform ist es, Voranschläge und Rechnungsabschlüsse der

- österreichischen Gemeinden transparent zu machen und übersichtlich darzustellen. Nur 66 steirische Gemeinden (und 1.008 Gemeinden insgesamt in Ö) hatten im August 2018 ihre Daten freigeschaltet (Hödl & Krabina, 2018, S. 24 f.).
- 10 Aktuell ist es weder dem Steiermärkischen Landesrechnungshof noch dem (Bundes-)Rechnungshof möglich, eine flächendeckende Analyse der Gemeindestrukturreform durchzuführen. Der Steiermärkische Landesrechnungshof darf Gemeinden seit dem 1.6.2015 prüfen, wodurch sich die Prüfbefugnis nicht auf den Zeitraum vor der Gemeindestrukturreform erstreckt (Gartner, Hellmann, Hametner & Müller, 2019).

## Literatur

- Amt der Steiermärkischen Landesregierung (2015a). *Die neue Steiermark ab 2015*. Graz-Burg, Verfügbar unter: [http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11558564\\_69294774/9b5c4fc/Brosch%C3%BCre.pdf](http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11558564_69294774/9b5c4fc/Brosch%C3%BCre.pdf) [25.10.2018].
- Amt der Steiermärkischen Landesregierung (2015b). *Gemeinderatswahl vom 22. März 2015*. Verfügbar unter: [https://egov.stmk.gv.at/wahlen/GR2015/index\\_GR2015.html](https://egov.stmk.gv.at/wahlen/GR2015/index_GR2015.html) [25.10.2018].
- Allers, Maarten A. & Geertsema, J. Bieuwe (2016). The effects of local government amalgamation on public spending, taxation, and service levels: Evidence from 15 years of municipal consolidation. *Journal of Regional Science*, 56 (4), 659-682.
- Askim, Jostein, Klausen, Jan Erling, Vabo, Signy Irene & Bjurström, Karl (2016). What Causes Municipal Amalgamation Reform? In Sabine Kuhlmann & Geert Bouckaert (Eds.). *Local Public Sector Reforms in Times of Crisis. National Trajectories and International Comparisons* (pp. 59-79). London: Palgrave Macmillan Publishers Ltd.
- Askim, Jostein, Klausen, Jan Erling, Vabo, Signy Irene & Bjurström, Karl (2017). Territorial upscaling of local governments: a variable-oriented approach to explaining variance among Western European countries. *Local Government Studies*, 43 (4), 555-576.
- Auer, Rainer (2017). *Gemeindefusionen als Chance für alle?* Bericht vom 11.1.2017. Verfügbar unter: [https://www.meinbezirk.at/grieskirchen-eferding/c-politik/gemeindefusionen-als-chance-fuer-alle\\_a1985909](https://www.meinbezirk.at/grieskirchen-eferding/c-politik/gemeindefusionen-als-chance-fuer-alle_a1985909) [25.10.2018].
- Blesse, Sebastian & Rösel, Felix (2018). Merging county administrations – cross-national evidence of fiscal and political effects. *Local Government Studies*, 44, 1-21. <https://doi.org/10.1080/03003930.2018.1501363>
- Blom-Hansen, Jens, Houlberg, Kurt, Serritzlew, Søren & Treisman, Daniel (2016). Jurisdiction Size and Local Government Policy Expenditure: Assessing the Effect of Municipal Amalgamation. *American Political Science Review*, 110 (4), 812-831.
- Bouckaert, Geert & Kuhlmann, Sabine (2016). Introduction: Comparing Local Public Sector Reforms: Institutional Policies in Context. In Sabine Kuhlmann & Geert Bouckaert (Eds.). *Local Public Sector Reforms in Times of Crisis* (pp. 1-20). London: Palgrave Macmillan Publishers Ltd.
- Dunst, Martin (2014). *Für viele Einwohner ist der Ortsname wichtig*. Bericht vom 9.9.2014. Verfügbar unter: <http://www.nachrichten.at/oberoesterreich/Fuer-viele-Einwohner-ist-der-Ortsname-wichtig;art4,1494230> [4.10.2014].
- Fallend, Franz, Mühlböck, Armin & Wolfgruber, Elisabeth (2001). Die österreichische Gemeinde. In Forum Politische Bildung (Hrsg.). *Regionalismus, Föderalismus, Supranationalismus* (S. 45-61). Wien/Innsbruck: Studien-Verlag.
- Gartner, Gerald, Hellmann, Gabriel, Hametner, Markus & Müller, Monika (2019). *Drei Jahre danach: Kaum Einsparungen durch die Gemeindereform*. Bericht vom 27.1.2019. Verfügbar unter: <https://www.addendum.org/gemeindefusionen/kaum-einsparungen/> [25.4.2019].
- Gläser, Jochen & Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse* (4. Aufl.). Wiesbaden: VS Verlag.
- Gödl, Ernst (2013). *Die Gemeindestrukturreform in der Steiermark*. Zwarung-Pöls: Eigenverlag.

- Hecke, Bernd & Rossacher, Thomas (2019). *Bilanz der Gemeindereform. Was die steirischen Gemeindefusionen gebracht haben*. Bericht vom 27.1.2019. Verfügbar unter: [https://www.kleinezeitung.at/steiermark/chronik/5567329/Bilanz-der-Gemeindereform\\_Was-die-steirischen-Gemeindefusionen](https://www.kleinezeitung.at/steiermark/chronik/5567329/Bilanz-der-Gemeindereform_Was-die-steirischen-Gemeindefusionen) [30.1.2019].
- Heinisch, Reinhard, Lehner, Thomas, Mühlböck, Armin & Schimpf, Christian H. (2018). How do municipal amalgamations affect turnout in local elections? Insights from the 2015 municipal reform in the Austrian state of Styria. *Local Government Studies*, 44 (4), 465-491.
- Hödl, Clemens & Krabina, Bernhard (2018). Fünf Jahre Offenerhaushalt.at. *Österreichische Gemeindezeitung (ÖGZ)*, Heft 10, 24-25.
- Hörmann, Hans-Jörg (2016). Die Umsetzung der Gemeindegebietsreform Steiermark 2015. *Österreichische Gemeindezeitung (ÖGZ)*, Heft 10, 19-20.
- Kaiser, Claire (2015). Top-down versus Bottom-up: Comparing Strategies of Municipal Mergers in Western European Countries. *dms – der moderne staat*, 7 (1), 113-127.
- KDZ – Zentrum für Verwaltungsforschung (2019). *Offener Haushalt. Finanzdaten aus dem öffentlichen Sektor in Österreich – Anger*. Verfügbar unter: <https://www.offenerhaushalt.at/gemeinde/anger> [25.4.2019].
- Krenn-Aichinger, Michaela (2018). *Eferdings Bürgermeister Mair will die Bürger über Fusion abstimmen lassen*. Bericht vom 27.9.2018. Verfügbar unter: <https://www.nachrichten.at/oberoesterreich/wels/Eferdings-Buergermeister-Mair-will-die-Buerger-ueber-Fusion-abstimmen-lassen;art67,3018082> [25.10.2018].
- Kuhlmann, Sabine, Seyfried, Markus & Siegel, John (2018a). Was bewirken Gebietsreformen? Eine Bilanz deutscher und europäischer Erfahrungen. *dms - der moderne staat*, 11 (1), 119-141.
- Kuhlmann, Sabine, Seyfried, Markus & Siegel, John (2018b). Wirkungen kommunaler Gebietsreformen. Stand der Forschung und Empfehlungen für Politik und Verwaltung. Baden-Baden: Nomos Verlagsgesellschaft.
- Land Steiermark (2011). *Steirische Gemeindestrukturreform – Ausgangslage*. Verfügbar unter: [http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11558627\\_69294537/de4516e8/Gemeindestrukturreform%20Ausgangslage.pdf](http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11558627_69294537/de4516e8/Gemeindestrukturreform%20Ausgangslage.pdf) [14.1.2013].
- Land Steiermark (2012). Gemeindestrukturreform Steiermark – Leitbild: Stärkere Gemeinden – Größere Chancen. Verfügbar unter: [http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11615233\\_100167017/525d4b39/GSR\\_Leitfaden\\_30012012\\_Vs2.pdf](http://www.gemeindestrukturreform.steiermark.at/cms/dokumente/11615233_100167017/525d4b39/GSR_Leitfaden_30012012_Vs2.pdf) [14.1.2013].
- Lüchinger, Simon & Stutzer, Alois (2002). Skalenerträge in der öffentlichen Kernverwaltung. Eine empirische Analyse anhand von Gemeindefusionen. *Swiss Political Science Review*, 8 (1), 27-50.
- Mayring, Philipp (2010). *Qualitative Inhaltsanalyse* (11. Aufl.). Weinheim u.a.: Beltz.
- OÖNachrichten (2017). *Drei Gemeindefusionen für 2019 geplant*. Bericht vom 8.5.2017. Verfügbar unter: <https://www.nachrichten.at/oberoesterreich/Drei-Gemeindefusionen-fuer-2019-geplant;art4,2560819> [25.10.2018].
- Österreichischer Gemeindebund (2019). *Struktur der Gemeinden*. Verfügbar unter: <https://gemeindebund.at/struktur-der-gemeinden/> [5.4.2019].
- o. V. (2017). Sehr gutes Zeugnis für die Gemeinden. Gemeindefinanzbericht 2017. *Kommunal*, Heft 12, 39-43.
- Pitlik, Hans & Wirth, Klaus (2012). Gemeindestrukturreformen und Gemeindekooperation. *WIFO Monatsberichte*, Heft 12, 957-967.
- Pitlik, Hans, Wirth, Klaus & Lehner, Barbara (2010). *Gemeindestruktur und Gemeindekooperation*. Wien: WIFO & KDZ.
- Prettenthaler, Franz, Kernitzkyi, Michael & Kurzmann, Raimund (2011). Gemeindestrukturreform in der Steiermark. Eine modellbasierte Schätzung des Einsparpotenzials und Bewertung der makroökonomischen Effekte. Graz: Joanneum Research.
- Schauer, Reinbert (2019). *Öffentliche Betriebswirtschaftslehre – Public Management* (4. Aufl.). Wien: Linde Verlag.
- Seuberlich, Marc (2017). Warum fusionieren Gemeinden freiwillig? *Verwaltung & Management*, Heft 1, 26-32.

- Stangl, Michaela & Stötzer, Sandra (2015). Gemeindestrukturreformen in Österreich – Auswirkungen aus Sicht der Praxis. *Verwaltung & Management*, Heft 2, 89-95.
- Statistik Austria (2013a). *Ein Blick auf die Gemeinde Anger*. Verfügbar unter: <http://www.statistik.at/blickgem/blick1/g61702.pdf> [14.1.2013].
- Statistik Austria (2013b). *Ein Blick auf die Gemeinde Baierdorf bei Anger*. Verfügbar unter: <http://www.statistik.at/blickgem/blick1/g61704.pdf> [14.1.2013].
- Statistik Austria (2013c). *Ein Blick auf die Gemeinde Feistritz bei Anger*. Verfügbar unter: <http://www.statistik.at/blickgem/blick1/g61707.pdf> [14.1.2013].
- Statistik Austria (2013d). *Ein Blick auf die Gemeinde Naintsch*. Verfügbar unter: <http://www.statistik.at/blickgem/blick1/g61732.pdf> [14.1.2013].
- Statistik Austria (2019a). *Gemeindegrößenklassen mit der Einwohnerzahl 2019*. Verfügbar unter: [https://www.statistik.at/web\\_de/klassifikationen/regionale\\_gliederungen/gemeinden/index.html](https://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/gemeinden/index.html) [10.7.2019].
- Statistik Austria (2019b). *Politische Bezirke*. Verfügbar unter: [https://www.statistik.at/web\\_de/klassifikationen/regionale\\_gliederungen/politische\\_bezirke/index.html](https://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/politische_bezirke/index.html) [4.2.2019].
- Steiner, Reto (1999). Kooperationen und Fusionen der Gemeinden in der Schweiz. Konzeptionelle Grundlagen und Ergebnisse aus einer empirischen Studie. Bern: Institut für Organisation und Personal der Universität Bern.
- Steiner, Reto, Kaiser, Claire & Eythórsson, Grétar Thór (2016). A Comparative Analysis of Amalgamation Reforms in Selected European Countries. In Sabine Kuhlmann & Geert Bouckaert (Eds.). *Local Public Sector Reforms in Times of Crisis* (pp. 23-42). London: Palgrave Macmillan Publishers Ltd.
- Swianiewicz, Pawel (2018). If territorial fragmentation is a problem, is amalgamation a solution? – Ten years later. *Local Government Studies*, 44 (1), 1-10.
- Tavares, Antonio F. (2018). Municipal amalgamations and their effects: a literature review. *Miscellanea Geographica – Regional Studies on Development*, 22 (1), 5-15.
- Wirth, Klaus (2011). Strukturreformen bei Gemeinden. *Wirtschaftspolitische Blätter*, Nr. 4, 717-728.

### *Anschriften der Autor\*innen*

Priv.-Doz. Dr. Sandra Stötzer, Johannes Kepler Universität Linz, Institut für Public und Nonprofit Management, Altenberger Straße 69, A-4040 Linz;  
E-Mail: [sandra.stoetzer@jku.at](mailto:sandra.stoetzer@jku.at)

Mag. Michaela Stangl, Johann-Böhm-Straße 17, A-4860 Lenzing;  
E-Mail: [michaelastangl@gmx.at](mailto:michaelastangl@gmx.at)

Mag. Christian Brödl, Kudlichstraße 18, A-3150 Wilhelmsburg;  
E-Mail: [broedlch@hotmail.com](mailto:broedlch@hotmail.com)

*Manfred Röber*

Sack, Detlef (2019). *Vom Staat zum Markt. Privatisierung aus politikwissenschaftlicher Sicht*. Wiesbaden: VS Verlag, XVII/379 Seiten. ISBN 978-3-658-26872-5; ISBN 978-3-658-26873-2 (eBook); <https://doi.org/10.1007/978-3-658-26873-2>.

### Abstract

The book by Detlef Sack provides a profound overview of the topic of “privatisation” from a political science perspective. First of all, different forms of privatisation are presented. Then the privatisation topic is placed in the context of political and economic theories. After an overview of the history of privatisation in Germany, Europe and selected countries in Africa and Asia, the author describes and analyses privatisation projects in various policy fields. He then points out the need for an effective regulatory regime for the privatisation of public tasks and services. In a next step, the economic, social and power-political effects of privatisation are outlined. Finally, current reactions to privatisation in the form of remunicipalisation and perspectives for further privatisation research in political science are presented. The book is characterized by a pleasant objectivity in dealing with a politically and scientifically often very controversially discussed topic and can therefore be recommended for reading to anyone interested in this subject.

Detlef Sack hat mit seinem Buch *“Vom Staat zum Markt. Privatisierung aus politikwissenschaftlicher Perspektive”* einen bedeutenden Beitrag zur Diskussion über ein wichtiges Thema vorgelegt, das sowohl wissenschaftlich als auch politisch nach wie vor kontrovers diskutiert wird. Die häufig die Diskussion dominierenden wirtschafts- und rechtswissenschaftlichen Diskurse werden um einen bislang weitgehend fehlenden fundierten Beitrag aus politikwissenschaftlicher Perspektive ergänzt und bereichert. Die Zielsetzung, die mit diesem Buch verfolgt wird, besteht darin, „Privatisierung (als ein Synonym für die Übertragung von Verfügungsrechten vom Staat und von öffentlichen Gebietskörperschaften auf gewerbliche Unternehmen) in ihren unterschiedlichen Facetten darzustellen und politikwissenschaftlich zu erklären“ (Sack, 2019, S. 7). Mit diesem Anspruch unterscheidet sich das Buch von Anfang an wohlthuend von normativen oder krypto-normativ eingefärbten Glaubensbekenntnissen für oder gegen Privatisierungen.

Das Buch ist in Bezug auf das Thema sinnvoll in insgesamt 14 Kapitel untergliedert. Nach dem einleitenden 1. Kapitel geht es im 2. Kapitel um für das Thema wichtige begriffliche und konzeptionelle Klärungen von unterschiedlichen Erscheinungsformen der Privatisierung (formal, materiell, funktional) und um Verfahrensoptionen, wie der Prozess der materiellen und funktionalen Privatisierung organisiert werden kann. Im 3. Kapitel widmet sich der Verfasser der vor allem in den Wirtschaftswissenschaft-

ten geführten Debatte über die Frage, was das besondere Charakteristikum öffentlicher Güter ist, und damit implizit auch der Frage, inwiefern sich öffentliche Güter von privaten Gütern unterscheiden. Das 4. Kapitel enthält eine sehr kurze und relativ allgemein gehaltene Darstellung von Methoden, die in der politikwissenschaftlichen Privatisierungsforschung verwendet werden. Im 5. Kapitel wird das Thema „Privatisierung“ in den politik- und staatswissenschaftlichen Kontext, im 6. Kapitel dann in den Kontext ökonomischer Theorien (insbesondere der Public-Choice-Theorie und der Institutionenökonomie) eingeordnet. Das 7. Kapitel bezeichnet Sack selbst als „das konzeptionell-methodische Herzstück des theoretischen Abschnittes“ (Sack, 2019, S. 15), weil hier Ansätze präsentiert werden, die Aufschluss über die Gründe für Privatisierungen und Hinweise für politikwissenschaftliche Erklärungen geben. Dem schließt sich im 8. Kapitel ein historischer Überblick über die Entwicklung der Privatisierungen während der letzten ca. 50 Jahre in Deutschland an, der dann in weiteren Abschnitten dieses Kapitels um historische Darstellungen der Privatisierung in Europa (mit Schwerpunkt auf Großbritannien seit 1979 und auf die mittel- und osteuropäischen Transformationsländer seit 1989) erweitert wird. Danach wird im 9. Kapitel der Blick auf Entwicklungen der Privatisierung im globalen Kontext in Bezug auf sich entwickelnde Länder in Lateinamerika, Afrika und Asien gerichtet. Nach den regional ausgerichteten Überblicken widmet sich der Verfasser im 10. Kapitel der Privatisierung in einzelnen Politikfeldern, wobei er sich auf Fragen der Privatisierung beim staatlichen Gewaltmonopol, bei netzgebundenen Infrastrukturen und im Krankenhausbereich konzentriert. Da Entstaatlichung nicht bedeuten kann, die Wahrnehmung öffentlicher Aufgaben und die Erbringung öffentlicher Leistungen einfach privaten Akteuren zu überlassen und damit der Gefahr ausgesetzt zu sein, dass aus öffentlichen Angebotsmonopolisten private Monopole oder Oligopole entstehen, bedarf es – sowohl bei Entscheidungsprozessen über die Privatisierung als auch bei der sich daran anschließenden Leistungsproduktion – eines wirksamen Regulierungsregimes. In diesem Zusammenhang beschäftigt sich Sack im 11. Kapitel mit komplementären und kompensatorischen Formen der Regulierung sowie den Gründen, die für die Implementierung von Regulierungsorganisationen und -mechanismen sprechen. Folgerichtig geht es im 12. Kapitel dann um die ökonomischen, aber auch sozialen und machtpolitischen Effekte von Privatisierungen sowie um die daraus resultierenden Konsequenzen für die politikwissenschaftliche Privatisierungsforschung. Vor dem Hintergrund einer zunehmenden Privatisierungsskepsis beschäftigt sich der Verfasser im 13. Kapitel mit den aktuellen Trends der Re-Kommunalisierung. Das Buch endet mit einem ausführlichen Fazit im 14. Kapitel, in dem Erklärungsfaktoren und Kernaussagen der politikwissenschaftlichen Privatisierungsforschung differenziert nach funktionalistischen, diskursanalytischen und institutionalistischen Theorien sowie Akteurstheorien skizziert, Verbindungen zwischen Privatisierung und sich verändernden Formen politischer Herrschaft beleuchtet und Perspektiven der weiteren politikwissenschaftlichen Privatisierungsforschung aufgezeigt werden.

Sack ist es gelungen, einen ausgezeichneten und lesenswerten Überblick über Entwicklungstendenzen der Privatisierung und über unterschiedliche Facetten der Privatisierungsdiskussion zu geben, der nicht nur für Politikwissenschaftlerinnen und Politikwissenschaftler informativ und aufschlussreich ist. Das Buch bietet sowohl denjenigen, die sich zum ersten Mal mit dieser Thematik beschäftigen wollen, als auch den Kennern der Privatisierungsdiskussion eine Fülle von wichtigen und interessanten Informationen und Einsichten.



In Bezug auf einige Komponenten des Buches gibt es allerdings ein paar kritische Fragen und Anmerkungen. So ist zunächst überraschend, dass der Verfasser darauf verzichtet hat, seine Reflexionen und Ausführungen explizit in das Modell des Gewährleistungsstaates einzuordnen und das konzeptionelle Potenzial, das mit diesem Modell für die Privatisierungsdiskussion verbunden ist, zu nutzen. In diesem Kontext hätte das Thema „Privatisierung“ zum Beispiel sehr gut in die Überlegungen zur institutionellen Differenzierung und in den Kanon institutioneller Wahlmöglichkeiten eingeordnet werden können, die Staat und Kommunen bei der Wahrnehmung öffentlicher Aufgaben zur Verfügung stehen. In diesem Zusammenhang fällt auch eine unzureichende Abgrenzung von öffentlichen Aufgaben und öffentlichen Gütern auf. Dies führt zu einer konzeptionellen Inkonsistenz, weil auf der einen Seite in den Ausführungen über öffentliche Güter im 3. Kapitel der Eindruck erweckt wird, dass es objektive Bestimmungsmerkmale für die Abgrenzung dieser Güter von den privaten Gütern gibt, und weil der Verfasser auf der anderen Seite mehrmals in seiner Arbeit ausdrücklich und zurecht darauf hinweist, dass über das, was öffentliche Aufgabe ist, nur politisch in den dafür zuständigen und legitimierten Gremien entschieden werden kann. Insofern sind die Passagen über öffentliche Güter politikwissenschaftlich wenig zielführend, weil sie keinen Zugang für die Analyse von Handlungsspielräumen interessen- und machtleiteter Akteurskonstellationen eröffnen.

Wenn es um Themen geht, die in den wirtschaftswissenschaftlichen Bereich hineinragen, dann enthält der Text einige problematische Passagen und Ungenauigkeiten. Problematisch sind beispielsweise die fehlende Unterscheidung von Wettbewerbs- und Privatisierungsfähigkeit öffentlicher Unternehmen, die hinsichtlich ihrer Marktposition und ihres Leistungsportfolios erhebliche Unterschiede aufweisen, sowie die undifferenzierte Gleichsetzung von Ökonomisierung und Profitmaximierung, die auf einem weit verbreiteten und bedauerlichen Missverständnis des nicht nur auf die kapitalistische Marktwirtschaft bezogenen betriebswirtschaftlichen Erkenntnisinteresses beruht. Zu den Ungenauigkeiten gehört die an einigen Stellen der Arbeit unpräzise und zu Missverständnissen führende Verwendung von inhaltlich eindeutigen betriebswirtschaftlichen Begriffen (wie beispielsweise Ausgaben, Kosten, Einnahmen, Erlöse und Rentabilität). Aus solchen Fehlinterpretationen und Ungenauigkeiten können sehr schnell – wie einige der politischen und populärwissenschaftlichen Diskussionsbeiträge zeigen – verzerrte und falsche Schlussfolgerungen hinsichtlich des Erkenntnispotenzials wirtschaftswissenschaftlicher Analysen und der Konsequenzen betriebswirtschaftlich sinnvoller Entscheidungen resultieren.

Schließlich wirkt die Auswahl der Länder, auf die mit Informationen und Fallbeispielen Bezug genommen wird und die von Deutschland über Großbritannien, Mittel- und Osteuropa, Russland, Nigeria, Ghana und Lateinamerika bis nach China reicht, zu ambitioniert und eklektisch. Angesichts der Tatsache, dass sich die politischen, ökonomischen, gesellschaftlichen und technologischen Rahmenbedingungen in den erwähnten Ländern gravierend unterscheiden, ist fraglich, welche vergleichenden Erkenntnisse zum Beispiel aus der Wasserprivatisierung in Ghana und der Krankenhausprivatisierung in Hamburg gewonnen werden können. Gerade aus politikwissenschaftlicher Sicht wäre es interessant und dringend geboten, die Rolle und Funktion von Privatisierungen in demokratischen Industrieländern, Transformationsländern, kommunistischen Regimen und Entwicklungsländern, die überdies auch ganz unterschiedlichen Kulturkreisen zuzuordnen sind, viel grundsätzlicher zu thematisieren und in Bezug auf die politischen Systeme dieser Länder präziser zu analysieren.

Insgesamt ändern die kritischen Hinweise aber nichts an dem sehr positiven Urteil über das Buch, das aufgrund der tiefen Sachkenntnis des Verfassers und seiner abgewogenen Argumentation zur Versachlichung einer zum Teil emotional aufgeladenen Privatisierungsdiskussion beitragen kann und dem deshalb eine weite Verbreitung zu wünschen ist.

*Anschriften des Autors:*

Prof. (em.) Dr. Manfred Röber, Universität Leipzig, Professur für Public Management, Universitätsstr. 16, 04109 Leipzig, E-Mail: roeber@wifa.uni-leipzig.de; Andrassy-Universität Budapest, Professur für Betriebswirtschaftslehre, H-1088 Budapest, Pollack Mihály tér 3, E-Mail: manfred.roeber@andrassyuni.hu.

*Werner Jann*

Falk, Svenja, Glaab, Manuela, Römmele, Andrea, Schober, Henrik & Thunert, Martin (Hrsg.) (2019). *Handbuch Politikberatung*. Wiesbaden: VS Verlag. 623 Seiten. ISBN: 978-3-658-03482-5.

### Abstract

After 13 years, this is the second edition of the *Handbuch der Politikberatung* – and the first problem is the translation of the title. The typical English translation would be Handbook of Policy Advice, but the term in Germany is used much more loosely, covering also all kinds political advice, i. e. public affairs, public relations, polls and survey research etc., and even what is called “participative advice” of and by citizens, non-governmental organizations etc. In 39 contributions by more than 50 authors, disciplinary and theoretical aspects are covered, but especially all kinds of actors, both looking for and supplying this kind of advice are presented or present themselves. The handbook offers a good overview about the German scene, but it does not quite succeed in its aim to clarify and enhance the academic debate. Many contributions are by actors directly involved in this field, which offers interesting insights, but is something different than critical and detached analysis.

Politikberatung ist ein klassisches Thema der Sozialwissenschaften, weil sie ja doppelt betroffen sind: Zum einen wollen sie verstehen und erklären, ob und wie Politikberatung zustande kommt und wirkt, und gleichzeitig sind sie auch als Politikberater gefragt, oder würden jedenfalls gern gefragt werden. In den letzten Jahren ist Politikberatung aber auch zunehmend ein „Modethema“ geworden, weil alle und alles sich als Politikberater deklarieren und verkaufen, oder dies zumindest versuchen. Politikberater ist keine geschützte Berufsbezeichnung. Sven Siefken hatte daher schon vor ein paar Jahren ironisch gefragt: „Ist denn alles Politikberatung?“, und vor einer inflationären und opportunistischen Verwendung des Begriffs gewarnt (Siefken, 2010).

Es ist daher nur zu begrüßen, dass das „*Handbuch Politikberatung*“ nach 13 Jahren endlich in einer „völlig neu bearbeiteten Auflage“ erschienen ist. In der Einführung skizzieren die Herausgeber<sup>1</sup> ihre Ziele: „Gerade weil Politikberatung eine dezidiert praktische Betätigung ist, die auf vielfältige Arten ausgeübt werden kann, benötigt dieses Feld begleitende Forschung, die zur Verortung und Versachlichung der Debatte beitragen, Trends skizzieren, Fragen aufwerfen und Ausblicke geben kann. Dazu möchte der vorliegende Sammelband einen Beitrag leisten“ (Falk, Glaab, Römmele, Schober & Thunert, 2019, S. 20). Zu fragen ist, ob und wie diese Ziele erreicht werden.

Die Einleitung versucht einen ersten Überblick über Begrifflichkeiten, Schwerpunkte und Trends der aktuellen Debatten zu geben. Hier werden viele grundlegende

Fragen kurz angerissen, also z. B. über die Vielfalt und Unklarheiten des Konzepts, die Unterscheidung der Beratung von Entscheidungsträgern oder Öffentlichkeits- und Gesellschaftsberatung, von *Policy*-, *Politics*- und *Polity*-Beratung, und es gibt auch eine erste hilfreiche Übersicht über Anbieter und Nachfrager in zentralen Beratungsfeldern, aber leider verbleibt das eher eklektisch. Eine Synthese, die z. B. auch die Struktur des Bandes und die zentralen Fragen veranschaulichen könnte, wird nicht gegeben, und auch die Frage, was denn Politikberatung ist, und was nicht, verbleibt offen. Die gewählte Gliederung, so heißt es, sei nur „eine von vielen möglichen Antworten“ auf die Frage der Darstellung des sich immer weiter differenzierenden Bereichs. Die Forschungslandschaft wird als „eher unübersichtlich“ bezeichnet, „ersichtlich wird die Vielfalt der Zugänge und Fragestellungen, die im vorliegenden Band vielfach aufgegriffen, aber nicht vollumfänglich behandelt werden können“, und es sei „im Rahmen eines Handbuches dieser Art nicht möglich [...], alle relevanten theoretischen Zugänge und Akteure zu behandeln“. Das ist alles sehr defensiv. Als Zukunftsaufgabe der nächsten Jahre wird „die Verbesserung der Kartographie des Fachs“ angemahnt – ehrlich gesagt hätte ich das eigentlich von der Neuauflage des Handbuchs, das immerhin 39 Beiträge von über 50 Autoren enthält, erwartet.

Der erste Teil „*Theoretische Grundlagen und Zugänge*“ liefert dann aber doch hilfreiche wissenschaftliche Einordnungen, z. B. aus juristischer und ethischer Sicht und insbesondere die Beiträge von Andreas Blätte zu unterschiedlichen sozialwissenschaftlichen Perspektiven und von Peter Weingart zum Verhältnis von Wissenschaft und Politikberatung sind – wie immer – lesenswert und liefern einen guten Überblick. Wenn es aber z. B. im Beitrag über „*demokratische Teilhabe in einer digitalen Gesellschaft*“ vor allem um direkte und repräsentative Demokratie, um Bürgerbeteiligung und Partizipation geht, verschwindet jegliche Differenz zwischen Demokratie und Politikberatung. Wer u. a. Bürgerbeteiligung und Repräsentation als Politikberatung behandelt, hat offenbar eine merkwürdige Auffassung von Demokratie. Hier rächt sich, dass die Herausgeber darauf verzichten, klar zu sagen, was für sie Politikberatung ist, und was nicht.

Auch der zweite Teil „*Ressourcen und Akteure*“ liefert wiederum viele informative Übersichten, u. a. Siefken über Expertenkommission oder Patrick Bernhagen über Lobbyismus. Allerdings wechselt die Perspektive etwas erratisch zwischen Anbietern (Wissenschaftliche Dienste des Bundestages, Stiftungen, Demoskopien) und Nachfragern von Politikberatung (Fraktionen, Ausschüsse, Parteien). Die Autoren sind zum Teil unmittelbar beteiligte Akteure der Politikberatung, das liefert interessante Einsichten, ist aber natürlich etwas anderes als kritische wissenschaftliche Forschung und Reflexion. Insbesondere die „Politikberatung von innen“, also durch den bürokratischen Apparat, durch Ressortforschung etc., hätte eine differenziertere Betrachtung verdient. Hier wird die These vertreten, „die Entscheidungsvorbereitung lebt vom Einfluss personaler Faktoren“ (Falk, Glaab, Römmele, Schober & Thunert, 2019, S. 135), also Personen seien wichtiger als Organisation und institutionelle Faktoren, und es wird dann von Machtmaklern, Selbstkasteiung, karitativen Funktionen, bürokratischer Fingerhakelei, Trickserie und ähnlichem erzählt. Das mag unterhaltend sein, trifft aber nicht die zentralen Mechanismen der Politikberatung in Regierung und Verwaltung.

Im dritten Teil geht es dann um „*Politikberatung in ausgewählten Politikfeldern*“. Hier gibt es anschauliche Übersichten über zentrale Institutionen der Politikberatung in der Wirtschafts-, der Arbeitsmarkt-, der Forschungs- und Technologiepolitik, oder auch dem Verbraucherschutz, zum Teil wiederum verfasst von direkt involvierten Akt-

euren. Der Beitrag zur Gesundheitspolitik liefert ein umfassendes historisches Narrativ über die unterschiedlichsten an den verschiedenen Reformen des Gesundheitswesens beteiligten, mehr oder weniger unabhängigen Politikberater. Aber ist das das zentrale Thema der Politikberatung in der Gesundheitspolitik? Natürlich konnte niemand die aktuelle Coronakrise vorhersehen, aber dass das Robert-Koch-Institut nur einmal *en passant* erwähnt wird, zeigt den *Bias* des Beitrages. Gerade anhand der Gesundheitspolitik hätte das komplizierte Verhältnis zwischen Wissenschaft und Politik erläutert werden können (und auch die damit verbundenen organisatorischen Probleme der Politikberatung), angefangen mit der HIV-Krise und der Auflösung des alten Bundesgesundheitsamtes, über die verschiedenen Epidemien der letzten Jahre (Rinderwahnsinn, Vogel- und Schweinegrippe, EHEC, SARS etc.). All dies wird nicht einmal erwähnt. Gerade die Gesundheitspolitik, aber auch die Klimapolitik (die leider fehlt) verdeutlichen die zentralen Kontroversen der wissenschaftlichen Politikberatung, gerade hier gibt es wieder einmal technokratische Vereinfachungen und Wunschvorstellungen. Hierzu hätte man gern mehr erfahren.

Der vierte Teil ‚Politikberatung international‘ behandelt zentrale Akteure in diesem Bereich (Stiftung Wissenschaft und Politik, Deutsche Gesellschaft für Auswärtige Politik, Deutsches Institut für Entwicklungspolitik, Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) u. a.), wiederum zum Teil von direkt Beteiligten vorgestellt. Zudem gibt es die exemplarische Darstellung der Erfahrungen in zwei Ländern, in der Schweiz und in Großbritannien. Und im fünften Teil, *„Methoden und Instrumente“* geht es dann vor allem um Akteure und Anbieter von Politikberatung, die erklären, warum und wie sie in der Politikberatung aktiv sind, und warum gerade dies besonders relevant sei. So schreiben die Inhaberin einer Agentur für *„Coaching und Moderation“* über Coaching, ein Senior Partner von *Berlin Risk Ltd.* über Risikoanalyse, eine Lobbyistin in Brüssel über Lobbying, Kommunikationsberater über *Public-Affairs*-Arbeit usw. Das ist durchaus instruktiv, aber keine kritische, distanzierte wissenschaftliche Analyse von Politikberatung, ihren Akteuren und Methoden, insbesondere deren Stärken und Schwächen.

Bei einem Handbuch ist es immer wohlfeil, sich zu beschweren, was aus Sicht des Rezensenten alles fehlt. Aber dennoch: Wünschenswert wäre ein Beitrag zur Ressortforschung (die nur im Beitrag über die Schweiz im Detail vorkommt). Das Robert-Koch-Institut mag erst in der Corona-Krise ein besonders bekannter Akteur geworden sein, aber auch schon vorher war diese Art der Politikberatung für Deutschland von entscheidender Bedeutung, und hierzu gibt es auch neuere Untersuchungen und kompetente Autoren. Die Rolle von Politikberatung in *Policy Networks* wäre einen eigenen Beitrag wert, und auch ein so zentrales, modisches Konzept wie *evidence based policy making* hätte eine eigene, kritische Diskussion verdient, also mehr als die kursorische Behandlung in dem Beitrag über Evaluation für und von Politikberatung.

Handbücher können bekanntlich sehr unterschiedliche Funktionen wahrnehmen. Sie können (1) breite Nachschlagewerke sein, (2) disziplinäre Überblicke „von Grundlagen zu ausgefächerten Vertiefungspfaden“ als Einführung in das Feld und (3) kritische Analysen des aktuellen Forschungsstandes zur Erschließung von Desideraten, also der Versuch, das Feld voranzubringen oder gar neu zu strukturieren. Wie kann man dieses Handbuch einordnen?

Ein Überblick über den Stand der Forschung wird vor allem im ersten Teil mit einigen sehr informativen Beiträgen geliefert, aber hier wäre Raum für weitere systema-

tische Einordnungen. Große Teile des Handbuchs sind tatsächlich in erster Linie ein informatives Nachschlagewerk, in dem verschiedene Akteure der Politikberatung dargestellt werden oder sich selbst darstellen. Die dritte Funktion, also die kritische Reflektion und die Identifikation von Forschungs- und Wissenslücken, wird leider zu wenig systematisch zu erfüllen versucht. Das ist schade, denn gerade die Politikberatung könnte einen solchen Anschub dringend vertragen. Es bleibt uns also nichts anderes übrig, als auf die dritte Auflage des Handbuchs zu warten, die hoffentlich bald erscheinen wird.

## Anmerkung

- 1 Aufgrund der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Selbstverständlich sind jeweils alle Geschlechter gleichermaßen gemeint.

## Literatur

Siefken, Sven T. (2010). Ist denn alles Politikberatung? *Politische Vierteljahresschrift*, 51(1), 127-136.

### *Anschrift des Autors:*

Prof. em. Dr. Werner Jann, Universität Potsdam, Seniorprofessur für Politikwissenschaft, Verwaltung und Organisation, August-Bebel-Str. 89, 14482 Potsdam, E-Mail: jann@uni-potsdam.de.

*Nadin Fromm*

McCowan, Martina (2019). *Public Administration in Times of Transnationalisation and Crisis. The Dynamics of Organisational Change in Ministries of Finance*. Baden-Baden: Nomos Verlagsgesellschaft. Paperback, 238 Seiten, ISBN: 9783848755592 (Print).

### Abstract

Recent policy and public administration studies have explored the emergence of organisational change of central government in turbulent times of reforms or crises to conclude repeatedly their manifest and persistent character. This book explores the influence of transnationalisation in fiscal policy and the European financial crisis on ministerial bureaucracies by comparing the German Federal Ministry of Finance and the Danish Ministry of Finance during the time period 1998 to 2014. The analysis is focusing on changes in organisational structure, civil servants' behavior and levels/mode of coordination. Two types of change are distinguished: change in terms of path dependencies and profound change. The study reveals that transnationalisation as reform leads to smaller adaptations and incremental change in the first phase of the study period (1998-2007). The financial crisis in the second phase (2008-2014), however, provokes profound changes in terms of structure, behaviour and coordination in both countries.

Die Fragen, ob sich die interne Organisationsstruktur öffentlicher Verwaltung angesichts einer Reform oder einer Krise wandelt und welche Folgen sich daraus für die internen Vollzugs- und Entscheidungsprozesse ergeben, dürfen als ein Kardinaldilemma der deutschen Verwaltungswissenschaft bezeichnet werden. Die (deutsche) Ministerialverwaltung ist hierbei weniger häufig Gegenstand der wissenschaftlichen Debatte (Veit, 2018, S. 121 f.), und es gibt bezüglich ihrer Wandlungs- und Innovationsfähigkeit unterschiedliche Befunde. Gemeinhin gilt, dass sich die Ministerialverwaltung aufgrund ihrer Autonomie im Gegensatz zu Verwaltungen auf Kommunal- und Landesebene eher als „resistent“ gegenüber Reformen charakterisiert (Bogumil & Jann, 2009, S. 259), sie „starke Beharrungskräfte“ beweist (ebenda) und ihre Mitglieder teilweise aktive Ablehnung zeigen (Hustedt & Veit, 2014). Demgegenüber belegen die Ergebnisse anderer Studien, dass sich ihre Organisationsstrukturen durchaus im Zeitverlauf verändern. Das betrifft insbesondere die Größe sowie die Ausdifferenzierung ihrer Aufbauorganisation, was als die zunehmende „Spezialisierung“ der Ministerialverwaltung angesichts eines externen Anpassungsdrucks einer immer komplexeren Umwelt bewertet wird (Bertels & Schulze-Gabrechten, 2019, S. 13, siehe dazu auch Hösl, Irgmaier & Kniep, 2019). Angesichts der scheinbar gegensätzlichen Erkenntnisse dreht sich die aktuelle wissenschaftliche Debatte insbesondere um die Frage, inwie-

fern Unterschiede im externen Anpassungsdruck erklärungskräftig für einen Wandel der Organisationsstruktur der Ministerialverwaltung und des Koordinationsverhaltens ihrer Mitglieder sind. Dies ist das Forschungsdesiderat, an dem die vorliegende Veröffentlichung mit dem Titel „Public Administration in Times of Transnationalisation and Crisis“ von Martina McCowan anschließt.

Die Publikation fasst die Ergebnisse des empirisch angelegten Forschungsprojektes mit dem Titel „Institutional Change within Governmental Organisations“ zusammen, welches am Munich Center on Governance, Communication, Public Policy and Law durchgeführt, als Dissertationsschrift im Jahr 2015 an der Ludwig-Maximilian-Universität München eingereicht wurde (McCowan, 2015) und nun als Monographie vorliegt. Die Veröffentlichung umfasst 238 Seiten und unterteilt sich in drei größere Teile (Einleitung, Haupt- und Schlussteil), jeweils durch zahlreiche Unterkapitel untergliedert. Im einführenden Textabschnitt werden der Forschungskontext, der heuristische sowie der theoretische Rahmen vorgestellt. Auf der Basis schließen sich die theoretisch-fundierte Herleitung der Vorannahmen und Hypothesen sowie die empirische Analyse an. Im Schlussteil werden die Ergebnisse zusammengefasst und Prämissen für die zukünftige Forschung umrissen. Die Untersuchung ist als komparative Fallstudie (auf Basis einer empirischen Untersuchung durch Interviews) angelegt und vergleicht die deutsche Ministerialverwaltung mit der Dänemarks im Zeitraum 1998-2014 (1998-2007 Reformphase der Transnationalisierung; 2008-2014 Finanzkrise) hinsichtlich der eingangs skizzierten Problemstellung.

Der übergeordnete Forschungskontext der Untersuchung ist die Transnationalisierung im Rahmen einer europäischen Finanzpolitik: Die Umsetzung der (gesamt-)europäischen Wirtschafts- und Währungsunion seit dem Jahr 1990 hat zur Folge, dass sich nationale Grenzen innerhalb der Europäischen Union hinsichtlich der Schaffung einer gemeinsamen Wirtschafts-, Fiskal- und Währungspolitik nach und nach auflösen. Damit besteht bei den (nationalen) Ministerialverwaltungen, insbesondere bei den Finanzministerien, die „funktionale Notwendigkeit“ (McCowan, 2019, p. 14) eines organisationalen Wandels sowie ein erhöhter Koordinationsbedarf. Um die Erklärungsfaktoren des externen Anpassungsdruckes eingehender studieren zu können, wird der Kontext um das interessante Momentum der Krise, hier die Europäische Finanzkrise im Jahr 2009/2010, ergänzt. Im Gegensatz zur Reform durch Transnationalisierung bedeutet die europäische Finanzkrise für die Finanzministerien, dass in kürzester Zeit Gegenmaßnahmen prompt, spontan und potentiell angemessen strategisch geplant und initialisiert werden (McCowan, 2019, p. 22). Dabei kann sich ein Wandel, gemäß der Autorin, auf unterschiedlichen Ebenen der Ministerialverwaltung (Organisationsstruktur, individuelles Entscheidungsverhalten des Personals) vollziehen, um mit dem erhöhten Koordinationsbedarf umzugehen. Im Rahmen der vorliegenden Veröffentlichung wird dies als „integrativer“ Erklärungsansatz klassifiziert und beschrieben (McCowan, 2019, pp. 55-57).

Als Folge fortschreitender europäischer Integration ist das deutsche Finanzministerium vor die Bewältigung einer Vielzahl neuer (transnationaler) Aufgaben gestellt und avanciert im Laufe der Zeit zur Durchsetzung nationaler Interessen zu einem zentralen Akteur der Koordination und der Umsetzung der europäischen Wirtschafts- und Währungsunion. Dies wird von den Interviewpartnern unisono bestätigt. Es zeigen sich trotz der zunehmenden Bedeutung und des Querschnittscharakters der Europäisierung keine Überschneidungen zwischen nationalen und europäischen Themengebieten we-



der auf Organisations- noch auf Handlungsebene, was laut Autorin eine Koordination je Einzelfall notwendig macht (McCowan, 2019, p. 99). Organisational führt die Europäisierung dazu, dass das hierarchisch-bürokratische Modell in seinen Grundsätzen weiterhin existiert, dabei durch u. a. neue Abteilungen und Referate ergänzt wird. Interessanterweise offenbart die hierarchische Struktur dabei Spielräume informellen Handelns (im Schatten der Hierarchie), welche sich in Form eines informellen Netzwerkes (entgegen der ursprünglichen Annahme) teilweise als flexibel im Umgang mit den ihnen gestellten Aufgaben erweist. Dies führt in der Reformphase, erstens, zu einer Entkopplung von Struktur und Koordinationsverhalten sowie, zweitens, zu einem Wandel, der im Rahmen einer (historisch bedingten) Pfadabhängigkeit beschrieben werden kann. In der europäischen Finanzkrise zeigt sich, dass angesichts eines hohen externen Drucks notwendig gewordene Sofortmaßnahmen und Sanktionen (u. a. für das in Finanznot geratene Griechenland) initialisiert werden. Das hat zur Folge, dass sich bestehende Abteilungen weiter ausdifferenzieren sowie neue Abteilungen eingerichtet werden. Im Gegensatz zur Reformphase bedeutet die Phase der Krise für das Bundesfinanzministerium, dass die Veränderungen auf organisationaler Ebene den Handlungen folgen (McCowan, 2019, p. 123). Die hierarchisch aufgebaute Organisationsstruktur erweist sich als probates Mittel, um (informelle) Netzwerk- und Kollektivstrukturen tatsächlich durchzusetzen. Der Wandel vollzieht sich hier nicht im Rahmen einer Pfadabhängigkeit, sondern zeigt Anzeichen tiefgreifender Veränderungen, insbesondere bei der „Identität“ der Ministerialbeamten: „Only through the catalysing factor of crisis do we observe a transnationalisation of the German ministerial bureaucracy, denoting the modification of the border between national and the transnational sphere, which takes the form of a cosmopolitanisation of identities” (McCowan, 2019, p. 147).

Im Gegensatz zum deutschen Ministerium ist das dänische Finanzministerium anfangs nicht als zentraler Akteur in die Umsetzung der europäischen Wirtschafts- und Währungsunion involviert.<sup>1</sup> Das ändert sich im Jahr 2001 mit einem Regierungswechsel. Nach dem Wahlsieg der konservativ-liberalen Partei „Venstre“ bringt sich die neu gewählte Regierung unter Ministerpräsident Anders Fogh Rasmussen (2001-2009) stärker gestaltend in europäische Wirtschafts- und Fiskalpolitik ein. Das hat Auswirkungen für das dänische Finanzministerium, welches nun einen Wandel „nachholt“, der sich in anderen EU-Mitgliedstaaten schon längst vollzogen hat: Zuständigkeiten werden transferiert (z. B. für das EU-Budget vom Außen- zum Finanzministerium) (McCowan, 2019, p. 150) und in Folge des Bedeutungsgewinns der Themen für das Finanzministerium etablieren sich neue Abteilungen und Referate (McCowan, 2019, p. 151). Ähnlich der deutschen Ministerialverwaltung gestaltet sich der Wandel auf organisationaler Ebene im Sinne einer Pfadabhängigkeit, hier ausgehend von einer Grundstruktur, die sich historisch am Weberianischen Bürokratiemodell anlehnt, welches im weiteren Verlauf erodiert und hin zu einer Netzwerkorganisation entwickelt, die sich durch Kollektiventscheidung kennzeichnet und nur dann Gebrauch von der Hierarchie macht, wenn die Notwendigkeit besteht. Die Flexibilität einer solchen Organisationsstruktur erweist sich auch auf Handlungsebene, also im Koordinationsverhalten zur Umsetzung der transnationalen Aufgaben, als Vorteil. Neue Aufgaben werden in bestehende Strukturen und Verhaltensweisen adaptiert. Anzeichen eines grundlegenden Wandels zeigen sich erst mit der Finanzkrise und äußern sich wie im deutschen Fall als ein Verschwinden der Grenzziehungen zwischen nationaler und europäischer Politik bei der Identität des Personals. Die Krise „befähigt“ die Ministerialbeamten, eine trans-

ationale Perspektive in Entscheidungsprozessen einzunehmen, was aus Sicht der Autorin zur Folge hat, dass Dänemark „becomes even more integrated in the aftermath of the crisis“ (McCowan, 2019, p. 185).

Aufgrund des umfangreichen wie auch komplexen Forschungsdesigns, welches zunächst von den graduellen Unterschieden beim externen Anpassungsdruck (Reform versus Krise) ausgeht und den potentiellen Wandel der Modi der Koordination im Hinblick auf organisationale Struktur und Entscheidungsverhalten des Personals der Ministerialverwaltung untersucht, fällt es dem Leser<sup>2</sup> nicht immer leicht, dem roten Faden der Argumentation zu folgen. Der Eindruck verstärkt sich, da alle (möglichen) Forschungsperspektiven (empirisch, theoretisch-konzeptionell, analytisch) gleichermaßen berücksichtigt werden und sich die Autorin beim Design der Studie nicht für ein heuristisches Erkenntnisinteresse entschieden hat, sich damit entsprechend methodologisch positioniert und somit riskiert, eine Ebene auszublenken. Es wird strukturell versucht, der Verwirrung des Lesers entgegen zu wirken und mit einer feingliedrigen, logischen Struktur, zahlreichen Zwischenüberschriften sowie grafischen Darstellungen des Forschungsdesigns zur besseren Systematik beizutragen (u. a. McCowan, 2019, pp. 31, 37). Hier muss anerkennend hinzugefügt werden, dass es der Autorin gelingt, sämtliche Perspektiven konsequent zu berücksichtigen. Mit einer kleinen Einschränkung: Leider wird wenig zum methodischen Vorgehen im Rahmen der empirischen Untersuchung gesagt.

Die Multiperspektive könnte sich im Rahmen einer wissenschaftlichen Anschlussfähigkeit auch als Vorteil herausstellen. Denn das Werk kann einerseits für Verwaltungswissenschaftler mit Fokus auf organisationalem Wandel interessant sein, andererseits könnte es in den Fokus der Forscher rücken, die sich mit inter- und intraministerieller Koordination in Reform- und Krisenzeiten beschäftigen. Aus theoretischer Perspektive könnte die Studie wiederum Verwaltungswissenschaftler ansprechen, welche sich konzeptionell mit organisationalem Wandel beschäftigen. Schließlich könnten die Erkenntnisse auch für die konzeptionelle Debatte bezüglich einer Transnationalisierung oder bezüglich eines „methodologischen Transnationalismus“ in der verwaltungs- und politikwissenschaftlichen Forschung (u. a. Stone & Lane, 2015) befruchtend sein. Für Studierende der ersten Studiensemester ist die Lektüre jedoch aufgrund der Komplexität weniger geeignet.

Angesichts eines sich ausweitenden Unilateralismus, der die Abschottung und Durchsetzung insbesondere nationaler Interessen zur Folge hat (z. B. in der Klimapolitik), scheinen die positiven Seiten einer Transnationalisierung generell plötzlich obsolet oder überholt. Die kritischen Stimmen gegenüber der Globalisierung, die sich insbesondere auf die negativen Folgen einer neoliberalen Wirtschafts- oder Arbeitsmarktpolitik beziehen, dominieren. Angesichts des derzeitigen Trends einer Re-Nationalisierung befinden wir uns aber nicht unbedingt in der Umkehr der Verhältnisse. Denn trotz der aktuellen Vorzeichen ist die Transnationalisierung oder die kosmopolitische Perspektive in zahlreichen Politikfeldern nicht mehr wegzudenken. Hier würde sich die erneute Lektüre von „Cosmopolitan political science“ (Grande, 2006) empfehlen, um die neu aufkeimenden nationalistischen Tendenzen besser zu verstehen und erneut zu hinterfragen. Dort heißt es: „Cosmopolitanism permanently provokes counter-attacks, and these counter-attacks can well – although they must not necessarily – take the form of the national. This ‘new’ type of the national, the neo-nationalist reaction and counter-movement within cosmopolitanism, constructs and stylizes the same boundaries and

demarcations, which already served as a defence mechanism against the very cosmopolitan transformation.” Dieses Argument verschiebt die Fragerichtung eines gegenwärtigen Analyseanspruchs mehr auf die Ursachen der Re-Nationalisierung und kann den Blick für zukünftige Entwicklungen klären. Vor diesem Hintergrund liest sich das Werk von Martina McCowan weniger als rückwärtsgewandte Auseinandersetzung mit einem sich längst vollzogenen Paradigmenwechsel, sondern mehr als Auftrag für zukünftige Forschung. Offenbar braucht die (Ministerial-)Verwaltung die Krise, damit eine Reform Erfolg hat.

## Anmerkungen

- 1 Im Jahr 2000 hatte man sich im Rahmen eines Referendums gegen die Einführung des Euro entschieden. Die Dänische Krone ist seit dem Jahr 1999 an die europäische Währung gekoppelt.
- 2 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Die weibliche Form ist selbstverständlich jeweils eingeschlossen.

## Literatur

- Bertels, Jana & Schulze-Gabrechten, Lena (2020). Mapping the black box of intraministerial organization: An analytical approach to explore structural diversity below the portfolio level, *Governance*, 1–19. <https://doi.org/10.1111/gove.12486> (online first).
- Bogumil, Jörg & Jann, Werner (2009). *Verwaltung und Verwaltungswissenschaft in Deutschland. Einführung in die Verwaltungswissenschaft*. Wiesbaden: VS Verlag.
- Hösl, Maximilian, Irgmaier, Florian & Kniep Ronja (2019). Diskurse der Digitalisierung und organisationaler Wandel in Ministerien. In Tanja Klenk, Frank Nullmeier & Göttrik Wewer (Hrsg.), *Handbuch Digitalisierung in Staat und Verwaltung*. Wiesbaden: VS Verlag. [https://doi.org/10.1007/978-3-658-23669-4\\_33-1](https://doi.org/10.1007/978-3-658-23669-4_33-1)(online first).
- Hustedt, Thuid & Veit, Sylvia (2014). Forschungsperspektiven auf Regierungs- und Verwaltungskoordination: Koordinationsprobleme und Erklärungsfaktoren. *dms – der moderne staat*, 7, 17-36.
- McCowan, Martina (2015). *Transnationalization, Crisis and Organizational Change in Ministerial Bureaucracies. A Comparative Analysis of the German and Danish Ministries of Finance*. Univ. Dissertation, Ludwig-Maximilians-Universität, München.
- Stone, Diane & Ladi, Stella (2015). Global Public Policy and Transnational Administration. *Public Administration*, 93, 4, 839-855.
- Veit, Sylvia (2018). Verwaltungsreformen in der Bundesverwaltung: Eine kritische Bestandsaufnahme. In Edgar Grande & Michael W. Bauer (Hrsg.), *Perspektiven der Verwaltungswissenschaft* (S. 121-153). Baden-Baden: Nomos-Verlagsgesellschaft. <https://doi.org/10.5771/9783845288567-121>.

## Anschrift der Autorin:

Dr. Nadin Fromm, Universität Kassel, Fachgebiet Public Management, Kleine Rosenstraße 3, 34109 Kassel, E-Mail: [nfromm@uni-kassel.de](mailto:nfromm@uni-kassel.de).

Julia Schwanholz

Thiele, Alexander (Hrsg.) (2019). *Legitimität in unsicheren Zeiten. Der demokratische Verfassungsstaat in der Krise?* Mohr Siebeck, 225 Seiten. ISBN: 9783161557675.

### Abstract

Alexander Thiele's anthology on *Legitimacy in uncertain times* deals with the question of whether we are in a "general and serious crisis of legitimacy in the (western) democratic state" (Thiele, 2019, S. 3). Answers are provided on the theoretical foundations of legitimacy – from a historical, legal and economic perspective –, and on the foundation of empirical findings and solutions. The book offers readers interdisciplinary insights into the research of legal, economic and social scientists with the aim of sensitizing for the strengths and weaknesses of democracy.

„Wahlen allein“, so sagte es einmal der ehemalige US-Präsident Barack Obama, „machen noch keine Demokratie“. Eine Legitimation durch Wahlen ist demnach notwendig, aber nicht hinreichend für die Demokratie. Ansonsten wären zahlreiche Autokratien auch demokratisch. Und wenn von einer Krise der Demokratie die Rede ist, so geht es demnach um mehr als nur die sinkende Wahlbeteiligung: Soziale Ungleichheit, Vertrauensverluste in Rechtsstaatlichkeit, Rechtspopulismus u. a. m. Die Erzählung von der Krise der Demokratie ist in diesem Kontext so alt wie die Regierungsform selbst. In der Bundesrepublik, das belegen politikwissenschaftliche Studien, ist die Demokratie gewiss nicht perfekt, aber seit Jahrzehnten überaus robust und stabil.

Diese Einschätzung wird im Sammelband *„Legitimität in unsicheren Zeiten“* von Alexander Thiele hinterfragt. In der Einführung wirft er die Frage auf, ob wir uns in einer „allgemeinen und ernsthaften Legitimitätskrise des (westlichen) demokratischen Verfassungsstaates“ befinden (Thiele, 2019, S. 3). Legitimität ist für Thiele in Abgrenzung zu Legitimation (= demokratische Rückbindung) und Legalität (= rechtliche Bindung) die soziale Anerkennung einer bestehenden Herrschaftsform (Thiele, 2019, S. 4). Der Herausgeber blickt damit nicht singular auf Wahlen, sondern in einem weiteren Radius auf die Unterstützung und das Vertrauen der Bürger<sup>1</sup> in die für sie geltende Herrschaftsform.

Angetrieben von einigen negativen Zustandsbeschreibungen weltweiter Demokratien – es finden sich dazu Beispiele für etablierte Demokratien wie die USA, Italien, die Niederlande und Österreich, auch für postsozialistische wie Polen und Ungarn oder für fragile Demokratien wie Tunesien, Afghanistan, Libyen, Ukraine und Syrien oder

sogenannte demokratische Hoffnungsträger wie die Türkei und Russland sowie noch einige andere wie Indien, Südafrika und Brasilien – und geleitet von der Beobachtung, dass Rechtspopulismus mehr und mehr um sich greift, bringt Thiele seine Sorge um unsere Demokratie in der Einführung zum Ausdruck. Das erklärte Ziel des Bandes ist es dann auch, Wachsamkeit zu erzeugen, Krisensymptome zu identifizieren und legitimistische Stellschrauben neu zu justieren (Thiele, 2019, S. 18).

Zwar bestimmt der Herausgeber nur ungefähr, was er unter Demokratie versteht und was eigentlich eine Krise markiert bzw. wann diese als „*allgemein*“ und/oder „*ernsthaft*“, wie er schreibt, gelten kann. Er bietet seiner Leserschaft aber eine Positiv-Definition zur Ausgestaltung „denationalisierter demokratischer Verfassungsstaaten“ an. So kann er zeigen, worin Legitimitätschancen wurzeln (Thiele, 2019, S. 6), nämlich in einer Trias aus demokratischer Teilhabe an und der Begrenzung von Staatsgewalt sowie einer angemessenen Leistungsfähigkeit des Staates.

An dieser Trias arbeiten sich die interdisziplinären Kapitel des Buches in drei Teilen ab. Nach einem ersten Abschnitt (Teil 1) zu den theoretischen Grundlagen von Legitimität – aus historischer, rechtswissenschaftlicher und ökonomischer Perspektive – folgen zwei Abschnitte mit je drei Kapiteln zu empirischen Befunden (Teil 2) und Möglichkeiten der Abhilfe (demokratischer Schwächen) (Teil 3). Die Leserschaft erhält dazu Einsichten in die Forschungsarbeiten von Rechts-, Wirtschafts- und Sozialwissenschaftlern (letztere vor allem mit politikwissenschaftlicher und soziologischer Handschrift).

Die theoretischen Kapitel von Walter Reese-Schäfer und Friederike Wapler klären einesteils die ideengeschichtlichen Wurzeln von Legitimität und schärfen andernteils das Konzeptverständnis der Legitimitätskrise. Das dritte Theoriekapitel zur Rolle ökonomischer Wissenschaft in der Demokratie fällt etwas aus dem Rahmen der sonstigen Beiträge. Dabei unterstreicht Achim Trugers Plädoyer für eine demokratische Streitkultur die grundsätzliche Idee des Buches, eine lebendige und dauerhafte Debatte über Demokratie anzustoßen.

Woran dies in der Realität scheitern kann oder es bereits teilweise tut, zeigen die empirischen Befunde der Kapitel im zweiten Teil zu politischer Partizipation (in Wahlen oder Parteien) und sozialer Ungleichheit auf. Michael Kaeding untersucht politische Gleichheit und findet folgenden Zusammenhang: Je schlechter der sozioökonomische Status einer Person desto geringer sein Partizipationsverhalten und desto geringer schließlich auch die Legitimationskraft von Wahlen. Da Legitimität über Legitimation entsteht, nimmt am Ende auch diese ab. Die damit verbundene Frage nach der Bedeutung der Mittelschicht greift Cornelia Koppetsch im Kapitel über soziale Ungleichheit auf. Einen Glauben an die heilenden Kräfte des Marktes und Wohlstand für alle gibt es spätestens seit der vollständigen Internationalisierung des globalen Wettbewerbs nicht mehr. Die einfache Formel lautet: Reiche werden immer reicher und Arme immer ärmer. Von einer schmelzenden Mittelschicht profitieren laut Koppetsch am Ende vor allem (Rechts-)Populisten mit ihren unlauteren Versprechen. Auch Julian Krüper kommt in seinem Kapitel über politische Parteien zu keiner optimistischen Einschätzung und beschreibt deren Scheitern an sich selbst als eine logische Konsequenz organisationaler, funktionaler und systemischer Überforderungen.

Dass Legitimität trotz schwieriger Befunde in unsicheren Zeiten möglich ist, wird im dritten Teil aufgezeigt. Die Beiträge nehmen alternative Formen politischer Partizipation (Bürgerbeteiligung) in den Blick und schlagen Reformansätze für bestehende

(soziale) Systeme und Verhältnisse vor. Daniela Schweigler diskutiert Wege einer Neuordnung der sozialen Sicherung auf Basis der Beobachtung sich verändernder Lebensverhältnisse (bspw. Erwerbsverläufe). Mehr Akzeptanz sei am Ende nur zu erreichen, wenn die Regeln in der Wirtschafts- und Arbeitswelt auch den Bedürfnissen der Menschen angepasst würden.

Eine Ergänzung repräsentativ-demokratischer Elemente durch Bürgerbeteiligung wird seit einigen Jahren als Allheilmittel von Partizipationsbefürwortern erachtet. Daniel Oppold und Patrizia Nanz erklären dies in ihrem Kapitel und zeigen auf, welchen Stellenwert im Vergleich zu direkter Demokratie dialogorientierte Beteiligungsverfahren haben. Aus Sicht der Autoren ist es dringend wünschenswert, die repräsentative Demokratie um dialogorientierte Partizipationsangebote zu ergänzen, weil darüber Verfahren qualitativ zu verbessern und Legitimität zu erhöhen sind. Ob dies in der Realität umzusetzen ist, hinterfragen Simon Fink und Eva Ruffing im letzten Beitrag des Bandes am Beispiel von Stromnetzen. Sie überprüfen, ob sich eine additive Legitimationswirkung entfaltet, wenn Beteiligungsverfahren und Entscheidungsarenen miteinander gekoppelt werden und also sowohl die Öffentlichkeit beteiligt wird als auch herkömmliche repräsentativ-demokratische Prozesse stattfinden. Die Antwort der Autoren fällt differenziert aus, weil politische Fragen und Mitbestimmungsszenarien stets hochkomplex sind und aufgrund dessen unter Umständen das Gegenteil von mehr Legitimität, nämlich eine Überforderung und Enttäuschung, bei den Bürgern auslösen.

Welche Erkenntnisse stellen sich nach der Lektüre des Bandes ein? Die Leserschaft kann einiges über das Konzept der Legitimität, deren Herausforderung für die Demokratie und Auswege (so etwa alternative Formen der Partizipation) lernen. Mit der Kapitelsammlung hat Alexander Thiele – und darin liegt sein großer Verdienst – wertvolle Einsichten und interessante Befunde zu der überaus aktuellen und wichtigen Frage nach Legitimität in unsicheren Zeiten zusammengetragen.

Was dem Sammelband fehlt, ist eine abgewogene Krisendiagnostik: Denn die einschlägige Literatur stellt sich gegen negative Pauschalurteile. Dort werden eher Vertrauensverluste in einzelne politische Institutionen oder gegenüber der politischen Elite anstatt einer ausgewachsenen Demokratiekrise konstatiert (Merkel, 2015; Pickel, 2013; van Ham, Thomassen, Aarts & Andeweg, 2017). Die einzelnen Kapitel des Bandes sprechen – trotz thematischer Verwandtschaft – nicht zueinander (es gibt keine Querverweise) und auch der Wert und Nutzen von Interdisziplinarität wird nirgendwo ausgeführt. Ein geeigneter Ort hierfür wäre ein Fazit gewesen, um auch die Frage des Bandes nach der demokratischen Legitimitätskrise in Hinblick auf die Trias seiner Einführung hinlänglich zu beantworten. Die Frage, wie es denn nun um unsere Demokratie aktuell steht, bleibt am Ende offen. Dies lädt wiederum dazu ein, über die Stärken und Schwächen von Demokratien ins Gespräch zu kommen; und genau das war ja auch die zentrale Intention des Herausgebers.

## Anmerkung

- 1 Aus Gründen der besseren Lesbarkeit wird im Folgenden ausschließlich die männliche Sprachform verwendet. Gleichwohl sind alle Geschlechter gleichermaßen gemeint.

## Literatur

- Merkel, Wolfgang (2015). Schluss. Ist die Krise der Demokratie eine Erfindung?. In Wolfgang Merkel (Hrsg.), *Demokratie und Krise* (S. 473-498), Wiesbaden.
- Pickel, Susanne (2013). Politische Kultur, Systemvertrauen und Demokratiezufriedenheit. Wann fühlen sich die Bürger gut regiert?. In Karl-Rudolf Korte & Timo Grunden (Hrsg.), *Handbuch Regierungsforschung* (S. 161-174), Wiesbaden.
- van Ham, Carolien, Thomassen, Jacques, Aarts, Kees & Andeweg, Rudy (2017). *Myth and Reality of the Legitimacy Crisis: Explaining Trends and Cross-National Differences in Established Democracies*. Oxford.

### *Anschrift der Autorin:*

Dr. Julia Schwanholz, Professur für Ethik in Politikmanagement und Gesellschaft (i. V.), NRW School of Governance, Institut für Politikwissenschaft, Universität Duisburg-Essen, Lotharstraße 53, 47057 Duisburg, E-Mail: [julia.schwanholz@uni-due.de](mailto:julia.schwanholz@uni-due.de).



# Schriftenreihe Politik und Recht



## Verfassungsgerichtliche Entscheidungsfindung und ihre Folgen

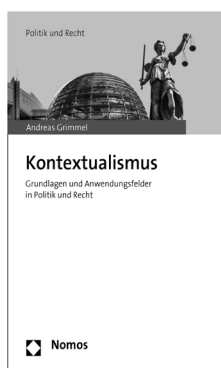
Das Türkische Verfassungsgericht zwischen Mehrheitslogik und Konsensverfahren

Von Dr. Maria Abad Andrade

2020, 314 S., brosch., 69,- €

ISBN 978-3-8487-6378-8

In diesem Buch wird erstmalig die Entscheidungsfindung am Türkischen Verfassungsgericht untersucht. Die Autorin entwickelt denkbare Modelle höchstrichterlicher Entscheidungslogiken und wendet diese in einer detailreichen Fallstudie auf das Türkische Verfassungsgericht an.



## Kontextualismus

Grundlagen und Anwendungsfelder in Politik und Recht

Von PD Dr. Andreas Grimmel

2019, 314 S., brosch., 59,- €

ISBN 978-3-8487-5815-9

Ändert sich der Kontext, so ändert sich auch der Horizont möglicher Wahrnehmungen, Deutungen und Handlungsmöglichkeiten. Dieser Band entwickelt einen neuen kontextualistischen Ansatz für die sozialwissenschaftliche Forschung, der in einer Reihe von Fallstudien auf aktuelle Entwicklungslinien der Europäischen Union (EU) und der Weltgesellschaft angewandt wird.

 **Nomos**  
eLibrary [www.nomos-elibrary.de](http://www.nomos-elibrary.de)

Bestellen Sie im Buchhandel oder versandkostenfrei online unter [nomos-shop.de](http://nomos-shop.de)

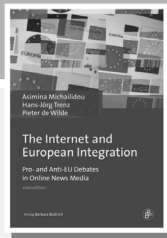
Bestell-Hotline (+49)7221.2104-37 | E-Mail [bestellung@nomos.de](mailto:bestellung@nomos.de) | Fax (+49)7221.2104-43

Alle Preise inkl. Mehrwertsteuer



**Nomos**





Asimina Michailidou, Hans-Jörg Trezn, Pieter de Wilde

## The Internet and European Integration

Pro- and Anti-EU Debates in Online News Media

This book offers a wealth of original empirical data on how online media shape EU politics. The authors highlight myths and facts about the nature of audience-driven online media content and show how public demands for legitimacy are at the heart of the politicization of European integration. Drawing on innovative and rigorous analysis of online media ownership, journalistic content, and readers input, the authors piece together the convergence toward Euroskepticism across EU member states.

2020 • approx. 250 pp. • Pb.  
approx. 33,00 € (D) • approx. 34,00 € (A)  
978-3-8474-2375-1 • available as eBook



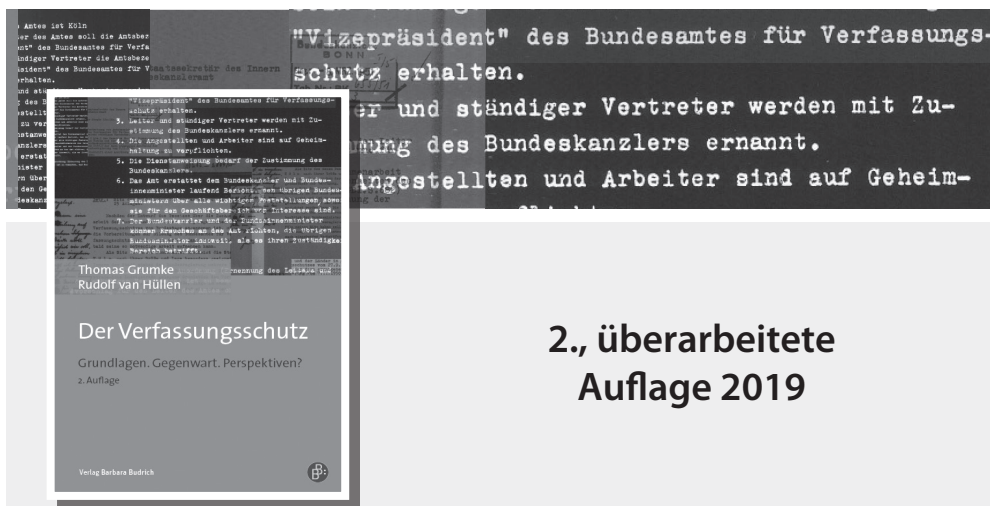
Marianne Kneuer, Helen V. Milner (eds.)

## Political Science and Digitalization – Global Perspectives

Digitalization is not only a new research subject for political science, but a transformative force for the discipline in terms of teaching and learning as well as research methods and publishing. This volume provides the first account of the influence of digitalization on the discipline of political science including contributions from 20 different countries. It presents a regional stocktaking of the challenges and opportunities of digitalization in most world regions.

2019 • 290 pp. • Hc.  
60,00 € (D), 61,70 € (A)  
ISBN 978-3-8474-2332-4 • eISBN 978-3-8474-1488-9

[www.shop.budrich.de](http://www.shop.budrich.de)



## 2., überarbeitete Auflage 2019

Thomas Grumke • Rudolf van Hüllen

# Der Verfassungsschutz

Grundlagen. Gegenwart. Perspektiven?

Wie kann ein zukunftsfähiger Verfassungsschutz aussehen? Welche zum Teil längerfristigen Fehlentwicklungen müssen behoben werden? Kritik an den Verfassungsschutzbehörden gehört mindestens seit der Entdeckung des NSU-Rechtsterrorismus zum guten Ton öffentlicher Empörung. Weniger bekannt ist aber, wie diese Institutionen aufgebaut sind, wie sie in der Realität funktionieren und was sie tatsächlich leisten können. Die Autoren schließen diese Wissenslücke, formulieren Verbesserungsbedarf und entwickeln Perspektiven zwischen den nicht selten zu hörenden Forderungen "Abschaffen" und "Weiter so".

Die Autoren:

Prof. Dr. Thomas Grumke | Professor für Politik und Soziologie an der Fachhochschule für öffentliche Verwaltung NRW, Studienort Gelsenkirchen

Dr. Rudolf van Hüllen | Freiberuflicher Dozent und Extremismusforscher, Krefeld

2019 • 250 S. • Kart. • 24,90 € (D) • 25,60 € (A)

ISBN 978-3-8474-2280-8 • eISBN 978-3-8474-0694-5

[www.shop.budrich.de](http://www.shop.budrich.de)

- **Allgemeines:**

„dms – der moderne staat“ veröffentlicht im Aufsatzteil grundsätzlich nur Originalarbeiten. Mit der Einsendung des Manuskripts erklärt die Autorin/der Autor, dass ihr/sein Beitrag nicht (auch nicht in Teilen) bereits an einem anderen Ort erschienen oder zur Veröffentlichung vorgesehen ist.

„dms – der moderne staat“ wendet das Verfahren der anonymen Begutachtung an. Jedes Manuskript wird wenigstens zwei Gutachter(inne)n vorgelegt. Name und Anschrift/Telefonnummer/E-Mail-Adresse der Autorin/des Autors sollten also nur auf einem gesonderten Deckblatt bzw. im Anschreiben stehen. Auffällige Selbstzitate und Anmerkungen sind zum Zeitpunkt der Begutachtung zu vermeiden, um die Anonymität der Autorin/des Autors zu wahren. Im Falle der Nichteinhaltung dieser Empfehlungen kann die Redaktion die Anonymität nicht verbürgen.

- **Einsendung von Manuskripten:**

Manuskripte sind der Redaktion (dms@hertie-school.de) nur in elektronischer Form (per E-Mail oder Datenträger) zuzusenden. Die Zusendung in Print-Form ist nicht erforderlich. Manuskripte können in allen gängigen Textverarbeitungsprogrammen übermittelt werden. Manuskripte sollen den Umfang von 8000 Wörtern nicht überschreiten.

Die Zeitschrift enthält verschiedene Rubriken, von allgemeinen Abhandlungen über spezielle Aufsätze zur Forschungsagenda, Literaturberichte und auch Kontroversen. Der Umfang kann zwischen 4000 und 8000 Wörtern variieren.

Nähere Hinweise zur Anfertigung der Manuskripte unter <http://www.budrich-journals.de/index.php/dms/about/submissions#authorGuidelines>

# Risiken besser im Blick



## Management von Risiken in Behörden

Konzepte – Instrumente – Praxislösungen

Von Prof. Dr. Bernhard Hirsch,  
Prof. Dr. Dr. h. c. Jürgen Weber,  
Fabienne-Sophie Schäfer und Josef Schmid

2020, ca. 220 Seiten, ca. € (D) 39,95,  
ISBN 978-3-503-19150-5

Steuerung & Controlling in öffentlichen Institutionen

Auch als eBook erhältlich

Das Risikomanagement und die daraus gewonnenen Informationen sind auch für Behörden von essentieller Bedeutung. Doch nur wenige öffentliche Institutionen betreiben ein behörden spezifisches Risikomanagementsystem. Durch weitgehend fehlende gesetzliche Vorgaben ist es ihnen selbst überlassen, ob und wie sie mit Risiken in der Verwaltung umgehen.

## Eine einzigartige Innensicht

Worauf es ankommt, beleuchtet der neueste Band des Arbeitskreises „Steuerung und Controlling in öffentlichen Institutionen“. Im Fokus u. a.:

- ▶ **Aktueller Stand des Risikomanagements** in öffentlichen Institutionen
- ▶ **Möglichkeiten und Limitationen** bei der Etablierung einer geeigneten **Risikokultur**
- ▶ Ansätze und Grenzen der **systematischen Risikosteuerung** im Verwaltungskontext
- ▶ **Praktische Anforderungen**, insbesondere auch an ein **risikobasiertes Berichtswesen**

Der Arbeitskreis, dem neben renommierten Wissenschaftlern hochrangige Vertreter aus 20 Bundes- und Landesbehörden angehören, ist ein **Teilprojekt des Regierungsprogramms „Digitale Verwaltung 2020“** der Bundesregierung.

Online informieren:

 [www.ESV.info/19150](http://www.ESV.info/19150)

**ESV** ERICH  
SCHMIDT  
VERLAG

Auf Wissen vertrauen

Bestellungen bitte an den Buchhandel oder: Erich Schmidt Verlag GmbH & Co. KG · Genthiner Str. 30 G · 10785 Berlin  
Tel. (030) 25 00 85-265 · Fax (030) 25 00 85-275 · [ESV@ESVmedien.de](mailto:ESV@ESVmedien.de) · [www.ESV.info](http://www.ESV.info)

