



Frank Greuel | Franziska Heinze |
Frank König (Hrsg.)

Was wirkt wie und warum?

**Wirkungsevaluationen in
pädagogischen Handlungs-
feldern für Demokratie
und gegen Extremismus**

BELTZ JUVENTA

Frank Greuel | Franziska Heinze | Frank König (Hrsg.)
Was wirkt wie und warum? Wirkungsevaluationen in pädagogischen
Handlungsfeldern für Demokratie und gegen Extremismus

Eine Veröffentlichung des Deutschen Jugendinstituts e.V. (DJI)

Das DJI ist eines der größten sozialwissenschaftlichen Forschungsinstitute Europas. Seit 60 Jahren erforscht es die Lebenslagen von Kindern, Jugendlichen und Familien, berät Bund, Länder und Gemeinden und liefert wichtige Impulse für die Fachpraxis. Aktuell sind an den beiden Standorten München und Halle (Saale) etwa 460 Beschäftigte tätig, darunter rund 300 Wissenschaftlerinnen und Wissenschaftler. Finanziert wird das DJI überwiegend aus Mitteln des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) und der Bundesländer. Weitere Zuwendungen erhält es im Rahmen von Projektförderungen u. a. vom Bundesministerium für Bildung und Forschung (BMBF), der Europäischen Kommission, Stiftungen und anderen Institutionen der Wissenschaftsförderung.

Frank Greuel | Franziska Heinze | Frank König
(Hrsg.)

**Was wirkt wie und warum?
Wirkungsevaluationen
in pädagogischen Handlungs-
feldern für Demokratie
und gegen Extremismus**

BELTZ JUVENTA

Die Veröffentlichung stellt keine Meinungsäußerung des BMFSFJ oder des BAFzA dar. Für inhaltliche Aussagen tragen die Autorinnen und Autoren die Verantwortung.

Gefördert vom

im Rahmen des Bundesprogramms



Bundesministerium
für Familie, Senioren, Frauen
und Jugend

Demokratie *leben!*

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Der Text dieser Publikation wird unter der Lizenz **Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-NC-SA 4.0)** veröffentlicht. Den vollständigen Lizenztext finden Sie unter:

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.de>

Verwertung, die den Rahmen der **CC BY-NC-SA 4.0 Lizenz** überschreitet, ist ohne Zustimmung des Verlags unzulässig. Die in diesem Werk enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Quellenangabe/Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Dieses Buch ist erhältlich als:

ISBN 978-3-7799-8070-4 Print

ISBN 978-3-7799-8071-1 E-Book (PDF)

DOI 10.3262/978-3-7799-8071-1

1. Auflage 2025

© 2025 Beltz Juventa

in der Verlagsgruppe Beltz · Weinheim Basel

Werderstraße 10, 69469 Weinheim

service@beltz.de

Einige Rechte vorbehalten

Satz: xerif, le-tex

Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza

Beltz Grafische Betriebe ist ein Unternehmen mit finanziellem Klimabeitrag

(ID 15985-2104-100)

Printed in Germany

Weitere Informationen zu unseren Autor:innen und Titeln finden Sie unter: www.beltz.de

Inhalt

Statt eines Vorworts: Danksagung	7
I Einführung	
1 Wirkungsevaluation in der Arbeit für Demokratie und gegen Extremismus <i>Frank Greuel, Franziska Heinze und Frank König</i>	10
II Wirkungen und Gegenstandsangemessenheit	
2 Wie ernst ist es uns wirklich mit der Wirkung? <i>Jan Ulrich Hense</i>	28
3 Gegenstandsangemessenheit in Wirkungsevaluationen <i>Frank Greuel</i>	55
III Evaluation von Wirkungen im Bundesprogramm „Demokratie leben!“	
4 Wirkungsorientierte Evaluation von Beratung nach der Realist Evaluation <i>Carmen Figlestahler, Katrin Haase und Katja Schau</i>	70
5 Wirkungsevaluation im Bereich Demokratieförderung <i>Aline Rehse, Max Hemmann und Marina Braun</i>	89
6 „Das hätte ich, dann jetzt rückblickend betrachtet, noch anders gestaltet“ <i>Jonas Euteneuer und Stella Rürger</i>	108
7 Resonanz ist mehr als die Sehnsucht nach Wiederhall <i>Franziska Heinze, Sarah Langer, Steffen Loick Molina, Stefanie Reiter, Kornelia Sammet und Ellen Schroeter</i>	127
8 Wie bringt ein Bundesprogramm die Zivilgesellschaft „in Aktion“? <i>Frank König, Alexander Stärck und Diana Zierold</i>	145

9	Die Analyse von Bedarfen in den Bereichen Demokratieförderung und Extremismusprävention <i>Laura Meijer und Pia Sauermann</i>	164
---	--	-----

IV Statt eines Fazits

10	Anforderungen an wirkungsorientierte Evaluationen im Bereich der Demokratieförderung, politischen Bildung und Prävention von Extremismus <i>Es diskutierten Helle Becker, Jörg Faust, Hemma Mayrhofer, Björn Milbradt und Thomas Widmer</i>	182
	Verzeichnis der Autorinnen und Autoren	207

Statt eines Vorworts: Danksagung

An dieser Stelle möchten wir uns bei allen Personen bedanken, die durch ihr *Wirken* zum Entstehen und Gelingen der vorliegenden Publikation beigetragen haben. Zunächst danken wir allen Autor:innen aus den Projekten der Programmevaluation „Demokratie leben!“ und der „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“ dafür, dass sie ihre Erfahrungen mit wirkungsorientierter Evaluation aufbereitet und reflektiert haben und ihr Wissen mit einem breiteren Fachpublikum teilen. Wir danken Christian Lüders, der die Bundesprogramme zur Demokratieförderung und Extremismusprävention seit über zwei Jahrzehnten engagiert begleitet und als ehemaliger Leiter der Abteilung Jugend und Jugendhilfe am Deutschen Jugendinstitut auch die Programmevaluation „Demokratie leben!“ und ihre evaluationstheoretische Grundlegung viele Jahre geprägt hat.

Ein besonderer Dank geht an Jan Ulrich Hense und die Mitwirkenden an einem Fachgespräch zum Thema Wirkungsevaluation, Helle Becker, Jörg Faust, Hemma Mayrhofer, Björn Milbradt und Thomas Widmer, die mit ihren Perspektiven die Praxisbeiträge dieses Sammelbandes rahmen und einordnen helfen.

Ein solches Projekt wäre nicht umsetzbar ohne die oftmals unsichtbare Unterstützung einer Reihe von Personen im Hintergrund. Wir danken dem Verlag Beltz Juventa für das Interesse, diesen Sammelband mit uns zu realisieren, und Susanne Opitz für die Unterstützung bei der Akquise und der Kommunikation mit dem Verlag. Wir bedanken uns sehr herzlich bei Susanne Timling, die uns organisatorisch bei der Umsetzung des Fachgesprächs unterstützt hat, und bei Petra Göricke, die durch gründliches Korrekturlesen der Beiträge und die Manuskriptgestaltung maßgeblich zur Entstehung dieses Bandes beigetragen hat.

Allen Leserinnen und Lesern danken wir für das Interesse an dieser Publikation und wünschen eine anregende Lektüre!

Halle (Saale) im Sommer 2024,

Frank Greuel, Franziska Heinze und Frank König

I Einführung

1 Wirkungsevaluation in der Arbeit für Demokratie und gegen Extremismus

Einleitung und Überblick über den Band

Frank Greuel, Franziska Heinze und Frank König

Seit über zwei Jahrzehnten unterstützen sowohl der Bund als auch die Länder vor allem zivilgesellschaftliche Akteur:innen mittels staatlicher Förderprogramme dabei, demokratische Beteiligung zu steigern, die Akzeptanz kultureller und geschlechtlicher Vielfalt zu erhöhen sowie Extremismus vorzubeugen. Sowohl die Programmverantwortlichen als auch die interessierte politische und Fach-Öffentlichkeit stellen an die Evaluation dieser Programme in wachsendem Maße die Anforderung, über die Wirksamkeit und die Wirkungen der geförderten Maßnahmen auskunftsfähig zu sein. Dies steht im Einklang mit dem generell zu konstatierenden gestiegenen Interesse an Wirkungsuntersuchungen im Feld der Kinder- und Jugendhilfe (Begemann / Bleck / Liebig 2019, S. 7), die im Zusammenspiel mit der Kinder- und Jugendpolitik des Bundes einen wichtigen Bezugshorizont für entsprechende Programme darstellt.

Aus der Perspektive von Politik und Verwaltung sollen Wirkungsevaluationen in solchen Programmen folgende Fragen beantworten: Halten die Programme, was sie versprechen? Erzielen sie die intendierten Wirkungen? Kommt es zu unerwünschten Nebenwirkungen? Rentiert sich der Einsatz von öffentlichen Geldern? Grundlage sind hierfür § 7 der Bundeshaushaltsordnung sowie die entsprechende Verwaltungsvorschrift, die vorgeben, dass für jede finanzwirksame Maßnahme im Rahmen ihrer Wirkungskontrolle zu untersuchen ist, ob diese „für die Zielerreichung geeignet und ursächlich war. Hierbei sind alle beabsichtigten und unbeabsichtigten Auswirkungen der durchgeführten Maßnahme zu ermitteln“ (https://olev.de/w/vv-bho_7.pdf, S. 4). Die Relevanz dieser Fragen hat in letzter Zeit deutlich zugenommen (Clark/Dollinger/Heppchen 2024; Ottmann/König 2023b; Albus/Micheel/Polutta 2018). Auch Christian Lüders (2024) konstatiert diesbezüglich, dass sich im politischen und im politisch-administrativen Raum erkennbar Akzente in Richtung Wirkungsanalyse verschoben haben. Als Beispiel zieht er die im Frühsommer 2024 erfolgten Ausschreibungen für die Evaluationen der 3. Förderphase des Bundesprogramms „Demokratie leben!“ ab 2025 heran. Diese Akzentverschiebungen liegen zum einen daran, dass in den zurückliegenden Jahren staatliche Programme und Initiativen der Demokratiestärkung, Vielfaltgestaltung und Extremismusprävention einen deutlichen

Mittelaufwuchs erfahren haben und die Verwendung öffentlicher Mittel einer Reihe von Grundsätzen folgt (u. a. Sparsamkeit und Wirtschaftlichkeit). Zum anderen sind diese Programme verstärkt Gegenstand öffentlicher und politischer Auseinandersetzungen und damit in besonderer Weise begründungsbedürftig geworden, wofür wirkungsbezogene Evaluationen die nötigen Argumente beitragen sollen.

Darüber hinaus dienen Erkenntnisse zur Wirksamkeit der geförderten Aktivitäten und zu den Bedingungen, unter denen sie Wirkungen entfalten können, auch dazu, die evaluierten Programme – im Sinne evidenzbasierter Politik – an sich stetig ändernde Gegebenheiten anzupassen und weiterzuentwickeln.

Diese Lern- und Anpassungsfunktion wirkungsorientierter Evaluation kommt auch den geförderten Akteur:innen selbst zugute. Hier trägt Evaluation mit ihren Wirkungsuntersuchungen vor allem zur Qualitätssicherung der zivilgesellschaftlichen Arbeit und zur Unterstützung von Professionalisierungsprozessen in der Fachpraxis bei. Dabei kann aus der Kontroll- und Legitimationsfunktion von Wirkungsevaluation auf der einen und deren Lern- und Anpassungsfunktion auf der anderen Seite ein Spannungsfeld insbesondere dann entstehen, wenn Auftraggeber:innen oder andere Stakeholder in diese Evaluationen überzogene, gegenstandsunangemessene Wirkungserwartungen hineinprojizieren.

Davon unabhängig ist zu konstatieren, dass das Thema der Evaluation von Wirkungen und Wirkungsforschung in den hier betrachteten Themenfeldern nicht neu ist und durchaus kontrovers verhandelt wird. Jenseits zunehmender Wirkungsorientierung (Albus/Micheel/Polutta 2018) und gewachsener politischer Aufmerksamkeit sind Fragen nach Wirkungen dem pädagogischen Handeln generell eingeschrieben. Pädagogisch Handelnde streben immer danach, ihre Adressat:innen verändernd oder stabilisierend zu beeinflussen. Die Erfassung von Wirkungen und das Verstehen von Wirkweisen sind damit für die geförderten Akteur:innen und ihre pädagogische Praxis hoch relevant. Gleichzeitig ist die empirische Erfassung bzw. Feststellung von Wirkungen alles andere als trivial.

Im Hinblick auf (methodisch) angemessene Wirkungsevaluationen muss zunächst das konkrete Erkenntnisinteresse der beteiligten Stakeholdergruppen bestimmt werden.¹ Unter Abwägung der konkret vorgefundenen Bedingungen, der Spezifika des in einer Wirkungsevaluation zu untersuchenden Gegenstandes (z. B. einzelne Maßnahmen, Maßnahmebündel oder Programmbereiche) und

1 Es handelt sich hierbei um „ein breites Spektrum von Akteur:innen und Institutionen – von dem politischen System über die politische Verwaltung über die Medien, die beteiligte und nicht beteiligte Fachpraxis [...] bis hin zu den betroffenen Adressat:innen [...] (sowie) [...] die an der Evaluation der Programme und den einschlägigen Fachdiskursen beteiligten wissenschaftlichen Akteur:innen und Institutionen“ (Lüders 2024).

der dafür verfügbaren Ressourcen gilt es dann, das konkrete Untersuchungsdesign für Wirkungsevaluation zu konkretisieren. Dabei geht es nicht zuletzt darum, das Problem zu lösen, Wirkungszusammenhänge bzw. entsprechende Kausalbeziehungen vor allem zwischen der Inanspruchnahme pädagogischer sowie anderer Angebote einerseits und den festgestellten Resultaten bzw. Veränderungen aufseiten der Zielgruppen andererseits adäquat abzubilden. Weitere Fragen stellen sich hinsichtlich der relevanten, zu untersuchenden Wirkdimensionen und zu den Faktoren, die beeinflussen, was, wie und wann eine Wirkung hervorbringt oder zu ihr beiträgt.

Inzwischen liegen zahlreiche Erfahrungen mit der Umsetzung von Wirkungsevaluationen in der Arbeit für Demokratie und gegen Extremismus vor (z. B. Böttcher/Dicke/Ziegler 2009; Widmer 2012; Beierle u. a. 2013; Bischoff/Zimmermann/König 2021; Bohn/Saßmannshausen 2021; Milbradt u. a. 2021; Johansson u. a. 2022). Hinzu kommen Beiträge aus der Wirkungsforschung und -evaluation in der Sozialen Arbeit, beispielsweise in der Kinder- und Jugendhilfe (Mayrhofer 2017; Begemann/Bleck/Liebig 2019; Ottmann/König 2023a, b; Baumgartner/Haunberger 2023), der politischen Bildung (Uhl/Ulrich/Wenzel 2004; Balzter/Ristau/Schröder 2014) und weiterer angrenzender Arbeitsfelder. Aus ihnen geht hervor, dass die Entscheidung für das jeweils passende, gegenstandsangemessene Evaluationsdesign vor allem in bildungs-, beratungs- und präventionsorientierten Arbeitsfeldern, die sich durch Koproduktion und oftmals fehlende Standardisierbarkeit auszeichnen, nicht auf der Basis wissenschaftstheoretischer und methodologischer Dogmen erfolgen sollte. Stattdessen verlangen verschiedenartige Untersuchungs- bzw. Evaluationsgegenstände sowie -fragestellungen unterschiedliche, auf sie angepasste Evaluationsdesigns (Nehlsen u. a. 2020; Bischoff/Zimmermann/König 2021).

Entsprechend kreisen fachliche Debatten um die Frage, wie gegenstandsangemessene Evaluationsdesigns und -methoden zum Aufspüren, Nachweis oder zur Plausibilisierung von Wirkungen auf verschiedenen Ebenen (z. B. Individuen, Projekte, Programmbereiche) in den Feldern der (pädagogischen) Demokratieförderung, Vielfaltgestaltung und Extremismusprävention aussehen sollten und können.

Hier setzt der vorliegende Sammelband an. In dessen Mittelpunkt stehen Beiträge zur Umsetzung von Wirkungsevaluationen im Bundesprogramm „Demokratie leben!“ in seiner zweiten Förderphase (2020–2024). Diese Beiträge, die den Hauptteil des Buches bilden, stammen aus dem Projekt „Programmevaluation Demokratie leben!“, das am Deutschen Jugendinstitut am Standort Halle (Saale) angesiedelt war und mehrere Teilbereiche des Bundesprogramms evaluiert hat. Eine Ausnahme bildet ein Beitrag aus einem nahestehenden Projektkontext („Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“). In den Beiträgen steht nicht die Darstellung von Ergebnissen (*Was wirkt?*) im Vordergrund, sondern vielmehr die Frage, *wie* die in den verschiede-

nen Handlungsbereichen des Bundesprogramms ermittelten Wirkungen zustandekommen und wie diese Prozesse evaluiert wurden. Damit will der Band dazu beitragen, verschiedene Ansätze der wirkungsorientierten Evaluation in der Arbeit für Demokratie und gegen Extremismus zu reflektieren und Anregungen für deren Weiterentwicklung im Kontext zukünftiger Evaluationsvorhaben zu geben.

Gerahmt werden die angesprochenen Beiträge durch einen einführenden ersten Teil, der den evaluationstheoretischen und begrifflichen Rahmen für die Praxisbeiträge absteckt, sowie durch einen Ausblick in Form eines interdisziplinären Fachgesprächs. Jenes vereint die Perspektiven ausgewiesener Evaluations- und Praxisfeldexpert:innen auf die Frage der gegenstandsangemessenen Evaluation von Wirkungen in komplexen Förderprogrammen wie „Demokratie leben!“. Der Sammelband knüpft an die Publikation *Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen* (Milbradt u. a. 2021) an. Diese hatte ausgewählte Fragen und Aspekte aus dem Kontext der Evaluation der ersten Förderphase des Bundesprogramms „Demokratie leben!“ zum Gegenstand.

Um den Leser:innen zunächst einen Eindruck von den in der Programmevaluation „Demokratie leben!“ vorgefundenen Evaluationsgegenständen zu vermitteln, wird nachfolgend kurz in die Genese der hier betrachteten programmgeförderten zivilgesellschaftlichen Arbeit für Demokratie und gegen Extremismus eingeführt und das Bundesprogramm „Demokratie leben!“ in seiner zweiten Förderphase (2020–2024) sowie seine wissenschaftliche Programmbegleitung vorgestellt. Daran schließt sich ein Überblick über die im Band verfolgten Fragestellungen und die hier versammelten Beiträge an.

1.1 Gegenstände der Programmevaluation

1.1.1 Staatlich geförderte zivilgesellschaftliche Arbeit für Demokratie und gegen Extremismus

Rechtsextremismus, Rassismus und Antisemitismus sind seit langem Teil der deutschen Gesellschaft. Das erste Programm der Bundesregierung zur Bekämpfung von Rechtsextremismus war das „Aktionsprogramm gegen Aggression und Gewalt“ (AgAG). Es stellte eine Reaktion des Staates auf die hohe Belastung der deutschen Gesellschaft mit rassistisch motivierten Straftaten und eine Serie rechtsextremer Ausschreitungen und Gewalttaten Anfang der 1990er-Jahre dar und konzentrierte sich in der Programmumsetzung auf die ostdeutschen Bundesländer (Möller 2003). Ab dem Jahr 2001 widmeten sich unterschiedliche Bundesmodellprogramme mit verschiedenen fachlichen Schwerpunktsetzungen der Förderung der Arbeit für Demokratie und gegen Extremismus (Programm-

evaluation DJI 2021, S. 352). Rechte, rassistische und antisemitische Gewalt sind auch heute weiterhin in der deutschen Gesellschaft präsent. Ebenso sind die diesen Phänomenen zugrundeliegenden Ungleichwertigkeitsvorstellungen in der Bevölkerung (z. B. in Form von Antisemitismus, Rassismus, Homosexuellen- und Trans*feindlichkeit oder Sozialdarwinismus) nach wie vor weit verbreitet. Inzwischen werden sie sogar offener als noch vor zehn Jahren artikuliert (Decker u. a. 2022; Groh-Samberg u. a. 2022). Zudem haben sie sich zum konstituierenden Motiv populistischer und antidemokratischer Bewegungen und Parteien entwickelt, die ihrerseits eine immer breitere Anhängerschaft finden (Zick/Küpper/Mokros 2023). Das Internet und die Sozialen Medien beschleunigen und verstärken diese Entwicklung (ebd., S. 108 f., 114 f.).

Neben Rechtsextremismus und -populismus hat sich mit der Zeit auch der islamistische Extremismus zu einem im öffentlichen Bewusstsein und auf der politisch-administrativen Ebene präsenten Phänomen entwickelt. Insbesondere die Terroranschläge im Jahr 2001 und verschiedene Erscheinungsformen des „homegrown-terrorism“ erzeugten eine stark erhöhte gesellschaftliche Relevanz des Phänomens. Diese steigerte sich mit der Etablierung des sogenannten „Islamischen Staates“ im Jahr 2014 und den folgenden Terroranschlägen in Europa sowie der Ausreise zahlreicher deutscher Jugendlicher und junger Erwachsener in die Kampfgebiete in Syrien und dem Irak nochmals deutlich (Ceylan/Kiefer 2018). Im Ergebnis sahen sich sowohl der Bund als auch die Länder veranlasst, neben sicherheitspolitischen Maßnahmen verstärkt Präventionsarbeit in zivilgesellschaftlicher Trägerschaft zu fördern (Figlesthler/Schau 2021).

Die Bearbeitung des Themas Islamistischer Extremismus erfolgte zunächst in einem eigenen Programm, der „Initiative Demokratie stärken“ (2010–2014), zusammen mit der Prävention von Linksextremismus. Seit dem Start des Bundesprogramms „Demokratie leben!“ im Jahr 2015 werden sowohl Rechtsextremismus als auch islamistischer und Linksextremismus unter einem Dach bearbeitet.²

Betrachtet man die Umsetzungspraxis der verschiedenen Programme genauer, ist zu sehen, dass sie nie „nur“ Präventionsarbeit gefördert haben, sondern eine Doppelstrategie verfolgten. Diese bestand und besteht bis heute darin, Präventionsarbeit zu politisch-weltanschaulichem Extremismus (und Gruppenbezogener Menschenfeindlichkeit) mit der Förderung von demokratischem Denken und Handeln sowie von Strukturen und Verfahren zu verbinden, die diese stützen (Bischoff u. a. 2012, S. 24 f.; Programmevaluation DJI 2021, S. 352). Inzwischen hat sich der inhaltliche Schwerpunkt des Bundesprogramms „Demokratie leben!“ weiter von der Extremismusprävention weg hin zur Stärkung von Demokratie, einschließlich der Resilienz gegenüber Demokratiegefährdungen,

2 Versuche, zivilgesellschaftliche Akteur:innen mittels Programmförderung für eine breite Auseinandersetzung mit Linksextremismus zu gewinnen, waren zumindest bisher wenig erfolgreich (Grunow 2022).

verschoben. Demokratieförderung – im gerade umrissenen Sinn – hat sich neben Extremismusprävention als eigenständiges Ziel fest etabliert (Jütz u. a. 2023, S. 11; BMFSFJ 2020, S. 100).

Darüber hinaus verknüpften die einschlägigen Bundesprogramme die Förderung demokratischer Kompetenzen, Verfahren und Strukturen frühzeitig auch mit migrationsbezogenen Themen (Programmevaluation DJI 2021, S. 352). Ausgangspunkt war dabei die Feststellung, dass Migration und die damit einhergehende kulturelle, religiöse, ethnische und staatsbürgerschaftliche Diversifizierung der Bevölkerung die Gesellschaft kontinuierlich neu ordnen (Foroutan/Ikiz 2016, S. 139). Zugleich erwuchs daraus im politischen Diskurs zunehmend die Bereitschaft anzuerkennen, dass Deutschland ein Einwanderungsland ist (ebd., S. 138). Mit dem Bundesprogramm „Demokratie leben!“ entwickelte sich aus der zunächst gemeinsamen Bearbeitung von Demokratie und Vielfalt und der Prävention verschiedener Elemente des Syndroms der Gruppenbezogenen Menschenfeindlichkeit (vor allem Rassismus, Antisemitismus, Antiziganismus und Homosexuellen- und Trans*feindlichkeit) ein eigener thematischer Schwerpunkt unter dem Label Vielfaltgestaltung.

Im Folgenden sollen kurz die wesentlichen Charakteristika der zweiten Förderphase des Bundesprogramms „Demokratie leben!“ und seiner Evaluation vorgestellt werden, da diese den Kontext für die im Band enthaltenen Beiträge darstellen.

1.1.2 Das Bundesprogramm „Demokratie leben!“ (2020–2024)

„Demokratie leben!“ ist eine zentrale Säule der von 2016 bis 2024 geltenden Strategie der Bundesregierung zur Extremismusprävention und Demokratieförderung (BMFSFJ/BMI 2016). Es soll kinder- und jugendpolitische „Maßnahmen zur Stärkung von Demokratie und Vielfalt im Wege der Modellprojektförderung mit bundeszentraler Ausstrahlung vor allem bei Kindern und Jugendlichen“ (BMFSFJ 2022, S. 12) fördern. Darüber hinaus adressiert das Programm eine Reihe an sozialisationsrelevanten Akteur:innen, vor allem in der Kinder- und Jugendhilfe sowie im Bildungssystem, und fördert Projekte, die weitere Zielgruppen wie junge Erwachsene, von Rassismus und rechter Gewalt Betroffene oder auch breite Teile der Bevölkerung ansprechen.

Im Einzelnen geht es darum,

- innovative pädagogische Ansätze der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention zu entwickeln und zu erproben,
- die Vernetzung bundeszentraler zivilgesellschaftlicher Träger sowie von Engagement auf der kommunalen Ebene und auf der Ebene der Bundesländer anzuregen,

- die Beratung von Betroffenen rechter, rassistischer und antisemitischer Gewalt sowie die Mobile Beratung gegen Rechtsextremismus sicherzustellen und
- die Ausstiegs- und Distanzierungsberatung in den Kontexten von Rechtsextremismus sowie islamistischem Extremismus und „(sozial-)pädagogische Arbeit mit rechtsextrem oder islamistisch orientierten Jugendlichen“ zu gewährleisten (BMFSFJ/BMI 2016).

Um die gerade umrissenen Zielstellungen zu verfolgen, sind in „Demokratie leben!“ mehrere Programmbereiche eingerichtet (im Programm „Handlungsbereiche“ genannt). Die Projekte in den einzelnen Bereichen verfolgen je eigene Handlungsstrategien und sind auf verschiedene föderale Ebenen hin ausgerichtet:

- Im *Handlungsbereich Kommune* werden Partnerschaften für Demokratie gefördert, die das kommunale Engagement für Demokratie unterstützen und vernetzen.
- Im *Handlungsbereich Land* werden Landes-Demokratiezentren gefördert, die die Arbeit im Bundesprogramm in der Auseinandersetzung mit Rechtsextremismus und Islamismus auf der Landesebene koordinieren und die Beratungsarbeit in der Auseinandersetzung mit Rechtsextremismus und Islamismus sowie von Betroffenen rechter, rassistischer und antisemitischer Gewalt unterstützen.
- Im *Handlungsbereich Bund* wird die Arbeit von Kompetenznetzwerken und -zentren gefördert, die zur Qualifizierung von Fachkräften und zum Transfer von Wissen, Erkenntnissen und Arbeitsansätzen in Regelstrukturen beitragen soll.
- In den *Handlungsbereichen Modellprojekte und Innovationsfonds* sollen neue Arbeitsansätze in den Themenfeldern des Bundesprogramms entwickelt, erprobt und verbreitet werden.

Die Projektaktivitäten in den Handlungsbereichen werden durch die Förderung von zahlreichen Begleit- und Forschungsprojekten, von Maßnahmen des programmbezogenen Fachaustauschs und der Öffentlichkeitsarbeit sowie von wissenschaftlicher Begleitung und Evaluation flankiert.

Damit vereint das Programm eine Reihe unterschiedlicher Phänomene, die zahlreiche Akteur:innen mit unterschiedlichsten Handlungsansätzen bearbeiten. Als Programm der Bundesregierung konzentriert es sich im Rahmen der föderalen Kompetenzordnung darauf, Anregungen bzw. Impulse für die Arbeit in seinen Themenfeldern zu geben. Es ist nicht dafür gedacht, *flächendeckend* Extremismus *direkt* zu bekämpfen oder demokratische Einstellungen sowie die Akzeptanz von Vielfalt *bundesweit* in der Breite *unmittelbar* zu fördern. Eine der Hauptaufgaben von „Demokratie leben!“ besteht „in der *beispielhaften* Entwicklung und Erprobung von Konzepten, Strategien und Arbeitsformen zur Extremismusprävention und Demokratieförderung. Bundesprogramme sind in diesem Sinne vorrangig

als ‚exemplarische Feldexperimente, in denen stellvertretend für den jeweiligen Bereich Erfahrungen insbesondere mit innovativen Ansätzen gesammelt werden‘ [Lüders 2003, S. 4] zu verstehen. Genau in diesem Sinne sind sie dann auch als eine Form der Umsetzung der Anregungsfunktion des Bundes (§ 83 SGB VIII) gegenüber der Fachpraxis zu betrachten [vgl. Haubrich/Lüders/Struhkamp 2007]⁴ (BT-Drucks. 18/12743, S. 4, Hervorhebung d. V.). Das ist gerade mit Blick auf Wirkungsuntersuchungen im Rahmen der wissenschaftlichen Programmbegleitung von Bundesprogrammen wie „Demokratie leben!“ bedeutsam.

Aufgrund seiner Anlage als Modellprogramm, das dazu motiviert und hilft neue Arbeitsansätze, Organisationsstrukturen und Netzwerke in seinen Themenfeldern hervorzubringen, sind es insoweit nicht globale Veränderungen auf der Ebene des Bundesprogramms als Ganzes oder gesellschaftliche Veränderungen, die in den Fokus der in diesem Band vorgestellten Wirkungsuntersuchungen geraten. Stattdessen wird untersucht, inwieweit sich die Wirksamkeitserwartungen an die Projekte in den konkreten Aktivitäten mit Zielgruppen als auch in Hinblick auf ihre Anregungsfunktion für die Fachpraxis erfüllen. In Bezug auf Letzteres sorgen die Projekte im Idealfall dafür, dass die modellhaften Erprobungen, die sich als wirksam herausstellen, über das Bundesprogramm hinausgehend in der Fachpraxis aufgenommen und umgesetzt werden.

Angesichts der Vielgestaltigkeit der programmgeförderten Gegenstände bzw. Projekte sind die Ansätze für deren Evaluation, um gegenstandsangemessen zu sein, ebenfalls vielgestaltig. Ein einheitliches Vorgehen würde hier zu kurz greifen. Das spiegelt sich auch in der Ausrichtung der in diesem Band vorgestellten (Wirkungs-)Evaluationsdesigns wider.

Damit sind zentrale Eckpfeiler für die Evaluation der verschiedenen Teilbereiche von „Demokratie leben!“ umrissen. Dies gilt insbesondere für Fragen von Wirkungsevaluationen, die untersuchen, welche Veränderungen oder Stabilisierung bzw. Wirkungen eine Aktivität, Maßnahme oder ein Programm hervorruft. Das Bundesprogramm beeinflusst hierbei nicht nur den Wirkungsradius der geförderten Praxisprojekte, sondern auch deren Untersuchung durch die wissenschaftliche Programmbegleitung.

1.2 Die wissenschaftliche Programmbegleitung von „Demokratie leben!“

Die einzelnen Handlungsbereiche des Bundesprogramms „Demokratie leben!“ werden unabhängig voneinander wissenschaftlich begleitet und evaluiert. Für die Handlungsbereiche Kommune, Land und Bund und den Innovationsfonds ist jeweils eine wissenschaftliche Begleitung zuständig. Im Handlungsbereich Modellprojekte sind insgesamt vier wissenschaftliche Begleitungen für jeweils

einen spezifischen Themenbereich verantwortlich. Hinzu kommt mit der „Gesamtevaluation“ eine Arbeitseinheit, deren Gegenstand das Bundesprogramm als Ganzes ist und die programmbereichsübergreifenden Fragen nachgeht.

Die Evaluation im Bundesprogramm erfolgt durch vier Institutionen: Neben dem Deutschen Jugendinstitut (DJI) sind das Institut für Sozialarbeit und Sozialpädagogik (ISS), das Deutsche Zentrum für Integration und Migration (DeZIM) und Camino – Werkstatt für Fortbildung, Praxisbegleitung und Forschung im sozialen Bereich gGmbH beteiligt.

Alle am DJI angesiedelten Evaluationseinheiten sind im Projekt „Programmevaluation ‚Demokratie leben!‘“ am Standort Halle (Saale) zusammengefasst. Dort werden die Handlungsbereiche Land und Bund sowie die Modellprojekte im Handlungsfeld der Demokratieförderung und im Themenfeld Prävention und Deradikalisierung in Strafvollzug und Bewährungshilfe wissenschaftlich begleitet. Ebenfalls dort zu finden sind die Gesamtevaluation und ein Forschungsprojekt zur Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention im Kontext von „Demokratie leben!“.

1.3 Überblick über die Beiträge des Bandes

Der Band gliedert sich in drei Teile. Der erste Teil führt zunächst in das Thema Wirkungen und ihre Analyse sowie das damit verbundene Thema der Gegenstandsangemessenheit von Wirkungsevaluationen ein. Daran anschließend stellen die Autor:innen von fünf Beiträgen ihre jeweiligen Vorgehensweisen und Erfahrungen bei der Umsetzung von Wirkungsuntersuchungen im Bundesprogramm „Demokratie leben!“ vor und diskutieren diese. Ein weiterer Beitrag reflektiert die Untersuchung von Bedarfslagen als einen möglichen Ausgangspunkt von Wirkungsuntersuchungen. Der Band schließt mit einem interdisziplinären Fachgespräch, in dem sich namhafte Expert:innen zu ihren Erfahrungen mit und Perspektiven auf empirische Untersuchungen von Wirkungen im Bereich der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention austauschen sowie Anregungen für die Anlage und Umsetzung solcher Untersuchungen geben.

Wirkungen und Gegenstandsangemessenheit

Einen über die Evaluation des Bundesprogramms „Demokratie leben!“ hinausweisenden Einstieg in das Thema des vorliegenden Bandes ermöglicht Jan Ulrich Hense mit der Frage: *Wie ernst ist es uns wirklich mit der Wirkung? Haupt- und Nebenfolgen eines Begriffs mit hoher Anmutungsqualität.* Der übergreifende Beitrag befasst sich aus der Perspektive der Evaluationsforschung allgemein mit dem Thema der Wirkungen und ihrer Evaluation. Er nimmt die Beobachtung zum Ausgangspunkt, dass seit einiger Zeit in der Diskussion um Qualitätsfragen und Evalua-

tion zentrale Fachtermini leichtfertig, bisweilen alltagsweltlich und verkürzend verwendet würden. Das betreffe auch den Wirkungsbegriff, der allerdings, so Jan Ulrich Hense, einigen unsichtbaren Ballast mit sich führe. Vor allem die Notwendigkeit, ein als Wirkung verstandenes Ereignis kausal auf eine Ursache zurückzuführen – und dies oft in komplexen Bedingungsgefügen – würde in der Praxis regelmäßig ignoriert. Bestimmte Wirkungen hervorzurufen und nachzuweisen, werde nicht selten vorschnell versprochen, ohne den Anspruch später einzulösen. Vor diesem Hintergrund spannt der Beitrag einen Bogen über verschiedene Möglichkeiten, angemessen zum einen zu untersuchen, ob Wirkungen eingetreten sind und zum anderen, wie die entsprechenden Veränderungsprozesse zustandekommen. Zunächst geht Jan Ulrich Hense auf experimentelle Ansätze zur Wirkungsanalyse, deren Potenziale und Fallstricke, ein. Diese Ansätze messen Unterschiede in der Veränderung der Zielvariablen zwischen mindestens zwei idealerweise zufallsverteilt zusammengesetzten Gruppen, jeweils mit und ohne „Treatment“, und weisen so nach, ob eine bestimmte Intervention gewirkt hat und ggf. wie stark. Solche Verfahren sind in der Umsetzung methodisch anspruchsvoll und die Grundvoraussetzungen für ihre Umsetzung nach Ansicht des Autors in den Feldern der pädagogischen Demokratieförderung und Extremismusprävention oft nicht gegeben. Außerdem tragen jene Ansätze in Evaluationskontexten nur begrenzt dazu bei, die für die Praxis relevante Frage zu klären, wie beobachtete Wirkungen konkret zustande kommen. Vor diesem Hintergrund setzt sich der Autor kritisch mit dem mitunter artikulierten Anspruch auseinander, dass es sich hierbei um den „Goldstandard“ der Wirkungsevaluation handele.

Ist die Umsetzung experimenteller oder quasi-experimenteller Designs nicht möglich, mit Blick auf die Untersuchungsfrage oder den -gegenstand nicht sinnvoll oder nicht ausreichend, kommen, so Jan Ulrich Hense, als Alternativen oder Ergänzungen dazu u. a. Konzepte infrage, die den Wirkungsuntersuchungen konkrete „Theorien“ (empirisch oder theoretisch begründete Annahmen) zum Evaluationsgegenstand zugrunde legen. Diese Konzepte können sowohl *credible evidence* erzeugen als auch der untersuchten Praxis vertieftes Lernen ermöglichen, wenn sie sich insbesondere der Frage danach widmen, wie Wirkungen zustandekommen. Der Autor legt in seinen Ausführungen einen Schwerpunkt auf drei Ansätze: Realist Evaluation, Contribution Analysis und Process Tracing, in die er kurz mit ihren Potenzialen und Beschränkungen einführt. In seinen Ausführungen bricht Jan Ulrich Hense eine Lanze dafür, zum einen mit dem Wirkungsbegriff selbstbewusster umzugehen und zum anderen nicht nur die Analyse von Wirkungen theoriebasiert durchzuführen, sondern bereits die Planung und Implementation der analysierten bzw. evaluierten Programme soweit es geht theoretisch zu fundieren. Das ermögliche sowohl eine sachgerechtere Praxis, im konkreten Fall hier in der Demokratieförderung und Extremismusprävention, als auch deren gegenstandsangemessene Evaluation.

Die Frage, wie Wirkungsuntersuchungen in den Themenfeldern Demokratieförderung, Vielfaltgestaltung und Extremismusprävention gegenstandsangemessen ausgestaltet werden können, steht im Mittelpunkt des Beitrags von Frank Greuel. Der Autor stellt einleitend mit einem Fokus auf qualitative Forschungsansätze dar, was unter Gegenstandsangemessenheit in der empirischen Sozialforschung sowie der Evaluation verstanden wird. Auf dieser Basis spezifiziert er dann, was *Gegenstandsangemessenheit in Wirkungsevaluationen*, insbesondere im Bereich der Evaluation von Wirkungen pädagogischer Arbeit in den zentralen Themenfeldern des Bundesprogramms „Demokratie leben!“ (Demokratieförderung, Vielfaltgestaltung und Extremismusprävention), bedeutet. Dabei plädiert der Beitrag dafür, sicherzustellen, dass sowohl die untersuchten Wirkungsausschnitte bzw. -dimensionen als auch das Evaluationsdesign einschließlich des methodischen Zugriffs geeignet sind, mindestens die in den Maßnahmen beabsichtigten und realistisch zu erwartenden Wirkungen zu erfassen. Dabei sollten Evaluierende und auch Evaluierete genau prüfen, inwieweit jeweils die Voraussetzungen für die Anwendung wirkungsbezogener Evaluationsdesigns gegeben sind.

Evaluation von Wirkungen im Bundesprogramm „Demokratie leben!“

Wie Wirkungsevaluationen im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) gegenstandsangemessen in verschiedenen Handlungsfeldern von „Demokratie leben!“ umgesetzt wurden und welche Erfahrungen die Evaluator:innen dabei gemacht haben, steht im Mittelpunkt des zweiten Teils dieses Sammelbandes. Dieser beginnt mit einem Beitrag von Carmen Figlestahler, Katrin Haase und Katja Schau zu *Wirkungsorientierte[-r] Evaluation von Beratung nach der Realist Evaluation* und widmet sich der empirischen Rekonstruktion von Wirkmechanismen von Beratungsarbeit im Kontext Rechtsextremismus. Evaluationstheoretischer Bezugspunkt ist der Ansatz der Realist Evaluation von Pawson und Tilly (1997). Gegenstand der Betrachtung sind konkrete Fallkonstellationen aus der Beratung von Betroffenen rechter, rassistischer und antisemitischer Gewalt sowie der Ausstiegs- und Distanzierungsberatung im Bereich Rechtsextremismus. Anhand dieser stellen die Autorinnen dar, wie mithilfe der Identifikation spezifischer Kontext-Mechanismus-Outcome-Konfigurationen unter Einbezug von Theorien der Sozialen Arbeit praxisnahe und professionstheoretisch gerahmte Aussagen dazu möglich sind, wodurch Beratungshandeln in den betrachteten Kontexten wie für wen wirkt. Vor dem Hintergrund bestehender Diskussionen darum, was in der Realist Evaluation unter „Mechanismus“ verstanden werden soll (Lemire u. a. 2020), stellen die Autorinnen eine beratungsfeldbezogene, einzelfallübergreifende Konzeptualisierung von „Mechanismus“ als „haltungsbasiertes Handeln“ vor, die Ausgangspunkt für die konkrete Beschreibung fallspezifischer Mechanismen ist. Insoweit stellt der Beitrag nicht nur anschauliche

Anwendungsfälle von Realist Evaluation vor, sondern leistet einen Beitrag dazu, das Konzept des „Mechanismus“ kontextspezifisch zu schärfen.

Ebenfalls am Konzept der Realist Evaluation orientierten sich die empirischen Wirkungsuntersuchungen zu Modellprojekten der Demokratieförderung, die Aline Rehse, Max Hemmann und Marina Braun in ihrem Beitrag *Wirkungsevaluation im Bereich Demokratieförderung* präsentieren. Die Autor:innen kombinieren diesen Ansatz mit der Herangehensweise der Cluster-Evaluation (Haubrich 2006; Worthen/Schmitz 1997). Sie stellen ausgewählte Erkenntnisse aus der wissenschaftlichen Begleitung von (sozial-)pädagogischen Modellprojektaktivitäten auf Zielgruppenebene dar und diskutieren ihr Vorgehen. Dieses reflektieren sie vor allem bezogen auf Herausforderungen, die sich aus der Komplexität des von ihnen evaluierten Gegenstandes und den Rahmenbedingungen der Evaluation (z. B. Bindung an die Programmlaufzeit) in Verbindung mit begrenzten Ressourcen sowohl der Evaluator:innen als auch der Evaluierten ergeben haben. Der Beitrag erörtert vor diesem Hintergrund u. a. die Frage, inwieweit die verfolgten Evaluationsansätze modifiziert werden können, um zu generalisierbaren Ergebnissen in den Wirkungsuntersuchungen auf Clusterebene zu kommen oder wie längerfristige Wirkungen empirisch besser untersucht werden könnten.

Die von Jonas Euteneuer und Stella Rüger in diesen Band eingebrachten Erfahrungen mit Wirkungsuntersuchungen *„Das hätte ich, dann jetzt rückblickend betrachtet, noch anders gestaltet“* stammen ebenfalls aus dem Kontext der wissenschaftlichen Begleitung von Modellprojekten der Demokratieförderung. Sie interessieren sich in ihrem Beitrag dafür, inwieweit sich basierend auf der Kombination von Clusterevaluation und Realist Evaluation systematisch nicht-intendierten Wirkungen im Sinne von „Lernen aus Scheitern“ empirisch auf die Spur kommen lässt. Diese Frage ist für Modellprojekte besonders relevant, weil ihr Auftrag, Innovationen zu entwickeln, latent immer auch impliziert, mit ihren „exemplarischen Feldexperimenten“ (Haubrich/Lüders/Struhkamp 2007, S. 184) zu scheitern. Zugleich scheint diese Wirkdimension bei der Formulierung von Theories of Change – als Ausgangspunkt von Wirkungsuntersuchungen im Rahmen von Realist Evaluation – bisher eher selten in den Fokus von Evaluation zu rücken. Der Beitrag macht außerdem deutlich, dass sich Praxisprojekte durchaus auch schwer damit tun, Erfahrungen über ein Scheitern für Evaluationen zugänglich zu machen. Über die daraus resultierende Frage hinaus, wie mittels Realist Evaluation nicht-intendierte Wirkungen festgestellt und plausibilisiert werden können, diskutieren der Autor und die Autorin, inwieweit Evaluationsdesigns zu Gegenständen wie Modellprojekten auch die Folgewirkungen von Projektinterventionen (unabhängig davon, ob diese gelingen) konzeptionell berücksichtigen sollten.

Einen möglichen Ausgangspunkt dafür stellen Franziska Heinze, Sarah Langer, Steffen Loick Molina, Stefanie Reiter, Kornelia Sammet und Ellen Schroeter in ihrem Beitrag *Resonanz ist mehr als die Sehnsucht nach Wiederhall* vor. Sie machen

für die in ihrem Beitrag vorgestellten empirischen Wirkungsuntersuchungen ein der Akustik entlehntes und erkenntnistheoretisch übersetztes Konzept von „Resonanz“ fruchtbar, welches sich von dem prominenten Begriffsverständnis Hartmut Rosas (2016) absetzt. Dabei greifen sie auf Überlegungen von Breyer und Gerner (2017) zurück und stellen die Wechselverhältnisse zum einen zwischen Wirkungen hervorrufender Intervention und Adressat:in bzw. Zielstruktur dieser Intervention (Wirkort) und zum anderen zwischen verschiedenen Wirkmechanismen und Wirkungen, die zeitgleich oder in einer zeitlichen Abfolge innerhalb sozialer Systemen auftreten, in den Mittelpunkt ihrer Betrachtung. Insoweit spiegelt sich im Beitrag eine Perspektive auf Wirkungen und ihre Untersuchung wider, für die sich auch Rehse/Hemmann/Braun in diesem Band unter Verweis auf Bergeron/Gaboury (2020, S. 3) stark machen und die beinhaltet, einzelne festgestellte Wirkungen nicht isoliert, sondern als Element miteinander verwobener Wirkungsgefüge zu betrachten. Letztere als Resonanzbeziehungen zu verstehen, sensibilisiert nicht nur für das Moment der Koproduktion, aus dem heraus Wirkungen entstehen, sondern lenkt den Blick über die empirische Betrachtung singulärer Ursache(-n)-Wirkungs-Beziehungen hinaus auf die Interferenzen von Wirkungen, Folgewirkungen und deren Entstehungsprozesse.

Die Frage, *Wie bringt ein Bundesprogramm die Zivilgesellschaft „in Aktion“?* nehmen Frank König, Alexander Stärck und Diana Zierold in ihrem Beitrag zum Ausgangspunkt, um zu illustrieren, wie sich das Erreichen abstrakter Programmleitzielle anhand einer vergleichenden Studie auf regionaler Ebene feststellen und als Programmwirkung plausibilisieren lässt. Sie stellen anhand von zwei Beispielregionen zum einen dar, wie sie mithilfe von Outcome Harvesting (Wilson-Grau 2018) sowie Netzwerkanalysen die Beiträge untersucht haben, die das Bundesprogramm mit seiner Förderung zur Stärkung von zivilgesellschaftlichem Engagement in seinen drei Themenbereichen leistet. Zum anderen zeigen sie, dass eine Analyse von zentralen Kontextfaktoren dieser Förderung, hier des Zusammenspiels bundesprogrammgeförderter Akteur:innen mit den unabhängig vom Bundesprogramm bestehenden Aktivitäten und Strukturen, Aufschlüsse über Wirkmechanismen zu der untersuchten Wirkdimension geben kann, auch wenn dem keine explizite Programmtheorie auf der regionalen Ebene zugrunde lag. Anhand der Studie veranschaulichen die Autor:innen, dass derartige Evaluationsdesigns selbst auf der Ebene einzelner Fallstudien ressourcenaufwendig sind. Dabei verweisen sie nicht zuletzt auf das oftmals existierende Missverhältnis zwischen Ansprüchen an die Aussagekraft von Wirkungsuntersuchungen und den dafür nötigen Voraussetzungen.

Einer dieser Voraussetzungen, nämlich der einer evidenzbasierten und bedarfsorientierten Programmkonzeption und -planung, widmet sich der Beitrag *Die Analyse von Bedarfen in den Bereichen Demokratieförderung und Extremismusprävention* von Laura Meijer und Pia Saueremann. Auf Basis eigener Bedarfserhebungen stellen die Autorinnen grundsätzliche Überlegungen darüber an, wie solche Ana-

lysen zur Vorbereitung von Förderprogrammen angelegt und umgesetzt werden sollten, um nicht nur eine bedarfsangemessene Programmplanung zu ermöglichen, sondern um zugleich auch eine geeignete Grundlage für Wirkungsevaluationen zu bilden. In diesem Zusammenhang geht der Beitrag auf Potenziale und Fallstricke ein, die eine Verknüpfung von Wirkungsevaluation und Bedarfsanalysen mit sich bringen kann. Die Autorinnen greifen damit das zuletzt wieder intensiver diskutierte Thema der evidenzbasierten Politikgestaltung und Programmplanung sowie -steuerung auf.

Statt eines Fazits

Die Frage, inwiefern Wirkungsevaluationen relevantes Wissen für Lernprozesse der Programmpraxis sowie für die Steuerung und Verwaltung von Programmen bereitstellen können, wird auch im Rahmen eines *interdisziplinären Fachgesprächs* verhandelt, welches die Herausgeber:innen dieses Bandes mit fünf evaluationserfahrenen Expert:innen aus der Praxis und Wissenschaft führten. Helle Becker (für das Feld der politischen Bildung), Jörg Faust (für die Entwicklungszusammenarbeit), Hemma Mayrhofer (für die Soziale Arbeit), Björn Milbradt (für die Felder der Demokratieförderung und Extremismusprävention) und Thomas Widmer (für eine politikwissenschaftliche Sicht auf Wirkungsevaluationen im Feld der Demokratieförderung) berichten in diesem Gespräch über ihre Einsichten aus der wissenschaftlichen Befassung und praktischen Erfahrungen mit wirkungsorientierter Evaluation. Sie diskutieren Ansätze der Wirkungsevaluation und formulieren ihre je durchaus unterschiedlichen Anforderungen an valide Designs zur Untersuchung bzw. Messung der Wirkungen eines Programms. Neben zentralen Orientierungspunkten dafür sowie existierenden Fallstricken und Leerstellen im Bereich der Wirkungsforschung, vor allem in der politischen Bildung und in Programmen der Demokratieförderung, reflektieren die Expert:innen auch konkrete Anforderungen an die Evaluation von Bundesprogrammen wie „Demokratie leben!“. Dabei werden sowohl Bedingungen für eine gute institutionelle Aufstellung von Evaluation als auch die Grenzen für empirische Wirkungsuntersuchungen sowie die Verwendung der dabei gewonnenen Erkenntnisse in Prozessen der Politikformulierung deutlich. Alle Teilnehmenden sprechen sich entschieden dafür aus, empirischen Wirkungsuntersuchungen trotz aller Herausforderungen nicht von vornherein aus dem Weg zu gehen. Wesentlich ist allerdings, dass Evaluator:innen, Auftraggeber:innen (v. a. Fördermittelgeber:innen oder Praxisprojekte/-träger) sowie die evaluierten Projekte und Träger gemeinsam die Voraussetzungen dafür schaffen, dass solche Evaluationen gegenstandsangemessen durchgeführt werden können und sowohl für die Praktiker:innen als auch für die Auftraggeber:innen einen Mehrwert stiften.

Literatur

- Albus, S./Micheel, H.-G./Polutta, A. (2018): Evaluation und Wirkungsorientierung. In: Böllert, K. (Hrsg.): Kompendium Kinder- und Jugendhilfe. Wiesbaden: Springer VS, S. 1563–1579.
- Balzter, N./Ristau, Y./Schröder, A. (2014): Wie politische Bildung wirkt. Wirkungsstudie zur biographischen Nachhaltigkeit politischer Jugendbildung. Schwalbach/Ts.: Wochenschau Verlag.
- Baumgartner, E./Haunberger, S. (2023): Wirkungsevaluationen in der Sozialen Arbeit. Ein Orientierungsbuch für die Praxis. Bern: haupt Verlag.
- Begemann, M.-C./Bleck, C./Liebig, R. (Hrsg.) (2019): Wirkungsforschung zur Kinder- und Jugendhilfe. Grundlegende Perspektiven und arbeitsfeldspezifische Entwicklungen. Weinheim und Basel: Beltz Juventa.
- Begemann, M.-C./Bleck, C./Liebig, R. (Hrsg.) (2019): Empirische Wirkungsforschung zur Kinder- und Jugendhilfe. Konzeptionelle Vorüberlegungen und Einführung in den Sammelband. In: Begemann, M.-C./Bleck, C./Liebig, R. (Hrsg.): Wirkungsforschung zur Kinder- und Jugendhilfe. Grundlegende Perspektiven und arbeitsfeldspezifische Entwicklungen. Weinheim und Basel: Beltz Juventa, S. 7–20.
- Beierle, S./Bischoff, U./Greuel, F./Johansson, S./König, F./Reiter, S./Zierold, D./Zimmermann, E. (2013): Leistungen und Wirkungen. Ergebnisse der Programmevaluation des Bundesprogramms „TOLERANZ FÖRDERN – KOMPETENZ STÄRKEN“. Zwischenbericht 2013. Halle (Saale): DJI.
- Bergeron, D. A./Gaboury, I. (2020): Challenges Related to the Analytical Process in Realist Evaluation and Latest Developments on the Use of NVivo from a Realist Perspective. In: International Journal of Social Research Methodology, 23. Jg., H. 3, S. 355–365.
- Bischoff, U./Greuel, F./Johansson, S./König, F./Zierold, D./Zimmermann, E. (2012): Handlungsansätze in der Rechtsextremismusprävention. Ergebnisse der Programmevaluation des Bundesprogramms „TOLERANZ FÖRDERN – KOMPETENZ STÄRKEN“. Zwischenbericht 2012. Halle (Saale): DJI.
- Bischoff, U./Zimmermann, E./König, F. (2021): Erkennen, was wirkt. Die Erprobung von Ansätzen der Wirkungsuntersuchung in der Evaluation von Bundesmodellprogrammen der Demokratieförderung und Extremismusprävention und die damit gemachten Erfahrungen. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 244–268.
- BMFSFJ (Hrsg.) (2020): Abschlussbericht Bundesprogramm Demokratie leben! Erste Förderperiode (2015 bis 2019). Berlin.
- BMFSFJ (Hrsg.) (2022): Bundeshaushaltsplan 2022. Einzelplan 17. Berlin
- BMFSFJ/BMI (Hrsg.) (2016): Strategie der Bundesregierung zur Extremismusprävention und Demokratieförderung. Berlin.
- Bohn, I./Saßmannshausen, J. (2021): Wirkungen komplexer Programm – Möglichkeiten und Grenzen evidenzbasierter Evaluation am Beispiel der Partnerschaften für Demokratie. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa S. 269–295.
- Böttcher, W./Dicke, J. N./Ziegler, H. (Hrsg.) (2009): Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. 2. Aufl. Münster: Waxmann.
- Breyer, T./Gerner, A. (2017): Resonanz und Interaktion. Eine philosophische Annäherung anhand zweier Proben. In: Breyer, T./Buchholz, M. B./Hamburger, A./Pfänder, S./Schumann, E. (Hrsg.): Resonanz – Rhythmus – Synchronisierung. Bielefeld: transcript, S. 33–46.
- Ceylan, R./Kiefer, M. (2018): Radikalisierungsprävention in der Praxis. Antworten der Zivilgesellschaft auf den gewaltbereiten Neosalafismus. Wiesbaden: Springer VS.

- Clark, Z./Dollinger, B./Heppchen, S. (Hrsg.) (2024): „Kausalität und die Re-/Konstruktion von Folgen sozialer Hilfen“ (neue praxis, Sonderheft 18). Lahnstein: Verlag neue praxis.
- Decker, O./Kiess, J./Heller, A./Brähler, E. (Hrsg.) (2022): Autoritäre Dynamiken in unsicheren Zeiten. Neue Herausforderungen – alte Reaktionen? Leipziger Autoritarismus Studie 2022. Gießen: Psychosozial-Verlag.
- Figlesthler, C./Schau, K. (2021): Prävention und Sicherheit. Ein (Rück-)Blick auf die Präventions- und Distanzierungslandschaft in Deutschland. In: Ligante. Fachdebatten aus der Präventionsarbeit. H. 4, S. 13–22. https://www.bag-relex.de/wp-content/uploads/2021/12/BAG-RelEx_Ligante4_2021_Online.pdf (Abruf 24.06.2024).
- Foroutan, N./Ikiz, D. (2016): Migrationsgesellschaft. In: Mecheril, P. (Hrsg.): Handbuch Migrationspädagogik. Weinheim und Basel: Beltz, S. 138–151.
- Groh-Samberg, O./Gerlitz, J.-Y./Minkus, L./Lengfeld, H./Dilger, C./Reinecke, J./Verneuer-Emre, L./Wächter, M./Skolarski, S./Traunmüller, R. (2022): Gefährdeter Zusammenhalt. Polarisierung und Spaltungstendenzen in Deutschland. Ausgewählte Ergebnisse der FGZ-Pilotstudie 2020. Bremen: SOCIUM, Forschungsinstitut Gesellschaftlicher Zusammenhalt.
- Grunow, D. (2022): Linksextremismusprävention in Deutschland – Entstehungsgeschichte, Feldbeschreibung und fachliche Herausforderungen. In: Milbradt, B./Frank, A./Greuel, F./Herding, M. (Hrsg.): Handbuch Radikalisierung im Jugendalter: Phänomene, Herausforderungen, Prävention. Opladen, Berlin und Toronto: Barbara Budrich, S. 335–356.
- Haubrich, K. (2006): Wirkungsannahmen sichtbar machen: Cluster-Evaluation innovativer multi-zentrischer Programme. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 101–122.
- Haubrich, K./Lüders, C./Struhkamp, G. (2007): Wirksamkeit, Nützlichkeit, Nachhaltigkeit. Was Evaluationen von Modellprogrammen realistisch leisten können. In: Schröder, U./Streblov, C. (Hrsg.): Evaluation konkret. Fremd- und Selbstevaluationsansätze anhand von Beispielen aus Jugendarbeit und Schule. Opladen und Farmington Hills, MI: Barbara Budrich, S. 183–201.
- Johansson, S./Junk, J./Liebich, J./Walkenhorst, D. (2022): Klientenzentrierte Evaluation in Multi-Agency-Settings der Extremismusprävention. Möglichkeiten und Grenzen eines Wirkungsorientierten Vorgehens. PRIF Report 6/2022. Frankfurt am Main. DOI: 10.48809/prifrep2206.
- Jütz, M./Kolke, S./König, F./Stärck, A./Zierold, D./Greuel, F./Roscher, T./Milbradt, B. (2023): Halbzeitbilanz zum Bundesprogramm „Demokratie leben!“ (2020–2024). Zwischenbericht der Gesamtevaluation. Halle (Saale): DJI.
- Lemire, S./Kwako, A./Nielsen, S. B./Christie, C. A./Donaldson, S. I./Leeuw, F. I. (2020): What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations. In: New Directions for Evaluation, H. 167, S. 73–86.
- Lüders, C. (2003): Jugendhilfeforschung zwischen Wissenschaft, Politik und Praxis. In: DJI-Bulletin, H. 64, S. 4–5.
- Lüders, C. (2024): Entwurf eines Vorworts für den Sammelband „Was wirkt wie und warum?“ Unveröffentlicht.
- Mayrhofer, H. (Hrsg.) (2017): Wirkungsevaluation mobiler Jugendarbeit. Methodische Zugänge und empirische Ergebnisse. Opladen, Berlin und Toronto: Barbara Budrich.
- Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.) (2021): Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa.
- Möller, K. (2003): Aktuelle politische Programme und Aktivitäten der pädagogischen sozialarbeiterischen Bekämpfung von Rechtsextremismus und Fremdenfeindlichkeit – eine kritische Zwischenbilanz. In: Lynen von Berg, H./Roth, R. (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen. Opladen: Leske + Budrich, S. 27–50.

- Nehlsen, I./Biene, J./Coester, M./Greuel, F./Milbradt, B./Armborst, A. (2020): Evident and Effective? The Challenges, Potentials and Limitations of Evaluation Research on Preventing Violent Extremism. In: *International Journal of Conflict and Violence*, 14. Jg., H. 2, S. 1–20.
- Ottmann, S./König, J. (Hrsg.) (2023a): Wirkungen im Blick – Wirkungsorientierung in der Sozialen Arbeit und Sozialwirtschaft. Tagungsdokumentation der Transferkonferenz. Nürnberger Hochschulschriften Nr. 53. Nürnberg: Evangelische Hochschule.
- Ottmann, S./König, J. (2023b): Wirkungsorientierung in der Sozialen Arbeit. Eine Einführung für Studium und Praxis. Stuttgart: Kohlhammer.
- Pawson, R./Tilley, N. (1997): *Realistic evaluation*. Thousand Oaks, CA: Sage.
- Programmevaluation DJI (2021): Entwicklung der Bundesprogramme seit 2001. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.) (2021): *Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen*. Weinheim und Basel: Beltz Juventa, S. 352.
- Rosa, H. (2016): *Resonanz. Eine Soziologie der Weltbeziehung*. Berlin: Suhrkamp.
- Uhl, K./Ulrich, S./Wenzel, F. M. (Hrsg.) (2004): *Evaluation politischer Bildung. Ist Wirkung messbar?* Gütersloh: Verlag Bertelsmann Stiftung.
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./Lobermeier, O./Heitmeyer, W. (Hrsg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden: Springer VS, S. 41–68.
- Wilson-Grau, R. (2018): *Outcome Harvesting: Principles, Steps, and Evaluation Applications*. Charlotte, NC: Information Age Publishing Inc.
- Worthen, B. R./Schmitz, C. C. (1997): Conceptual Challenges Confronting Cluster Evaluation. In: *Evaluation*, H. 3, S. 300–319.
- Zick, A./Küpper, B./Mokros, N. (2023): Die distanzierte Mitte. Rechtsextreme und demokratiegefährdende Einstellungen in Deutschland 2022/23. Hrsg. für die Friedrich-Ebert-Stiftung v. Franziska Schröter. Bonn: Verlag J. H. W. Dietz Nachf.

II Wirkungen und Gegenstandsangemessenheit

2 Wie ernst ist es uns wirklich mit der Wirkung?

Haupt- und Nebenfolgen eines Begriffs mit hoher Anmutungsqualität

Jan Ulrich Hense

2.1 Das Versprechen des Wirkungsbegriffs

Seit vielen Jahren ist in Bezug auf Qualitätsfragen allgemein und die Evaluation im Speziellen zu beobachten, dass zentrale Begrifflichkeiten große, oft wellenhaft verlaufende Beliebtheit außerhalb des Fachdiskurses erlangen und dabei proportional zu ihrer Verbreitung zunehmend leichtfertig verwendet und zweckentfremdet werden. Begriffe wie „evaluieren“ oder „Wirkungsanalyse“ sowie viele im Wortstamm verwandte Begriffe oder damit gebildete Komposita sickern in Alltagsdiskurse ein und unterliegen dort Deutungsmustern, die nicht immer mit denen des Fachdiskurses konvergieren.

Zu erklären ist dies vermutlich unter anderem mit einer hohen Anmutungsqualität und Attraktivität dieser qualitätsbezogenen Begrifflichkeiten. Sie sind einerseits semantisch so eingängig, dass ihre Bedeutung auf der Hand zu liegen und keine vertiefte konzeptionelle Auseinandersetzung zu erfordern scheint. Andererseits sind sie symbolisch stark aufgeladen und erfüllen oft allein Kraft ihres inhärenten Qualitätsversprechens die Funktion, Qualitätsbewusstsein und -fürsorge sowie Attribute wie Objektivität oder Zuverlässigkeit nach innen und außen zu repräsentieren.

Es scheint, als habe sich der Wirkungsbegriff, ebenfalls gemeinsam mit verschiedenen Komposita und Lehnwörtern, in den vergangenen Jahren verstärkt in diese Tradition eingereiht. Teils ist dies inhaltlich zu erklären, wenn sich darin etwa ein im Bildungssektor häufig konstaterter Wandel von der Input- zur vermehrten Output- bzw. Outcome-Steuerung ausdrückt und in diesem Zuge zunehmend Wirkungen thematisiert werden. Teils drängt sich aber auch der Eindruck von Moden auf. Im Hochschulsektor etwa wirkt der Evaluationsbegriff infolge von vielerorts wenig durchdachten Lehrevaluationsverfahren und einer gewissen Trivialisierung des Begriffs, wenn etwa jede Form der Bewertung als „Evaluation“ geadelt wird, zunehmend verbrannt. Als Konsequenz ist dort neuerdings mancherorts von Wirkungsanalysen die Rede, wo man noch vor wenigen Jahren richtigerweise wohl eher von Evaluationen gesprochen hätte.

Problematisch ist ein solcher fachlich wenig fundierter Umgang mit Begrifflichkeiten immer spätestens dann, wenn deutlich wird, dass es hochgradig fahrlässig sein kann, die Bedeutung scheinbar leicht eingängiger und erschließbarer Konzepte, zum Beispiel „Selbstevaluation“, für gesetzt zu halten und in der Folge aneinander vorbei zu reden, weil es eben nicht selbsterklärend ist, wer eigentlich das „Selbst“ in der Selbstevaluation und was genau die Rolle dieses „Selbst“ in der Evaluation ist (Hense 2006). Ebenso kann es zu Problemen führen, wenn symbolisch hochgradig aufgeladene Versprechungen, wie sie etwa mit dem Wirkungsbegriff verbunden sind, in der Praxis erfüllt werden sollen.

Denn der Wirkungsbegriff, um den es in diesem Beitrag zentral gehen soll, ist begrifflich mit einem gewissen Ballast verbunden, der nicht immer offensichtlich auf der Hand liegt. Vielleicht wird er gelegentlich auch gerne dann semantisch als Versprechen mitgeführt, wenn er nicht wirklich eingelöst werden kann. Sehen wir uns diesen Ballast im folgenden Abschnitt näher an, bevor wir in den folgenden Abschnitten diskutieren, wie in der Evaluation mit diesem Ballast umgegangen werden kann (Abschnitte 2.3 und 2.4) und welche Implikationen daraus für das Feld der Arbeit für Demokratie und gegen Extremismus (im Folgenden verkürzt: Demokratieförderung und Extremismusprävention) resultieren (Abschnitte 2.5 und 2.6).

2.2 Ein Begriff mit gern vergessenem Ballast?

Das zuvor angesprochene implizite Versprechen des Wirkungsbegriffs geht darauf zurück, dass „Wirkung“ nur die eine Hälfte einer untrennbaren Dyade und nur ein Drittel einer untrennbaren begrifflichen Triade ist. Eine Dyade bildet jede Wirkung mit der sie hervorbringenden Ursache. Egal, ob diese bekannt oder unbekannt ist, es macht keinen Sinn von einer Wirkung zu sprechen, wenn man nicht etwas sie Hervorbringendes mitdenkt – und sei es nur als noch zu ergründende Unbekannte. Egal, ob bekannt, vermutet oder unbekannt, jede Wirkung hat eine Ursache. Dabei ist zunächst einmal nachrangig, ob man diese Ursache singular ansieht, davon ausgeht, dass Wirkungen immer von mehreren Ursachen gemeinsam hervorgebracht werden oder dass es komplexere Bedingungsgefüge sind, aus denen Wirkungen hervorgehen, eine Unterscheidung, auf die wir noch zurückkommen werden.

Denn unabhängig von dieser Unterscheidung können wir direkt zur Triade überleiten. Sie resultiert daraus, dass die Natur dieser Beziehung zwischen Ursache und Wirkung zumindest in der westlichen, aristotelisch geprägten Denktradition (Foerster / Ollrogge 1993) immer eine Kausalbeziehung ist. Wie gewohnt prägnant von Michael Scriven, dem kürzlich verstorbenen Altmeister der Evaluation, auf den Punkt gebracht: „Causation: the relation between Mosquitoes and mosquito bites“ (Scriven 1991, S. 77). Wirkung und Ursache bilden also eine Dya-

de, die von einer unterstellten Kausalität verbunden werden, die sie zur Triade ergänzen.

Die genaue Natur dieser Verbindung ist schon lange Gegenstand umfassender epistemologischer und ontologischer Auseinandersetzungen, die in starkem Maße mit methodologischen Überlegungen zur Untersuchung von Kausalzusammenhängen verschränkt sind (Reichardt 2022; Gates / Dyson 2017; Pearl 2013; Befani 2012). Dabei werden verschiedene Kausalitätskonzepte unterschieden, die sich – wie oben bereits angedeutet – unter anderem darin unterscheiden, wie in einem Wirkungszusammenhang die Ursache verstanden wird. Reichardt (2022) differenziert monokausale und linear-sukzessionistische Auffassungen, die nur eine Ursache berücksichtigen (*regularity approaches*), die Annahme von Kausalfeldern, in denen Bedingungskonfigurationen eine bestimmte Wirkung auslösen (*configurational approaches*), bis hin zu einem Verständnis, bei dem Ursachen durch das Wirken kontextabhängiger Mechanismen erklärt werden (*generative approaches*).

Unabhängig von diesen Überlegungen ist aus empirischer Sicht ein zentrales Grundproblem der Wirkungsevaluation die Tatsache, dass immer nur eine jeweils bestehende Ursache beobachtet werden kann. Egal, ob man die dabei untersuchte Maßnahme (das evaluierte Programm) als monokausale Ursache, als Bedingungskonfiguration oder als Auslöser von Mechanismen, die in einem konkreten Kontext wirksam sind, betrachtet, es gibt immer nur eine Welt mit oder eine Welt ohne die Maßnahme. Die Frage, welchen Unterschied eine Maßnahme bewirkt hat, kann also nie direkt beobachtet werden, da der sogenannte kontrafaktische Zustand, wie sich also die Welt ohne Vorhandensein des Programms entwickelt hätte, nie gleichzeitig existieren und zum direkten Vergleich herangezogen werden kann. Obwohl dieser Gedanke des kontrafaktischen Zustands dem experimentellen Paradigma entstammt, das dem sukzessionistischen Kausalitätsverständnis (*regularity approach*) zuzuordnen ist, lässt er sich analog auch für andere Auffassungen anwenden (Reichardt 2022). Das Grundproblem besteht darin, dass der kausal bewirkte Effekt der Maßnahme, landläufig „Wirkung“, exakt in der Differenz zwischen dem faktischen Zustand nach umgesetzter Maßnahme und dem prinzipiell unbekanntem kontrafaktischen Zustand ohne die Maßnahme besteht, was das Dilemma jeder Analyse von Wirkungen, das bewusst groß geschriebene „Fundamental Problem of Causal Inference“ (Holland 1986, S. 947), auf den Punkt bringt.

Besonders im Kontext wirkungsorientierter Evaluation ist es immer lohnend, sich mit dieser Grundproblematik näher zu befassen (z. B. Wolbring 2014). Auch kann man noch einen weiteren Schritt zurücktreten und meist selbstverständlich vorausgesetzte Prämissen vor dem Hintergrund ihrer ideengeschichtlichen Genese im westlichen Denken und ihrer kulturellen Einbettung reflektieren (LaFrance / Nichols / Kirkhart 2012). Unabhängig von den Schlüssen über Kausalität, die aus diesen Diskussionen zu ziehen sind, können wir uns nicht darüber

hinwegsetzen, dass die in Evaluationen beobachtete Gegenstandsebene im Regelfall einem instrumentellen Wirkungsverständnis im Sinne der oben skizzierten Triade folgt. Denn den Gegenständen, die wir in der Programmevaluation typischerweise untersuchen, liegt letztlich immer eine zweckrationale Logik zugrunde, und damit ist ihnen von Natur aus eine bestimmte Auffassung von Kausalität eingeschrieben (Shadish/Cook/Leviton 1991).

Schließlich werden Projekte und Programme in der Vielfalts- und Demokratieförderung und Extremismusprävention ebenso wie in anderen Politikbereichen nicht einfach so in die Welt gesetzt, sondern immer mit Absichten und Zielen, mögen diese auch noch so vage benannt oder zwischen verschiedenen Beteiligtegruppen umstritten sein. Projekten und Programmen liegt immer eine Wirkungsannahme zugrunde, auch wenn diese in vielen Fällen nicht weiter elaboriert sein mag als zum Beispiel: „die Maßnahme wird einen ursächlichen Beitrag zu einer als erstrebenswert erachteten Veränderung oder Stabilisierung leisten“. Wie sonst auch sollten sich öffentlich oder anders finanzierte Maßnahmen begründen lassen, wenn nicht die Annahme damit verknüpft ist, bestimmte Problemlagen lindern oder verhindern zu können? Die Triade aus Ursache (Programm), Wirkung (Zielerreichung, hier meist unter Vernachlässigung möglicher anderer Wirkungen) und angenommener Kausalität (Wirkungserwartung) ist hier also immer sinnstiftend vorausgesetzt, weswegen sie auch vonseiten der Evaluation naturgemäß in ihrem Blick auf diese Gegenstände vorausgesetzt wird.

Die Frage ist allerdings, um wieder zum Aspekt der Popularität von Begrifflichkeiten des Qualitätsdiskurses zurückzukommen, ob diese triadische Natur des Wirkungsbegriffs immer präsent und ernsthaft mitgemeint ist, wenn von Wirkung und ihren Komposita wie Wirkungsanalyse, Wirkungsmessung, Wirkungsevaluation etc. die Rede ist. Wie gezeigt: Wer Wirkung ernst nimmt, erhebt mit der Verwendung dieser Begriffe sachlogisch den Anspruch, nicht nur irgendwie irgendetwas im Zielbereich einer Maßnahme zu erfassen oder eine Tabellenvorlage auszufüllen und dies als „Wirkungsmodell“ zu deklarieren (siehe Abschnitt 2.4.1), sondern den Anspruch der Kausalattribution, die einen empirisch beobachteten Unterschied kausal der untersuchten Maßnahme zuschreibt.

Das Problem verschärfend kommt hinzu, dass Wirkungen keinesfalls auf intendierte Wirkungen im Sinne erreichter Programmziele reduziert werden können, sondern begrifflich jede Form von kausal verursachter Veränderung oder Stabilisierung umfassen, zum Beispiel auch nicht-intendierte und nicht-antizipierte Wirkungen (Widmer 2012). Auch das ist ein Fakt, der in der Wirkungsdiskussion oft nicht umfänglich berücksichtigt wird, ebenso wie die Tatsache, dass die eigentlich interessierenden Wirkungen meist erst mit zeitlicher Verzögerung auftreten. Bildlich werden diese als „Downstream“-Wirkungen bezeichnet, bei denen der „Fluss“ der Programmwirkungen durch viele externe

Einflüsse nach Programmende „verwässert“ wird (Earl/Carden/Smutylo 2001). Je näher man also den letztendlich intendierten Wirkungen eines Programms kommen will, desto schwerer ist es, diese von anderen Einflüssen zu separieren, ein Problem, das sich als „Rigorous Impact Evaluation Paradox“ (Widmer 2012, S. 45) bezeichnen lässt.

Es drängt sich also insgesamt der leise Verdacht auf, dass man das erkenntnistheoretische Versprechen des Wirkungsbegriffs oft etwas leichtfertig implizit mitführt und daher, etwa im Kontext von Antragsrhetorik oder beim Erstellen von Evaluationskonzepten, im Bereich der Versprechungen, Ankündigungen oder symbolisch-politischen Kommunikation verbleibt. Was von diesen Versprechungen aber theoretisch und empirisch tatsächlich eingehalten wird bzw. überhaupt eingehalten werden kann, steht denn vielleicht auf einem anderen Blatt.

Daraus resultiert natürlich die Frage, inwiefern und wie dieses erkenntnistheoretische Versprechen überhaupt eingelöst werden kann. In der sozialwissenschaftlichen Methodenlehre gibt es zwar schon seit Längerem nur noch in Nischen den Versuch, bestimmte Methoden als einzig validen Weg zur Erkenntnis zu verfechten. Der Streit zwischen qualitativen und quantitativen Methoden ist zwar nicht restlos und spurlos überwunden, wird aber glücklicherweise kaum noch in Reinform betrieben. Dennoch erhebt in der „klassischen“ sozialwissenschaftlichen Methodenlehre nach wie vor allein ein Forschungsdesign den Anspruch, das Problem des Kontrafaktischen im Sinne eines sicheren Nachweises von Kausalwirkungen lösen zu können, das randomisierte Kontrollgruppenexperiment (engl. randomized controlled trial – RCT), der selbstbewusst sogenannte „gold standard“ oder „Cadillac“ der „rigorosen“ Wirkungsforschung. Indem dabei Zielgruppenmitglieder rein zufällig einer Kontrollgruppe, die nicht am Programm teilnimmt, und einer teilnehmenden Versuchsgruppe zugelost und Wirkungsindikatoren bei beiden Gruppen vor Beginn und nach Ende der Maßnahme gemessen werden, wird eine Konstellation geschaffen, bei der gemessene Unterschiede in der Veränderung der Zielvariablen zwischen den beiden Gruppen eindeutig der Teilnahme am Programm zugeschrieben werden können (für eine praxisnahe Einführung in Evaluationskontexten vgl. Caspari 2012).

Es wundert daher nicht, dass das experimentelle Forschungsparadigma und vor allem seine quasi-experimentellen Näherungen (Shadish/Cook/Campbell 2013; Caspari 2012; Cook/Pohl/Steiner 2011; Shadish/Cook 2009; Cook/Shadish/Wong 2008; Cook/Campbell 1979) lange Zeit die Wirkungsevaluation als Leitbild dominierten. Ähnlich lange gibt es allerdings auch Kritik an diesem Paradigma, unter anderem bezüglich seiner erkenntnistheoretischen Prämissen, der Reichweite und dem Informationsgehalt seiner Ergebnisse, der Ausmittelung individueller Unterschiedlichkeiten und nicht zuletzt an seiner praktischen Anwendbarkeit für Evaluationszwecke. Für die Evaluation besonders fruchtbar war dabei eine Perspektive, an die uns der folgende Abschnitt heranzführt.

2.3 Grenzen experimenteller Ansätze der Wirkungsevaluation

Über Vor- und Nachteile, Grenzen und Potenziale experimentell angelegter Wirkungsevaluationen gegenüber anderen methodischen Ansätzen wurde in der Literatur bereits ausführlich diskutiert. Wichtige Debatten der Evaluationsgeschichte haben sich an dieser Konfliktlinie entzündet, so etwa die zwischen Lee Cronbach und Donald Campbell zur Frage, ob in Evaluationen die im (quasi-)experimentellen Paradigma besonders betonte interne Validität (Cook/Campbell 1979) oder die dabei oft vernachlässigte externe Validität von Befunden (Cronbach 1982, 1980) von größerer Relevanz sei und welche methodischen Konsequenzen aus dem jeweiligen Primat resultieren (Albright/Malloy 2000). Eine jüngere Iteration dieser Auseinandersetzung erfolgte im Kontext der Diskussion um die *evidence based policy*. Aus ihr wird hier nur ein Aspekt herausgegriffen, der für die weiteren Überlegungen zentral ist und an einem plakativen Beispiel veranschaulicht werden soll.

Der Gedanke der evidenzbasierten Politik, der Anspruch also, Maßnahmen kausalattributionierend auf ihre letztendlichen Wirkungen hin zu untersuchen und politische und administrative Entscheidungen auf Basis solchermaßen erzeugter wissenschaftlicher Evidenz, am besten akkumuliert in Form von Forschungssynthesen oder Meta-Analysen über mehrere Studien, zu gründen, stammt ursprünglich aus dem Feld der Medizin (Hansen/Rieper 2009). In der Auseinandersetzung um die *evidence based medicine* haben Smith und Pell einen äußerst pointierten Beitrag geleistet, der hier kurz skizziert werden soll, da sich aus ihm wichtige Lehren – auch über die Medizin hinaus – ziehen lassen (Smith/Pell 2003).

Äußerlich dem gängigen Publikationsformat eines systematischen Reviews folgend, fassen Smith und Pell die Befundlage randomisierter Kontrollgruppenversuche zur Wirksamkeit von Fallschirmen zur Vermeidung von Tod und Verletzungen nach schwerkraftbedingten Abstürzen aus großer Höhe zusammen. Wenig überraschend müssen sie feststellen, dass keine rigorosen Studien vorliegen, in denen zufallsbedingt Versuchspersonen vor dem Absprung aus einem Flugzeug entweder mit einem Fallschirm ausgestattet wurden oder nicht. Gleichzeitig zeigen sie auf empirischer Basis, dass einerseits nicht jeder Absturz aus großer Höhe mit Tod oder Verletzung endet und andererseits Fallschirme nicht in jedem Fall zuverlässig davor schützen. Sie kommen daher zum (aus Perspektive der *evidence based medicine*) vernichtenden Urteil, dass der Gebrauch des Fallschirms als gesundheitsbewahrende Maßnahme allein auf Beobachtungsdaten basiert, also nur mit einer empirisch äußerst schwachen Begründung erfolgt. Sie schließen mit der sehr britischen, sarkastischen Empfehlung zur Verbesserung der Studienlage, „that those who advocate evidence based medicine and criticise use of interventions that lack an evidence base will not hesitate to demonstrate their commitment by volunteering for a double blind, randomised, placebo controlled, crossover trial“ (Smith/Pell 2003, S. 1460).

Die durchweg süffig zu lesende Abrechnung von Smith und Pell mit dem exzessiven Anspruch, allein und ausschließlich „rigoros“ wirkungsevaluierte Maßnahmen als Grundlage informierter Entscheidungen gelten zu lassen, deutet die praktisch erforderliche Alternative nur cursorisch an. Sie besteht ihnen zufolge darin, den gesunden Menschenverstand (*common sense*) bei der Beurteilung von Nutzen und Risiken einer Intervention anzuwenden. Tatsächlich lässt sich aber mehr aus ihrem plakativen Beispiel lernen, wenn wir danach fragen, *warum* wir denn trotz fehlender experimenteller Evidenz an die Wirksamkeit von Fallschirmen glauben, also lieber mit als ohne Fallschirm aus einem Flugzeug springen würden. Betrachten wir dazu drei zentrale Kriterien, die für den empirischen Nachweis von Kausalschlüssen aus Sicht eines sukzessionistischen Kausalitätsverständnisses gegeben sein müssen und im rigorosen RCT erfüllt sind (Shadish/Cook/Campbell 2002; Cook/Campbell 1979): Wirkung und vermutete Ursache treten erstens gemeinsam auf oder bleiben gemeinsam aus (Kovariation), die angenommene Ursache liegt zweitens zeitlich vor der Wirkung und drittens sind andere Kausalfaktoren abwesend.

Bei der Beobachtung des Einzelfalls können wir bei einem funktionierenden Fallschirm alle Kriterien mit hinreichender Sicherheit als erfüllt ansehen. Wir beobachten das Herabschweben und Landen mit Fallschirm bei gleichzeitiger Verletzungsfreiheit, wir sehen, dass erst der Fallschirm aufgeht und dann die sichere Landung erfolgt und wir können keine plausible Alternativerklärung – wie etwa ein plötzliches Aussetzen der Schwerkraft – beobachten. Interessanter wird es allerdings bei der Frage, warum wir jenseits der Beobachtung des Einzelfalls, ja sogar ohne jede eigene Augenscheinevidenz die *generelle* Überzeugung herausbilden, dass Fallschirme Leben retten. Warum sind wir der festen Meinung, dass Fallschirme unter der Annahme ihrer Funktionstüchtigkeit und sachgemäßen Verwendung unter gleichen Bedingungen *immer* wirken werden?

Die sehr einfache Erklärung liegt in diesem plakativen Fall darin, dass wir exakt verstehen und erklären können, *warum* der Fallschirm wirkt. Wir kennen sehr genau die physikalischen Gesetzmäßigkeiten von Schwerkraft, Luftwiderstand und Aerodynamik, die hier am Werk sind, und können daher auch zum Beispiel sehr genau bestimmen, welche Größe der Fallschirm für ein bestimmtes Gewicht bei einer gegebenen Luftdichte haben muss, welche Form besonders vorteilhaft ist und unter welchen Bedingungen er versagen wird. Mit anderen Worten: Wir haben eine plausible *Theorie* von der Wirkungsweise des Fallschirms, die wir für so gut halten, dass wir ihr auch in Abwesenheit experimenteller Evidenz vertrauen.

Es stellt sich damit die Frage, ob und unter welchen Bedingungen Theorie auch in anderen Konstellationen dabei helfen kann, Wirkungsanalysen zu stützen, und als Alternative zu RCTs dienen kann. Dies ist nicht zuletzt deswegen relevant, weil RCTs in komplexen Programmen mit einer Vielzahl von unterschiedlichen Aktivitäten, Zielgruppen und Zielen auf verschiedenen Ebenen, wie sie auch für die Felder der Demokratieförderung und Extremismusprävention typisch sind,

oft praktisch gar nicht oder nur für kleine Ausschnitte zu realisieren sind. Der Blick wendet sich daher einer Klasse von Evaluationsansätzen zu, in denen eine wie auch immer konkret geartete und dargestellte Theorie des Evaluationsgegenstands zentrales Werkzeug der Wirkungsanalyse ist.

2.4 Programmtheoriegestützte Ansätze in der Evaluation

Schon allein wegen der Notwendigkeit, Zielgruppenmitglieder randomisiert einer Maßnahme zuzuteilen bzw. sie explizit von dieser auszuschließen, sind experimentelle Designs sehr voraussetzungsreich. Zwar gibt es einige geschickte Lösungsansätze, zum Beispiel Wartegruppendesigns, auch diese stoßen aber in vielen Programmevaluationskontexten auf Grenzen der praktischen Anwendbarkeit. Auch dies hat schon seit langem die Frage aufkommen lassen, welche Näherung an die Attributionsproblematik durch andere Forschungsdesigns zu erreichen ist. Innerhalb des experimentellen Paradigmas sind dabei die quasi-experimentellen Designs zu nennen, denen unter bestimmten Bedingungen ähnlich sichere Kausalschlüsse wie den RCTs zugetraut werden (Cook/Pohl/Steiner 2011; Cook/Shadish/Wong 2008; Shadish/Cook/Campbell 2002). Aber auch eine Reihe anderer nicht-experimenteller Ansätze zur Ermittlung von Programmeffekten und Untersuchung von Kausalzusammenhängen wird in der Evaluationsliteratur diskutiert. Reichardt (2022) nennt die folgenden sieben Ansätze (vgl. auch Davidson 2005, 2000), die alle zumindest implizit das Problem des unbekanntes kontrafaktischen Zustands adressieren:

1. Vergleich mit einer nicht-teilnehmenden Kontrollgruppe (experimentelles Paradigma);
2. Vorher-Nachher-Vergleiche ohne Kontrollgruppe;
3. subjektive Schätzung des kontrafaktischen Zustands, der ohne das Programm vorliegen würde, durch Zielgruppenmitglieder oder Sachverständige („what if“-Fragen);
4. direkte Einschätzung von Programmeffekten durch Stakeholder in Vorwärts-Richtung, zum Beispiel „Was hat das Programm bei Ihnen bewirkt?“, oder Rückwärts-Richtung, zum Beispiel „Woran lag es, wenn sich etwas in der Programmlaufzeit verändert hat?“ („just-tell-me“-Fragen);
5. direkte Beobachtung von Ursache und Veränderung und Inferenzschluss, wie sich die Situation ohne Ursache entwickelt hätte;
6. theoriebasierte Ansätze, in denen Wirkungen entlang von Kausalketten und komplexeren Kausalzusammenhängen nachgespürt wird;
7. Modus Operandi (nach Scriven): Ausgehend von den Programminhalten werden spezifische „verräterische Spuren“ identifiziert, an denen Arbeit und Wirkungen des Programms im Feld erkennbar sein sollten.

Selbstverständlich sind diese Ansätze hinsichtlich der Belastbarkeit der gezogenen Kausalschlüsse nicht äquivalent. Aus der klassischen evidenzbasierten Perspektive schlägt sich dies im Begriff der *hierarchies of evidence* nieder, bei denen RCTs bzw. auf ihnen basierende Metaanalysen an der Spitze gesehen werden (Hansen/Rieper 2009). Solange man ihre jeweiligen Stärken und Schwächen im Blick behält, erhalten sie ihre Berechtigung aber eben einerseits dadurch, dass experimentelle Ansätze oft praktisch nicht realisierbar sind, was sich auch in der Forderung nach (best) *credible evidence* (Donaldson 2022) ausdrückt, und andererseits, indem sie oft eine größere Informationsfülle bereitstellen, als es experimentelle Ansätze vermögen. Denn ein randomisiertes Kontrollgruppenexperiment stellt ja trotz allen Aufwands zu ihrer Realisierung letztlich „nur“ einen statistischen Signifikanzwert und gegebenenfalls ein Effektstärkemaß, im Extremfall zu einer einzelnen abhängigen Variablen, als Ergebnis zur Verfügung. Sie haben noch dazu nur unter der Annahme von relativ engen Voraussetzungen Gültigkeit, die oft nicht erfüllt sind (Lipsey 1990) oder deren Übertragbarkeit auf andere Kontexte fraglich ist (Cronbach 1982; Pawson/Tilley 1997).

Unter den von Reichardt (2022) oben aufgezählten Alternativen zum experimentellen Paradigma hat im Evaluationsdiskurs der vergangenen Jahrzehnte allerdings nur einen substanziellen Niederschlag gefunden. Es handelt sich dabei um die theoriebasierten Evaluationsansätze, die ideengeschichtlich schon sehr früh in der modernen Evaluationsgeschichte angelegt waren und bereits in den 1970er Jahren konzeptionell fundiert wurden, aber erst ab den 1990er Jahren zunehmend das Evaluationsfeld beeinflussten und zum Teil des „Mainstreams“ wurden. Wie auch am humorigen Bonmot „What to do until the random assigner comes“ von Carol Weiss (2002, S. 198) zu erkennen, wurden diese in erster Linie als Antwort auf die Umsetzungsschwierigkeiten experimenteller Ansätze in der Praxis entwickelt. Vor allem mit ihrem zentralen Instrument, dem Wirkungsmodell, haben sie schnell auch breitere Anwendung gefunden, da es eine ganze Reihe von nützlichen Funktionen im Evaluationsprozess erfüllen kann (Hense/Taut 2021). Im Folgenden sehen wir uns daher vier Ansätze genauer an, die dem theoriebasierten Paradigma zuzuordnen sind und deren vermehrte Anwendung in der Evaluation von Programmen und Maßnahmen der Demokratieförderung und Extremismusprävention grundsätzlich vielversprechend erscheinen.

2.4.1 „Klassische“ theoriebasierte Evaluationsansätze

Als frühe Urheber theoriebasierter Evaluationsansätze werden oft Rossi und vor allem Chen genannt, der die erste Monographie zum Thema theoriegesteuerter Evaluation vorgelegt hat (Chen/Rossi 1992; Chen 1990; Chen/Rossi 1983). Maßgebliche Grundlagen und der Begriff selbst gehen aber auf Carol Weiss (1998) zu-

rück, ideengeschichtlich lässt sich das Konzept sogar bis in die Anfänge der modernen Evaluationsgeschichte zurückverfolgen (Gargani 2003).

Leitgedanke des Konzepts ist, der Evaluation von Programmen, also Maßnahmen aller Art, immer eine „Programmtheorie“ zugrunde zu legen, die die wesentlichen Annahmen zur Wirkungsweise enthält. Visualisiert werden diese gewöhnlich in Form eines Wirkungsmodells, das aufzeigt, wie und warum die vom Programm in Gang gesetzten Aktivitäten zu seinen Wirkungen führen. Der Theoriebegriff wird dabei eher alltagssprachlich verstanden. Es können zwar sozialwissenschaftliche Theorien herangezogen werden, ebenso gut kann es sich aber um eine rein aus der Praxis heraus geborene, „naive“ und feldnahe Theorie oder um Mischformen daraus handeln. Trotz verschiedener kritischer evaluationstheoretischer Stimmen (z. B. Stufflebeam/ Coryn 2014) und einer gewissen Diskrepanz zwischen dem eigentlichen Konzept und seiner Umsetzung in der Praxis (Coryn et al. 2011) war die verstärkte Aufmerksamkeit auf die theoretische Fundierung von Interventionen, die übrigens auch im experimentellen Paradigma eine notwendige Voraussetzung für verlässliche Ergebnisse wäre (Lipsey 1990), eine der einflussreichsten Innovationen der jüngeren Evaluationsgeschichte (Hense/Taut 2021).

Innerhalb einer Programmtheorie werden üblicherweise zwei Bereiche unterschieden. Die Theorie der Veränderung (*theory of change*) beschreibt, wie die im Programm umgesetzten Aktivitäten zu Veränderungen bei Zielgruppen führen bzw. zu ihnen beitragen. Je nach Elaborationsgrad erfolgt das im einfachsten Fall, im sogenannten Logischen Modell, als rein lineare Abfolge von Aktivitäten, Outputs (den aus den Aktivitäten resultierenden zählbaren Produkten und Leistungen), Outcomes (Veränderungen bei Zielgruppen) und Impacts (übergeordnete Wirkungen, z. B. auf gesellschaftlicher Ebene), die letztlich mithilfe einer einfachen Tabellenvorlage erstellt werden können. Üblicherweise wird aber unter Programmtheorie ein komplexeres Wirkungsgefüge erwartet, das nach Aktivitäten und Outcomes differenzierte, gegebenenfalls interagierende Wirkungspfade unter Einbezug mindestens von Mediatoren, seltener auch von Moderatoren (Donaldson 2001; Baron/Kenny 1986) enthält.

Da aber in Programmen nicht davon ausgegangen werden kann, dass Aktivitäten automatisch so erfolgen wie geplant, ist der *theory of change* eine *theory of action* vorgeschaltet, die die Bedingungen modelliert, die erforderlich sind, damit jene Aktivitäten in der notwendigen Qualität und Quantität ausgelöst werden, die die *theory of change* in Gang setzen können (Funnell/Rogers 2011). Enthalten sind daher vor allem notwendige Inputs (Ressourcen und Ausgangsbedingungen) sowie gegebenenfalls Kontextbedingungen, auf die das Programm zum Funktionieren angewiesen ist, und eine genaue Beschreibung seiner Aktivitäten, die durch die Inputs ermöglicht werden.

Verschiedene programmtheoriebasierte Ansätze bzw. die sie jeweils vertretenden Autor:innen sind sich darin einig, dass die Verwendung von Pro-

grammtheorie für die meisten Arten von Evaluation eine gute Sache ist. Weniger Eindeutigkeit besteht zu verschiedenen Fragen der praktischen Umsetzung, darunter etwa, wo im Kontinuum zwischen „echten“ sozialwissenschaftlichen Theorien und den eher „impliziten“ Theorien der Praxis Programmtheorien angesiedelt sein sollten oder wie sie konkret zu erarbeiten sind, falls noch nicht im Programm vorhanden. Auch Untersuchungsdesigns, Erhebungs- und Auswertungsmethoden sind im theoriebasierten Ansatz nicht per se festgelegt, auch wenn er oft eher mit einer quantitativen Methodik assoziiert ist. Von daher ist es – im Gegensatz zu den im Folgenden dargestellten Ansätzen – auch nicht ganz angemessen, „klassische“ Ansätze als etwas Einheitliches abzuhandeln. Wir können es hier aber dabei belassen, dass das Verbindende ist, dass eine Gegenstandstheorie eine wichtige Rolle in Planung und Durchführung der Evaluation spielt, dies sich aber in unterschiedlichen Realisierungsformen niederschlagen kann. Will man sich der Attributionsproblematik und der Wirkungsweise des Programms empirisch nähern, impliziert das allerdings mindestens, dass wesentliche angenommene Wirkungspfade des Programms operationalisiert und durch geeignete, meist quantitative Analysemethoden auf die angenommenen Kausalzusammenhänge hin überprüft werden (Funnell/Rogers 2011; Weiss 1998; Chen 1990).

2.4.2 Realistic Evaluation

Auch für Pawson und Tilley (1997) war bei der Entwicklung des Ansatzes der Realistic Evaluation ein zentraler Gedanke, die Limitierungen experimenteller Untersuchungsdesigns zu überwinden und den Fokus vom Wirkungsnachweis auf das Verstehen von Wirkungen zu lenken. Neben dem theoriebasierten Evaluationsdiskurs beziehen sie sich auf einen, dem philosophischen Realismus entlehnten, erkenntnistheoretischen Zugang, der insbesondere das Kausalitätsverständnis betrifft. Im Gegensatz zum stark simplifizierenden sukzessionistischen Verständnis, das dem Experiment unterstellt wird, bei dem im Prinzip nur Endzustände nach unterschiedlichen Ausgangsszenarios verglichen werden, gehen sie von einem generativen Kausalitätsmodell aus. In diesem werden Wirkungen von Mechanismen hervorgebracht, die nicht universell wirksam, sondern in starkem Maß kontextabhängig sind, woraus sich ihre grundsätzliche Kritik des experimentellen Paradigmas für Evaluationszwecke speist: „Programs are not disembodied treatments. Programs are always embedded into existing flows of social conduct and produce effects because of their place in the stratified layers of social reality. Policy on – policy off comparisons are a chimera“ (Pawson/Tilley 2005, S. 365).

Statt eines klassischen Kausalschlusses in der Form „Das Programm wirkt/wirkt nicht“ sucht die Realistic Evaluation nach Aussagen in der Form „Das

Programm(-element) wirkt, wenn im Kontext X der Mechanismus Y zum Tragen kommt“, wobei grundsätzlich mehrere unterschiedliche Mechanismen eine bestimmte Wirkung hervorbringen können. Ihr oft zitiertes, von ihnen selbst so bezeichnetes „*realist mantra*“ lautet dementsprechend: „What works for whom in what circumstances?“ (Pawson/Tilley 1997, S. 144). Eines ihrer klassischen Beispiele ist die Überwachung von Parkplätzen mit Kameras zur Reduktion von Fahrzeugeinbrüchen und -diebstahl. Sie wird durch eine Reihe denkbarer verschiedener Mechanismen bewirkt, etwa die Abschreckung potenzieller Täter, das Fassen von Tätern durch eine Streife nach Entdecken der Taten durch Echtzeitüberwachung oder die nachträgliche Aufklärung aufgrund einer Aufzeichnung. Die Wirksamkeit dieser Mechanismen wird aber hier eben nicht als universell angenommen, sondern als von Kontextbedingungen begünstigt oder behindert. Darunter fallen im Beispiel etwa die Distanz des Parkplatzes zur nächsten Polizeiwache, die Qualität und Speicherdauer von Videoaufzeichnungen oder die Anzahl von Gelegenheits- vs. Intensivtätern.

Zentrales Ziel der Evaluation ist hier daher das Identifizieren sogenannter Context-Mechanism-Outcome-Pattern-Configurations (CMOCs), die modellhaft zusammenfassen, wie durch eine Maßnahme Mechanismen ausgelöst werden, die unter welchen Umständen bei welchen Zielgruppen oder anderen Entitäten bestimmte Änderungen hervorbringen. Sie gilt es zunächst theoretisch zu bestimmen und im Rahmen der Evaluation empirisch anhand des Datenmaterials zu überprüfen. Genau wie in der ideengeschichtlich eng verwandten nordamerikanischen Variante der Realist Evaluation (Henry 2005; Mark/Henry/Julnes 2000), bei der der so verstandene Kontext eine weniger deutlich ausgeprägte Rolle spielt, können dabei sowohl qualitative als auch quantitative Methoden zum Einsatz kommen. Ein spezifischer methodischer Zugang existiert nur insofern, dass die Datenerhebung vor allem bei Stakeholdern des Programms so gestaltet wird, dass die resultierenden Daten zur Überprüfung der zuvor aufgestellten Theorien in Form von CMO-Konfigurationen dienen können. Oft impliziert das, die Befragten direkt oder verklausuliert mit den jeweiligen Wirkungsannahmen zu konfrontieren und aus ihren Reaktionen Schlüsse über einzelne oder mehrere CMO-Konfigurationen zu ziehen (Pawson/Tilley 2005).

Die Autoren selbst verorten ihren Ansatz innerhalb des theoriebasierten Paradigmas. Aufgrund ihrer Begriffsverwendung und Betonung der CMO-Konfigurationen braucht es allerdings einen gewissen Transfer, damit es nicht zu Missverständnissen kommt. Besonders *context* meint hier etwas anderes als in der Sprache anderer Programmtheorien und insbesondere der Logischen Modelle. Allgemein und auch im Logischen Modell wird Kontext üblicherweise als „alle Arten von Einflüssen auf das Programm außerhalb des Programms selbst“ verstanden und dient oft auch als vager Platzhalter für nicht näher spezifizierte, unbekannte oder nicht vorhersehbare Einflüsse. Im Rahmen von CMO-Konfigurationen bezeichnet *context* dagegen spezifisch benannte Bedingungen, unter

denen konkrete Mechanismen mehr oder weniger stark zum Tragen kommen. Im Gegensatz zum Kontextverständnis des Logischen Modells und vielen anderen programmtheoretischen Ansätzen können sie auch innerhalb des Programms liegen, zum Beispiel in Variationen der Programmumsetzung oder in Merkmalen von Zielgruppenmitgliedern. In pfadanalytisch orientierten Wirkungsmodellen, in denen einzelne Wirkungsfaktoren mit spezifischen Einflusspfaden modelliert werden, entsprechen sie am ehesten Moderatorvariablen (Baron/Kenny 1986), welche Stärke und Einfluss einer Wirkungsbeziehung verändern können und dementsprechend als Variablen dargestellt werden, deren Einfluss nicht auf andere Variablen zielt, sondern auf den Wirkungszusammenhang zweier anderer Variablen (Henry 2005; Donaldson 2001).

Ein weiterer Punkt, der bei der Anwendung in Evaluationen, vor allem von umfangreicheren Programmen, nicht übersehen werden sollte, ist die Granularität des Ansatzes. Die drei konstituierenden Elemente einer CMO-Konfiguration sind als distinkte, empirisch klar zu operationalisierende Konstrukte zu denken. Das impliziert, dass ein größeres Programm und selbst seine Einzelprojekte oft nicht als einzelne CMO-Konfigurationen zu modellieren sind, sondern je nach Anzahl der enthaltenen Aktivitäten und angestrebter Outcomes mehrere, oft sogar sehr viele CMOCs umfassen können.

2.4.3 Contribution Analysis

Contribution Analysis ist ein weiterer Ansatz, der dem theoriebasierten Evaluationsparadigma zuzuordnen ist und maßgeblich auf John Mayne (2019, 2012) zurückgeht. Auch hier liegen mehr oder weniger philosophisch fundierte erkenntnistheoretische Überlegungen zu Kausalschlüssen zugrunde, mit denen begründet wird, dass nicht mehr nach der kausalen Verursachung von Wirkungen durch das Programm gefragt wird, sondern nach dem Beitrag (*contribution*), den das Programm hatte. Auch hier wird zunächst eine Programmtheorie aufgestellt, welche diesen a priori vom Programmdesign implizit oder explizit unterstellten Beitrag illustriert. Gegenüber anderen Ansätzen allerdings handelt es sich dabei um ein relativ lineares Wirkungsmodell, das von einer Wirkungskette ausgeht, die Programmaktivitäten und letztendliche Wirkungen über mehrere vermittelnde Variablen, im pfadanalytischen Jargon wären das wieder Mediatoren, verbindet. Zur Programmtheorie wird diese Wirkungskette, indem für jede einzelne Kausalverbindung innerhalb dieser Kette die kritischen Annahmen ergänzt werden, die erfüllt sein müssen, um sie zu ermöglichen, was letztlich ähnlich zu pfadanalytischen Moderatoren bzw. dem ermöglichenden *context* der Realistic Evaluation zu lesen ist.

Das innovative Merkmal des Ansatzes besteht darin, welche Rolle er bereits vorhandener Evidenz einräumt. Die Annahme ist, dass vor allem jene Bereiche

des Wirkungsmodells, für die vorab noch wenig Evidenz vorliegt, intensiver im Rahmen der Evaluation empirisch untersucht werden sollten. Dementsprechend folgt nach der Formulierung der Programmtheorie, anders als in anderen theoriebasierten Ansätzen, nicht direkt die empirische Arbeit im Feld, sondern eine Recherche am Schreibtisch. Für jede im Wirkungsmodell dargestellte Wirkungsannahme wird gefragt, inwiefern sich diese bereits durch vorab vorliegende Evidenz, zum Beispiel aus der Forschung oder aus vorhergehenden Evaluationen, stützen lässt. Als plakatives Beispiel kann ein Präventionsprogramm dienen, das Mortalität und Gesundheitsschäden infolge des Tabakrauchens reduzieren will. Hier ist der Zusammenhang zwischen Rauchen und seinen gesundheitlichen Folgen so gut untersucht, dass eine Evaluation sich darauf beschränken kann, die Reduktion des Konsums empirisch zu untersuchen und den Rest der Wirkungskette mit vorhandener Evidenz zu belegen. Aus der Kombination von vorhandener Evidenz und neuer empirisch generierter Evidenz entsteht dann in Iterationen die sogenannte *contribution story*, die ein umfassendes Bild des Beitrags zeichnet, den das Programm im angenommenen Wirkungszusammenhang geleistet hat.

Im Prinzip steckt in der Forderung, vorhandene Evidenz planmäßig und ex ante zu nutzen, eine weitere implizite Kritik am experimentellen Ansatz, der Untersuchungsgegenstände in der Regel als *black box* behandelt als wären sie gleichermaßen unbekannt, und alles Vorwissen zu ähnlichen Maßnahmen ungenutzt lässt. Die Argumentation ähnelt an dieser Stelle übrigens frappierend jener, die aus Perspektive der Bayes-Statistik am frequentistischen Signifikanztest geübt wird (Austin/Brunner/Hux 2002), wie er in der Regel zur Auswertung experimenteller Designs zur Anwendung kommt: Durch das Ignorieren vorhandener Information, in diesem Falle zur Wahrscheinlichkeit, dass das Ergebnis in die eine oder andere Richtung ausschlägt, stelle man sich unnötig dumm und verschwende wertvolle Information, die zur Stärkung der Sicherheit eines Kausalschlusses dienen kann.

Resultat einer Contribution Analysis ist neben der *contribution story* ein *contribution claim*, die begründbare Behauptung, dass die untersuchte Maßnahme einen substanziellen Kausalbeitrag zu einer beobachteten Veränderung geleistet hat. Sie ist dem Ansatz zufolge berechtigt, wenn vier Bedingungen erfüllt sind (Mayne 2012, S. 272–273): (1) Eine plausible, ex ante von vorhandener Evidenz und Expertise sowie Stakeholdern gestützte Programmtheorie liegt vor, (2) die Programmaktivitäten wurden gemäß der Programmtheorie umgesetzt, (3) die Programmtheorie wird ex post entlang ihrer erwarteten Wirkungskette auf Basis von auf das Programm bezogenen Datenerhebungen und Expertise gestützt bzw. nicht widerlegt und (4) die Rolle möglicher alternativer Einflüsse wurde berücksichtigt.

2.4.4 Process Tracing

Process Tracing weist eine gewisse Nähe zur Contribution Analysis auf und kann in bestimmten Punkten als Weiterentwicklung begriffen werden (Befani/Mayne 2014). Der Ansatz will gegenüber anderen Ansätzen eine elaboriertere Methodologie bereitstellen, die einzelne Kausalbehauptungen in einer Programmtheorie retrospektiv und nicht-experimentell auf Basis eines einzelnen Falls unterfüttern kann. Im Fokus ist hier immer eine einzelne Ursache-Wirkungs-Beziehung innerhalb einer Programmtheorie, was in komplexen Evaluationsgegenständen mit entsprechend umfassenderen Wirkungsgefügen eine separate Analyse für jeden einzelnen Wirkungspfeil eines Kausalzusammenhangs impliziert.

Der Begriff des „Tracing“ speist sich daraus, dass in der Ex-post-Perspektive detektivähnlich nach empirischen „Spuren“ gesucht wird, die zur Stützung oder Widerlegung eines einzelnen angenommenen Kausalzusammenhangs dienen können (Befani/Stedman-Bryce 2017). Diese Spuren können einerseits aus dezidierten Datenerhebungen stammen, die für diesen Zweck im Rahmen der Evaluation vorgenommen wurden, aber auch aus jeder anderen Art von unabhängig davon verfügbaren Daten. Egal welcher Art, werden diese empirischen Spuren jeweils daraufhin geprüft, inwiefern, wie verlässlich und mit welcher Stärke sie den jeweiligen hypothetischen Kausalzusammenhang, zum Beispiel „Die Kampagne X beeinflusste die Institution Z, die Entscheidung Y zu treffen“, stützen oder gegen ihn sprechen. Zum Einsatz kommen dazu vier Arten von metaphorisch bezeichneten Tests, die an jedes aus der Programmtheorie abgeleitete „Beweisstück“ (*evidence*) angelegt werden, um zu bewerten, inwiefern sie die im Wirkungszusammenhang enthaltene Hypothese stützen oder sie widerlegen (Befani/Stedman-Bryce 2017, S. 45):

- *Smoking Gun*: Falls das Beweisstück beobachtet wird, bestätigt es die Hypothese; sein Fehlen reicht aber nicht (allein), um sie zu widerlegen.
- *Hoop Test*: Falls das Beweisstück fehlt (nicht beobachtet wird), widerlegt es die Hypothese; im umgekehrten Fall reicht es aber nicht (allein), um sie zu bestätigen.
- *Doubly Decisive*: Falls beobachtet, bestätigt das Beweisstück die Hypothese; falls fehlend, widerspricht es ihr.
- *Straw-in-the-Wind*: Das Beweisstück dient als Hinweis, gilt aber nicht als ausreichend, um die Hypothese zu belegen bzw. im umgekehrten Fall sie zu widerlegen.

Besonders jüngere Ergänzungen erweitern den Ansatz, um aus der Bayes-Statistik (Austin/Brunner/Hux 2002) entlehnte Überlegungen. Sie fokussieren die Frage, wie sicher man sich bereits aufgrund ex ante vorliegender Evidenz ist, dass eine Hypothese zutrifft, und welche Quantität und Qualität von „Beweisstücken“ daher gebraucht werden, um sie unter Einsatz von einem oder mehreren der vier

Tests zu bestätigen oder zu widerlegen (Befani/Stedman-Bryce 2017; Schmitt/Beach 2015). Resultat des Ansatzes ist dann idealerweise eine an jeder Stelle der Wirkungskette durch entsprechende Beweisstücke (empirisch beobachtete Sachverhalte) möglichst stark belegte Programmtheorie, die im Gesamtbild dann Auskunft über die grundsätzliche Wirkung der Maßnahme und gegebenenfalls bestimmte Schwachpunkte in der Wirkungskette geben kann.

Seine potenzielle Stärke, auch im Bereich der Evaluation von Maßnahmen der Demokratieförderung und Extremismusprävention, ist darin zu sehen, dass der Ansatz auch dann eingesetzt werden kann, wenn nicht ausreichend Fälle für eine quantitative Analyse der Programmtheorie, etwa im Rahmen von Pfadanalysen, Regressions- oder Strukturgleichungsmodellen und ähnlichen Verfahren, zur Verfügung stehen. Im Gegensatz zu rein qualitativen Verfahren, auf die zum Beispiel die Realistic Evaluation in starkem Maße setzt, steht hier ein Instrumentarium bereit, das die Verlässlichkeit der Wirkungsbehauptungen bis zu einem gewissen Grad quantifiziert und damit eindeutiger nachvollziehbar machen will.

Auch wenn der Ansatz sich darauf nicht explizit bezieht, hat die Grundidee der Spurensuche interessanterweise eine große Ähnlichkeit zu Michael Scrivens Idee der oben schon erwähnten Kausalanalyse mittels der „modus operandi“-Methode (Scriven 1974). Etwas übereifrig wirkt, dass Process Tracing sich aus einer teils oberflächlichen Kritik bestehender theoriebasierender Ansätze begründen will (Schmitt/Beach 2015). So werden etwa theoriebasierte Evaluationen implizit auf die Arbeit mit dem einfachen Logischen Modell reduziert, was rein empirisch in der Evaluationspraxis durchaus oft zutreffend sein mag, aber der umfassenden theoretischen und methodologischen Fundierungen des Ansatzes, unter anderem durch Weiss, Rossi, Chen, Donaldson oder Funnell und Rogers, nicht gerecht wird (Hense/Taut 2021). Auch wird der kausalprüfende Ansatz der Contribution Analysis zu stark auf reine Stakeholder-Einschätzungen reduziert (Mayne 2019). Diese sich abgrenzende Kritik erscheint aber unnötig, da Process Tracing erkennbar eine interessante Ergänzung bestehender Ansätze zur Prüfung von Kausalzusammenhängen bereitstellt, die auch in der Evaluation der Arbeit für Demokratie und gegen Extremismus zur Anwendung kommen kann.

2.5 Theorie als Grundlage der Wirkungsevaluation

Wie dargestellt, kann also die gezielte Nutzung von Programmtheorie in Evaluationen unter bestimmten Bedingungen nicht nur dabei helfen, mit einem gewissen Grad an *credibility* die Attributionsproblematik zu adressieren. Sie verspricht daneben einen über die Frage „Hat es gewirkt?“ hinausgehenden größeren Informationsgehalt im Sinne der Frage „Was wirkt für wen und unter welchen Bedingungen?“ Dieser ist einerseits für formative Verbesserungen von Maßnahmen

und andererseits für den Transfer von Erfolgsmodellen auf andere Kontexte essenziell.

Programmtheorien können dabei grundsätzlich ein breites Spektrum hinsichtlich des Grades an Wissenschaftlichkeit aufweisen, der in sie Eingang gefunden hat. Dieses reicht von erfahrungsbasierten Alltagstheorien der Praxis bis hin zu empirisch umfassend bewährten theoretischen Ansätzen zur Erklärung und Prognose von Veränderungen aus dem Gegenstandsbereich der Intervention bzw. Prävention.

Geht es allerdings um die wirkungsorientierte Konzeption und Planung von Maßnahmen, können die beiden Enden dieses Spektrums nicht gleichwertig behandelt werden. Steht vorab geeignetes theoretisches Wissen, unterfüttert mit robuster Empirie, zur Verfügung, das bei der Planung intendierter Wirkungen Orientierung geben kann, wäre jedes Programmdesign fahrlässig, das dieses Wissen ignoriert. Selbstverständlich soll dies nicht die Rolle des Erfahrungswissens der Praxis (Schön 1983) herunterspielen, das zum Beispiel immer dann benötigt wird, wenn es darum geht, Programmaktivitäten im Kontext der spezifischen Gegebenheiten vor Ort umzusetzen und dabei auch gegebenenfalls notwendige Anpassungen des ursprünglichen Konzepts vorzunehmen (Meyers/Nielsen 2012; Blakely et al. 1987). Dennoch sollte es auch bei innovativen Programmen eine Selbstverständlichkeit sein, diese nach Möglichkeit sowohl auf Basis des verfügbaren Wissens als auch unter Einbezug von Erfahrungswissen aus der Praxis zu konzipieren und zu planen, um so ihre Erfolgspotenziale voll auszuschöpfen und mögliche Wirkungshindernisse vorab zu erkennen.

Wenn Theorie in diesem Sinne stärker von Beginn an den Lebenszyklus von Programmen begleiten würde, wäre dies aber nicht nur aus Programm-, sondern auch aus Evaluationsperspektive wünschenswert. Zwar stammt Programmtheorie als Begriff aus dem Evaluationsdiskurs, und Wirkungsmodelle werden häufig erst zu Beginn von Evaluationsprojekten aus den impliziten Wirkungsannahmen des Programms rekonstruiert. Dies sollte aber nicht darüber hinwegtäuschen, dass sie primär Werkzeuge von Programmdesign und Maßnahmenplanung sind (Funnell/Rogers 2011). Sieht man also die verstärkte Nutzung von Programmtheorie als Schlüssel für möglichst nützliche, glaubwürdige und machbare Wirkungsevaluationen, erscheint es auch aus Evaluationsperspektive wichtig, dass sie bereits bei Programmdesign und -planung eine zentralere Rolle spielt.

Ebenso wie für andere Politikfelder ergeben sich daraus für das Feld der Demokratieförderung und Extremismusprävention je nach retrospektiver oder prospektiver Blickrichtung zweierlei Fragen, die in den folgenden Abschnitten thematisiert werden sollen. Die erste betrifft die theoretische Fundierung bestehender Programme und Einzelmaßnahmen, die zweite, inwiefern überhaupt etablierte Theorien zur Fundierung von Maßnahmen zur Verfügung stehen.

2.5.1 Theoretische Fundierung bestehender Maßnahmen

Die erste Frage ist für das Feld der Demokratieförderung und Extremismusprävention derzeit wohl nicht umfassend zu beantworten, da sie eine systematische empirische Analyse bestehender und abgeschlossener Maßnahmen erfordern würde. In der umfassenden Literatur zu Programmtheorien herrscht der Eindruck vor, dass Programme häufiger ohne als mit explizitem Theoriebezug geplant werden (Hense/Taut 2021). Daher ist damit zu rechnen, dass eine theoretische Fundierung oft nicht explizit zum Beispiel als theoretischer Teil eines Projektantrags vorliegt, sondern eher rekonstruktiv gewonnen werden müsste. Bestehende Programme und Projekte müssten also daraufhin untersucht werden, ob sich die zentralen Annahmen ihrer Wirkungslogik, zum Beispiel „Das gemeinsame Fußballspielen von Einheimischen mit Geflüchteten vermindert Vorurteile und erhöht Verständnis und Toleranz“, theoretisch begründen lassen, was im Beispiel etwa mit Verweis auf die psychologische Intergruppenkontakttheorie (Pettigrew et al. 2011) erfolgen könnte. Ein entsprechendes Forschungsprogramm würde angesichts der Vielfalt von Projekten und Einzelmaßnahmen, die beispielsweise allein in einem Programm wie „Demokratie leben!“ vertreten sind, eine Mammutaufgabe darstellen.

Einen Ausweg zeigt die Zusammenfassung von Projekten mit wiederkehrenden Handlungs- und Wirkungslogiken auf. Denn trotz aller Unterschiedlichkeit im Detail lassen sich ab einem gewissen Abstraktionsniveau gewisse Muster von Interventions- und Präventionsstrategien erkennen, die trotz lokaler Variationen übergreifend in unterschiedlichen Projekten ähnlich verfolgt werden. Im Kontext der Programmtheorie werden solche wiederkehrenden Muster „Programm-Archetypen“ (Funnell/Rogers 2011) oder „Wirkungsmodelltypen“ (Wachsmuth/Brinkmann/Hense 2019) genannt und gelten als wichtige Strategie zur Reduktion des Aufwands bei der Erarbeitung von Programmtheorien.

Dass dieser Ansatz auch im Kontext von Programmen der Demokratieförderung und Extremismusprävention anwendbar ist, zeigt das Beispiel der primär als pädagogische Maßnahmen an junge Menschen gerichteten Modellprojekte im Handlungsfeld Demokratieförderung im Bundesprogramm „Demokratie leben!“. Im Rahmen der wissenschaftlichen Begleitung wurden vier Typen von Maßnahmen mit jeweils spezifischen demokratiefördernden Strategien identifiziert, die clusterartig zusammengefasst werden können (Ehnert et al. 2021, S. 24).

Eine solche Clusterung erlaubt einerseits, dass die Rekonstruktion einer Programmtheorie nicht für jedes Projekt von neuem begonnen werden muss, sondern auf einem gemeinsamen Rahmengerüst basieren kann. Andererseits lässt sich der generelle Interventionsansatz übergeordnet theoretisch begründen, was am Beispiel der vier genannten Cluster etwa unter Rückgriff auf Theorien aus den Bereichen Kompetenzentwicklung, Partizipation, Empowerment und

Capacity Building erfolgen könnte. Welcher Theoriefundus dabei grundsätzlich zur Verfügung steht, ist auch für unsere zweite, prospektiv gerichtete Frage relevant.

2.5.2 Basis zur theoretischen Fundierbarkeit geplanter Maßnahmen

Die bereits hergeleitete Forderung, Wirkungsanalysen in stärkerem Maße auf solide theoretische Füße zu stellen, geht von der Annahme aus, dass für ein Handlungsfeld überhaupt geeignetes Wissen bereitsteht, das anschlussfähig für die Entwicklung von Interventions- bzw. Präventionsmaßnahmen im Feld ist. Gelegentlich entsteht in der Programm- und Projektförderung politikfeldübergreifend der Eindruck, dass dies nicht der Fall ist, wenn etwa Maßnahmen allein die Entwicklung und Erprobung (gänzlich) neuer Ansätze zum Ziel haben oder auf (gänzlich) neue oder sich ständig verändernde Rahmenbedingungen verwiesen wird, die eine solche Entwicklung erfordern würden. Zumindest für Veränderungen auf individueller und Organisationsebene, wie sie letztlich auch viele Maßnahmen der Demokratieförderung und Extremismusprävention zum Ziel haben, schiene es verfehlt, ein grundsätzliches Theoriedefizit zu konstatieren, das dazu zwingt, bei der Konzeption jedes Mal von Null auf neu zu beginnen.

Denn erstens stehen ja umfassende theoretische und Arbeiten feldspezifischer Art zur Verfügung, die für konzeptionelle Maßnahmenplanungen genutzt werden können. In diesem Sinne zählt beispielsweise Beelmann (2017, S. 20–23) als Quellen für die Entwicklung von Programmtheorien im Kontext der Prävention von Rechtsextremismus auf: Empirisch bestätigte Risiko- und Schutzfaktoren, bisherige Evaluationen bzw. deren Aggregation in Meta-Analysen, ätiologische Entwicklungsmodelle allgemeinen oder des spezifischen Problemverhaltens sowie Theorien über positive Entwicklung. Aus psychologischer Perspektive etwa könnten exemplarisch die Intergruppen-Kontakttheorie (Pettigrew et al. 2011) oder die Theorie der kognizierten Kontrolle (Hehnen et al. 2021) genannt werden. Viele weitere Ansätze aus den Bereichen Radikalisierungstheorien, Bindungstheorien, der Theorie relativer Deprivation und weiterer Forschungstraditionen wären zu ergänzen.

Aber selbst wenn für eine konkrete Problemlage oder für ein Entwicklungsziel kein passgenauer Ansatz verfügbar ist, bleiben immer noch übergreifende Ansätze. Denn schließlich ist Veränderlichkeit von Menschen und Organisationen seit Langem genuiner Forschungsgegenstand von Disziplinen wie Psychologie, Erziehungswissenschaft, Sozialpädagogik, (Organisations-)Soziologie oder Politik- und Verwaltungswissenschaft. Diese und weitere Disziplinen stellen einen umfassenden Fundus an Wissen darüber bereit, unter welchen Bedingungen Menschen beispielsweise Kompetenzen, Einstellungen, Motive und Verhal-

ten verändern, Institutionen ihre Ziele, Strukturen, Prozesse und Handlungslogiken anpassen oder wie es zu problematischen Entwicklungen auf diesen Ebenen kommt. Da für das Feld der Demokratieförderung und Extremismusprävention nicht davon auszugehen ist, dass Menschen und Institutionen hier grundsätzlich anders funktionieren als in anderen Bereichen, können auch solche Ansätze hier, ebenso wie in anderen Politikfeldern, grundsätzlich fruchtbar gemacht werden.

Dennoch bleibt der Schritt von der Theorie zum Programmkonzept mit fundiertem Interventionsdesign kein kleiner. Schließlich wollen Theorien in erster Linie einen bestimmten Phänomenbereich beschreibbar, erklärbar und vorhersehbar machen. Die gezielte Intervention im Phänomenbereich dagegen ist etwas, das oft eher abstrakt im Ausblick entsprechender Arbeiten abgehandelt wird, was den Transfer für das Programmdesign nicht leichter macht.

2.6 Konsequenzen für Programmförderung und Evaluation

Welche Schlüsse lassen sich aus dem Diskutierten für das Feld der Demokratieförderung und Extremismusprävention sowie seine Evaluation ziehen? Zunächst sind zwei mögliche Schlüsse zu erwähnen, die nicht gezogen werden sollten. Denn trotz seines Plädoyers dafür, den dem Wirkungsbegriff eingeschriebenen kausalattributionierenden Anspruch nicht zu leichtfertig zu vergessen, atmet der Beitrag ja eine gewisse Skepsis gegenüber der Reichweite des „klassischen“ kausalattributionierenden Ansatzes, dem experimentellen Paradigma der Wirkungsevaluation. Diese Skepsis sollte in zweierlei Hinsicht nicht missverstanden werden.

Erstens hat der experimentelle Ansatz wie jedes andere methodische Werkzeug, das uns in der Evaluation zur Verfügung steht, seine Stärken und Schwächen. Bei einem zweck- und kontextsensitiven Einsatz hat er seinen völlig berechtigten Platz, wie auch von anderen kritischen evaluationstheoretischen Stimmen eingeräumt (z. B. Scriven 2008). In vielen Evaluationskontexten sind es aber weniger die praktischen Probleme seiner Realisierbarkeit, die die Skepsis an dem Werkzeug speisen, als der letztlich begrenzte Informationsgehalt, den es bereitstellt. Wie aus theoriebasierter Perspektive schon immer kritisiert, liefert es zur Frage „Was wirkt für wen unter welchen Bedingungen?“ eben nur einen schmalen Informationsgewinn. Letztlich wird nicht einmal die aus dieser Perspektive kritisierte, eng geführte Frage „Was wirkt?“ (Pawson/Tilley 2005; Chen/Rossi 1992) so beantwortet, wie es die Frageformulierung suggeriert. Denn aufgrund der im RCT problematischen externen Validität wird eher die historische Frage „Hat es gewirkt?“ beantwortet, die eigentlich nur aus Rechenschaftsgründen von Interesse ist, als die für die Zukunft gerade auch im Kontext der Arbeit für Demokratie und gegen Extremismus viel wichtigere Frage „Wird es auch in anderen Kontexten wirken?“

Zweitens ist gelegentlich, wenn man die oben angeführte Skepsis zum experimentellen Ansatz als einzig validem Weg zur Wirkungsevaluation begründet, ein gewisses Aufatmen zu verspüren. „Wenn das mit den Kontrollgruppenexperimenten ohnehin so eine knifflige Sache ist und auch auf anderen Wegen plausible Wirkungsaussagen getroffen werden können, dann können wir uns die Sache ja auch einfacher machen“, scheint hier manches Mal der unausgesprochene Gedankengang zu sein, der sich wiederum mit der eingangs unterstellten Tendenz zu treffen scheint, dass man gelegentlich gern von Wirkung spricht, ohne den Anspruch des Begriffs einzulösen. Nichts wäre aber unangemessener.

Denn, wie gezeigt, sind erstens die Alternativwege zur Kausalanalyse in methodischer Hinsicht nicht unbedingt weniger anspruchsvoll. Egal, ob es die quantitative Modellprüfung in der „klassischen“ theoriebasierten Evaluation ist, das qualitative Aufspüren von CMO-Konfigurationen in der Realistic Evaluation, die Schreibtischrecherche existierender Evidenz bei der Contribution Analysis oder die akribische, theoriegeleitete Spurensuche beim Process Tracing, auch theoriebasierte Ansätze bieten keine verbilligte Annäherung an die Kausalfrage. Und trotz allem, was gegen ein Primat von RCTs spricht, wird der Attributionsanspruch nicht-experimenteller Verfahren immer begründungsbedürftiger bleiben als der des randomisierten Kontrollgruppenexperiments.

Hinzu kommt zweitens die Theoriearbeit. In der Praxis ist hier oft eine gewisse Trivialisierung zu beobachten, die man sich wohl teils selbst mit der Popularisierung des reduzierten, tabellarisch aufgebauten Logischen Modells mit seinem Ansatz des einfachen Ausfüllens (W. K. Kellogg Foundation 2004) zuschreiben muss. Nicht, dass diese einfache Form nicht auch ihre Berechtigung hätte, aber sie kann eben als sehr einfache Wirkungsmodellvariante nicht jede mögliche Funktion von Wirkungsmodellen ausfüllen (Hense/Taut 2021). Insbesondere bei der Überprüfung von Wirkungsannahmen hat sie ihre Grenzen, da diese üblicherweise einen höheren Auflösungsgrad und mehr Differenzierung verlangen. Zusätzlich kommt die inhaltliche Arbeit mit einschlägigen Gegenstandstheorien hinzu, wie im vorigen Abschnitt dargestellt.

Drittens besteht theoriebasierte Evaluation nicht daraus, ein Wirkungsmodell zu entwickeln und dieses dann im weiteren Verlauf der Evaluation weitgehend zu ignorieren. Dennoch war dies das Bild, das in systematischen Reviews in vielen Arbeiten gefunden wurde (Coryn et al. 2011). Es wurde zwar in allen untersuchten Studien eine Form von Programmtheorie präsentiert, im weiteren Verlauf der Evaluation war aber oft nicht nachzuvollziehen, inwiefern diese in den weiteren Evaluationsschritten oder in der Analyse von Wirkungen eine Rolle hatte.

Trotz dieser Herausforderungen liegen mit Blick auf die Gesamtsituation einige Konsequenzen nahe. Angesichts der Tatsache, dass Wirkungsanalysen in der Demokratieförderung und Extremismusprävention in vielerlei Hinsicht dringend benötigt werden, RCTs hier aber nur in Nischen anwendbar und nützlich im Sinne zukunftsgeleiteter Entscheidungen sind, braucht es Alternativen,

die *credible evidence* beisteuern und zugleich wichtige Hinweise zur formativen Verbesserung und für den Transfer auf andere Kontexte beisteuern können. Gute Theorie weist dafür einen Ausweg und sollte daher eine größere Rolle in der Programmförderung und -evaluation, nicht nur im Bereich der Demokratieförderung und Extremismusprävention, spielen.

Die Forderungen an die Evaluation liegen mehr oder weniger auf der Hand: Aus fachlicher Sicht sollten wir grundsätzlich bewusster mit dem Wirkungsbegriff und seinen Implikationen umgehen als dies in vielen Praxiskontexten geschieht. Wenn wir ihn hinsichtlich seines Attributionsversprechens ernst nehmen, sollten wir dies mit angemessenen wirkungsanalytischen Ansätzen, von denen eine Auswahl im Beitrag besprochen wurde, unterfüttern können. Dies verlangt vor allem eine Kompetenz im Umgang mit Programmtheorie, die nicht beim einfachen Logischen Modell aufhört oder sich auf das Anfertigen von Grafiken mit Kästchen und Pfeilen beschränkt. Idealerweise umfasst sie eine Kenntnis sowohl der Potenziale als auch der Kosten und Risiken ihres Einsatzes und eine entsprechend bewusste Nutzung in den verschiedenen Phasen einer Evaluation (Hense/Taut 2021).

Allerdings wäre das Potenzial des programmtheoretischen Ansatzes verschenkt, beschränkte man ihn nur auf die Evaluation. Denn manches Problem, das in Evaluationen aufgedeckt wird, könnte bei einer veränderten Projektplanung und -durchführung bereits frühzeitiger adressiert werden. Zugleich werden programmtheoriebasierte Wirkungsevaluationen natürlich erleichtert, wenn sie auf einer explizit dargelegten Wirkungslogik aufsetzen können und diese nicht erst rekonstruieren müssen. Die wichtigeren Forderungen richten sich somit an die Ebenen der Programmförderung und des Projektdesigns:

- Grundsätzlich empfiehlt sich eine Orientierung an einem ganzheitlich gedachten Ansatz der Wirkungsorientierung. Er beginnt bei der Maßnahmenplanung, erstreckt sich über die Maßnahmendurchführung und -evaluation und endet (vorläufig) beim Lernen aus der Evaluation im Sinne begründbarer Konsequenzen für Verbesserung, Fortsetzung oder Transfer von Maßnahmen.
- Bei der Maßnahmenplanung hilft die Auseinandersetzung mit der intendierten Wirkungslogik gegen Projektplanung nach dem „ISLAGIATT“-Prinzip (it seemed like a good idea at the time), vor allem indem sie das Prüfen kritischer Annahmen und Wirkungserwartungen provoziert.
- Um dies zu ermöglichen, bräuchten jene, die Maßnahmen wie Programme und Projekte konzipieren, entweder eigene Kompetenzen im Entwickeln von Programmtheorien oder entsprechend erfahrene externe Begleitung oder Schulung. Es ist nicht realistisch zu erwarten, dass diese überall vorhanden ist, ein Problem, das sich nicht kurzfristig und grundsätzlich lösen lassen wird. Ein Beitrag zur Lösung könnten zweistufige Antragsprozesse sein,

bei denen zur Ausarbeitung eines Vollantrags bei Bedarf entsprechende Ressourcen zur professionellen Begleitung zur Verfügung gestellt werden.

- Ein weiterer Beitrag könnten vorab definierte Projekt-Cluster sein, die einer gemeinsamen, bereits vorbereiteten Wirkungslogik folgen. Diesen könnten vorab erarbeitete und theoretisch fundierte Vorlagen für Wirkungsmodelle im Sinne der oben erwähnten Archetypen oder Wirkungsmodelltypen zur Verfügung gestellt werden.
- Davon unabhängig sollte mindestens auf Programmebene wissenschaftliche Expertise aus der jeweiligen Inhaltsdomäne, zum Beispiel in Form von programm vorbereitenden Gutachten, bereitgestellt werden, die einen Überblick über potenziell nutzbares theoretisches Wissen für die Projektkonzeption bereitstellt. Wie oben dargestellt, kann dieses empirisch bestätigte Risiko- und Schutzfaktoren, bisherige Evaluationen, Entwicklungsmodelle zum Problem-bereich oder Theorien über positive Entwicklung umfassen.
- Der Antrags- und Auswahlprozess für geförderte Projekte sollte einen Mindestgrad an theoretischer Fundierung der Projekte einfordern, was besonders praxisnahe und kleinere Einrichtungen ausschließen könnte. Diese benötigen daher entweder zusätzliche unterstützende Ressourcen oder es bedarf unterschiedlicher Förderlinien, etwa für „kleinere“ Projekte mit geringeren Ansprüchen an eine programmtheoretische Fundierung, vor allem dann, wenn der Antrags- und Auswahlprozess, was zu empfehlen wäre, differenzierter Wirkungsmodelle umfassen sollen, die über das Logische Modell hinaus die zentralen Wirkungsannahmen identifizieren und vorab prüfbar machen.
- Im Stiftungskontext haben sich Formate bewährt, in denen Vorhaben ex ante auf Basis eines Wirkungsmodells auf ihre Umsetzbarkeit und Wirkungspotenziale hin analysiert und bereits vor der Realisierung optimiert werden. Mögliches Format sind Workshops unter Beteiligung von Projektdesign, wissenschaftlicher Expertise, Expertise aus der Praxis und gegebenenfalls Zielgruppenmitgliedern (Wachsmuth/Brinkmann/Hense 2019). Ein solcher „plausibility check“ (Brousselle/Champagne 2011; Donaldson 2007) widerspricht der klassischen Antragslogik in Förderprogrammen, wäre aber ein äußerst vielversprechendes Mittel, um die Umsetzbarkeit und potenzielle Wirksamkeit von Maßnahmen bereits vorab zu steigern.

Eine Befürchtung mit Blick auf all diese Vorschläge könnte sein, dass sie besonders innovative Vorhaben ausbremsen könnten. Auch wird verschiedentlich gerne auf Methoden wie agiles Projektmanagement verwiesen, die dem Vorab-Formulieren einer Programmtheorie widersprechen würden. In Antwort darauf könnte man fragen, ob Innovation denn automatisch einen theoretischen Blindflug impliziert. Auch Projekte etwa, die Innovatives entwickeln und erproben wollen, verfolgen dabei einen (impliziten) theoretischen Ansatz, dessen Prämissen expliziert und überprüfbar gemacht werden können. Grundsätzlich gilt: Wer mit Aktivitä-

ten Ziele verfolgt, hat immer mindestens eine implizite Theorie, warum gerade diese Aktivitäten dabei helfen, diese Ziele zu erreichen. Wer sie nicht einmal in Ansätzen vorab explizieren kann, verdient Skepsis.

Denn wir sind es letztlich sowohl dem verantwortungsvollen Gegenstand der Demokratieförderung und Extremismusprävention als auch der Öffentlichkeit, die ihre Maßnahmen finanziert, dringend schuldig, dass wir uns nicht dem politisch immer noch verbreiteten Input-Output-Denken verschreiben, das zufrieden auf investierte Summen und die Anzahl von damit ermöglichten Aktivitäten verweist. Die Sorge um und Fürsorge für das demokratische Gemeinwesen gebietet es, dass wir sowohl auf Gegenstandsebene der Programm- und Projektförderung als auch auf Evaluationsebene alles dafür tun, dass das Richtige richtig getan wird. Und auch wenn das häufig zitierte Bonmot „Nichts ist so praktisch wie eine gute Theorie“ nicht von Kurt Lewin stammt, dem es in der Regel zugeschrieben wird (Bedeian 2016), ändert das nichts an seiner Anwendbarkeit auch auf diesen Anspruch.

Literatur

- Albright, L. / Malloy, T. E. (2000): Experimental Validity: Brunswik, Campbell, Cronbach, and Enduring Issues. In: *Review of General Psychology*, 4. Jg., H. 4, S. 337–353.
- Austin, P. C. / Brunner, L. J. / Hux, J. E. (2002): Bayeswatch: an overview of Bayesian statistics. In: *Journal of Evaluation in Clinical Practice*, 8. Jg., H. 2, S. 277–286.
- Baron, R. M. / Kenny, D. A. (1986): The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. In: *Journal of Personality and Social Psychology*, 51. Jg., H. 6, S. 1173–1182.
- Bedeian, A. G. (2016): A note on the aphorism „there is nothing as practical as a good theory“. In: *Journal of Management History*, 22. Jg., H. 2, S. 236–242.
- Beelmann, A. (2017): Grundlagen einer entwicklungsorientierten Prävention des Rechtsextremismus. Gutachten im Rahmen des Wissenschafts-Praxis-Dialogs zwischen dem Landespräventionsrat Niedersachsen und der Friedrich-Schiller-Universität Jena. Jena. dpt-statisch.s3.eu-central-1.amazonaws.com/dpt-digital/dpt-25/medien/dateien/368/Gutachten-LPR-Niedersachsen.pdf (Abruf 04.06.2024).
- Befani, B. (2012): Models of Causality and Causal Inference. In: Stern, E. / Stame, N. / Mayne, J. / Forss, K. / Davies, R. / Befani, B. (Hrsg.): *Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development, Working Paper 38, Anhang*, S. 92–115.
- Befani, B. / Mayne, J. (2014): Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. In: *IDS Bulletin*, 45. Jg., H. 6, S. 17–36.
- Befani, B. / Stedman-Bryce, G. (2017): Process Tracing and Bayesian Updating for impact evaluation. In: *Evaluation*, 23. Jg., H. 1, S. 42–60.
- Blakely, C. H. / Mayer, J. P. / Gottschalk, R. G. / Schmitt, N. / Davidson, W. S. / Roitman, D. B. / Emshoff, J. G. (1987): The fidelity-adaptation debate: Implications for the implementation of public sector social programs. In: *American Journal of Community Psychology*, 15. Jg., H. 3, S. 253–268.
- Brousselle, A. / Champagne, F. (2011): Program theory evaluation: Logic analysis. In: *Evaluation and Program Planning*, 34. Jg., H. 1, S. 69–78.

- Caspari, A. (2012): Chancen der Wirkungsorientierung für die entwicklungspolitische Bildungsarbeit. In: *Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik*, 35. Jg., H. 2, S. 11–17.
- Chen, H.-T. (1990): *Theory-Driven Evaluations*. Newbury Park: Sage.
- Chen, H.-T./Rossi, P. H. (1983): Evaluating With Sense. In: *Evaluation Review*, 7. Jg., H. 3, S. 283–302.
- Chen, H.-T./Rossi, P. H. (1992): *Using theory to improve program and policy evaluations*. New York: Bloomsbury Publishing.
- Cook, T. D./Campbell, D. T. (1979): *Quasi-experimentation design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cook, T. D./Pohl, S./Steiner, P. M. (2011): Die relative Bedeutung der Kovariatenwahl, Reliabilität und Art der Datenanalyse zur Schätzung kausaler Effekte aus Beobachtungsdaten. In: *Zeitschrift für Evaluation*, 10. Jg., H. 2, S. 203–224.
- Cook, T. D./Shadish, W. R./Wong, V. C. (2008): Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. In: *Journal of Policy Analysis and Management*, 27. Jg., H. 4, S. 724–750.
- Coryn, C. L. S./Noakes, L. A./Westine, C. D./Schröter, D. C. (2011): A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. In: *American Journal of Evaluation*, 32. Jg., H. 2, S. 199–226.
- Cronbach, L. J. (1980): *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L. J. (1982): *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Davidson, E. J. (2000): Ascertaining causality in theory-based evaluation. In: *New Directions for Evaluation*. Vol. 2000, H. 87, S. 17–26.
- Davidson, E. J. (2005): *Evaluation methodology basics. The nuts and bolts of sound evaluation*. Thousand Oaks: Sage.
- Donaldson, S. I. (2001): Mediator and moderator analysis in program development. In: Sussman, S. Y. (Hrsg.): *Handbook of program development for health behavior research & practice*. Thousand Oaks: Sage, S. 470–496.
- Donaldson, S. I. (2007): *Program theory-driven evaluation science. Strategies and applications*. Mahwah, NJ: Erlbaum.
- Donaldson, S. I. (2022): *Introduction to theory-driven program evaluation. Culturally responsive and strengths-focused applications. Culturally Responsive and Strengths-Focused Applications*. 2. Auflage. New York: Routledge.
- Earl, S./Carden, F./Patton, M. Q./Smutylo, T. (2001): *Outcome mapping. Building learning and reflection into development programs*. Ottawa: The International Development Research Centre (IDRC).
- Ehnert, K./Johann, T./Mielke, A./Rehse, A./Walter, E./Zimmermann, E. (2021): *Handlungsfeld und Handlungslogiken. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld Demokratieförderung im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von C. Wagner*. Halle (Saale): DJI.
- Foerster, H. v./Ollrogge, B. (1993): *Kybernetik*. Berlin: Merve.
- Funnell, S. C./Rogers, P. J. (2011): *Purposeful program theory. Effective use of theories of change and logic models*. San Francisco: Jossey-Bass.
- Gargani, J. (2003): *The history of theory-based evaluation: 1909 to 2003*. Reno, NV.
- Gates, E./Dyson, L. (2017): Implications of the Changing Conversation About Causality for Evaluators. In: *American Journal of Evaluation*, 38. Jg., H. 1, S. 29–46.
- Hansen, H. F./Rieper, O. (2009): The Evidence Movement. The Development and Consequences of Methodologies in Review Practices. In: *Evaluation*, 15. Jg., H. 2, S. 141–163. <https://doi.org/10.1177/1356389008101968>.

- Hehnen, M. M. /Frey, D. /Rez, H. /Fladerer, M. (2021): Analyse des Erstarkens rechtspopulistischer Bewegungen aus Perspektive der Theorie der kognizierten Kontrolle am Beispiel der Partei „Alternative für Deutschland“. In: Politische Psychologie, 9. Jg., H. 1, 6–19.
- Henry, G. T. (2005): Realist Evaluation. In: Mathison, S. (Hrsg.): Encyclopedia of evaluation. Thousand Oaks: Sage, S. 359–362
- Hense, J. U. (2006): Selbstevaluation. Erfolgsfaktoren und Wirkungen eines Ansatzes zur selbstbestimmten Qualitätsentwicklung im schulischen Bereich, Frankfurt am Main: Peter Lang.
- Hense, J. U./Taut, S. (2021): Wie Wirkungsmodelle zur Wirkung kommen. Nutzungsvarianten, Mehrwert und Kosten ihrer Verwendung in der Evaluationspraxis. In: Zeitschrift für Evaluation. 20. Jg., H. 2, S. 267–292. <https://doi.org/10.31244/zfe.2021.02.02>.
- Holland, P. W. (1986): Statistics and Causal Inference. In: Journal of the American Statistical Association, 81. Jg., H. 396, S. 945–960. <https://doi.org/10.1080/01621459.1986.10478354>.
- LaFrance, J. /Nichols, R. /Kirkhart, K. E. (2012): Culture writes the script: On the centrality of context in indigenous evaluation. In: New Directions for Evaluation, Vol. 2012, H. 135, S. 59–74. <https://doi.org/10.1002/ev.20027>.
- Lipsey, M. W. (1990): Design sensitivity. Statistical power for experimental research. Newbury Park: Sage.
- Mark, M. M. /Henry, G. T. /Julnes, G. (2000): Evaluation. An integrated framework for understanding, guiding, and improving public and nonprofit policies and programs. San Francisco: Jossey-Bass.
- Mayne, J. (2012): Contribution analysis: Coming of age? In: Evaluation, 18. Jg., H. 3, S. 270–280. <https://doi.org/10.1177/1356389012451663>.
- Mayne, J. (2019): Revisiting Contribution Analysis. In: Canadian Journal of Program Evaluation, 34. Jg., H. 2, S. 171–191.
- Meyers, M. K. /Nielsen, V. L. (2012): Street-level bureaucrats and the implementation of public policy. In: Peters, B. G. /Pierre, J. (Hrsg.): The Sage Handbook of Public Administration. 2. Auflage. London: Sage, S. 305–318.
- Pawson, R. /Tilley, N. (1997): Realist Evaluation. Thousand Oaks, CA: Sage
- Pawson, R. /Tilley, N. (2005): Realistic Evaluation. In: Mathison, S. (Hrsg.): Encyclopedia of evaluation. Thousand Oaks: Sage, S. 362–367.
- Pearl, J. (2013): Causality. Models, reasoning, and inference. 2. Auflage. Cambridge: Cambridge University Press.
- Pettigrew, T. F. /Tropp, L. R. /Wagner, Ulrich/Christ, Oliver (2011): Recent advances in intergroup contact theory. In: International Journal of Intercultural Relations, 35. Jg., H. 3, S. 271–280. <https://doi.org/10.1016/j.ijintrel.2011.03.001>.
- Reichardt, C. S. (2022): The Counterfactual Definition of a Program Effect. In: American Journal of Evaluation, 43. Jg., H. 2, S. 158–174. <https://doi.org/10.1177/1098214020975485>.
- Schmitt, J./Beach, D. (2015): The contribution of process tracing to theory-based evaluations of complex aid instruments. In: Evaluation, 21. Jg., H. 4, S. 429–447. <https://doi.org/10.1177/1356389015607739>.
- Schön, D. A. (1983): The reflective practitioner. How professionals think in action. New York: Basic Books.
- Scriven, M. (1974): Maximizing the power of causal investigations: The modus operandi method. In: Popham, W. J. (Hrsg.): Evaluation in education. Current applications. Berkeley, Calif.: McCutchan, S. 68–84.
- Scriven, M. (1991): Evaluation thesaurus. 4. Auflage. Newbury Park: Sage.
- Scriven, M. (2008): A Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research. In: Journal of Multi Disciplinary Evaluation, 5. Jg., H. 9, S. 11–24. <https://doi.org/10.56645/jmde.v5i9.160>.

- Shadish, W. R./Cook, T. D. (2009): The Renaissance of Field Experimentation in Evaluating Interventions. In: *Annual Review of Psychology*, 60. Jg., H. 1, S. 607–629. <https://doi.org/10.1146/annurev.psych.60.110707.163544>.
- Shadish, W. R./Cook, T. D./Campbell, D. T. (2002): *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shadish, W. R./Cook, T. D./Campbell, D. T. (2013): *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth.
- Shadish, W. R./Cook, T. D./Leviton, L. C. (1991): *Foundations of program evaluation. Theories of practice*. Newbury Park: Sage.
- Smith, G. C. S./Pell, J. P. (2003): Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. In: *BMJ*, 327. Jg., H. 7429, S. 1459–1461. <https://doi.org/10.1136/bmj.327.7429.1459>.
- Stufflebeam, D. L./Coryn, C. L. S. (2014): *Evaluation Theory, Models, and Applications*. 2. Auflage. San Francisco: Jossey-Bass.
- W. K. Kellogg Foundation (2004): *Logic Model Development Guide: Using Logic Models to Bring Together Planning, Evaluation, and Action* (Updated January 2004). Battle Creek, Michigan.
- Wachsmuth, E./Brinkmann, H./Hense, J. (2019): Wirkungsorientierung in der Non-Profit-Programmarbeit: Wirkungsmodelle und Wirkungsmodelltypen als Instrumente der Steuerung und Evaluation. In: *Zeitschrift für Evaluation*, 18. Jg., H. 2, S. 291–306.
- Weiss, C. H. (1998): *Evaluation. Methods for studying programs and policies*. 2. Auflage. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C. H. (2002): What to Do until the Random Assigner Comes. In: Mosteller, F./Boruch, R. (Hrsg.): *Evidence matters. Randomized trials in education research*. Washington, D. C.: Brookings Institution Press, S. 198–224.
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./Lobnermeier, O./Heitmeyer, W. (Hrsg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden: Springer VS, S. 41–68. https://doi.org/10.1007/978-3-531-19009-9_3.
- Wolbring, T. (2014): Kausalanalyse und Wirkungsevaluation. Potential Outcomes, Graphenmethodologie und ihre Anwendung am Beispiel der Bologna-Reform. In: *Zeitschrift für Evaluation*, 13. Jg., H. 2, S. 243–270.

3 Gegenstandsangemessenheit in Wirkungsevaluationen

Frank Greuel

„Gegenstandsangemessenheit“ bezeichnet im Allgemeinen einen erwünschten Zusammenhang zwischen einerseits dem Gegenstand, auf den sich eine wissenschaftliche Evaluation bezieht, und andererseits dem Zugriff, durch den der Gegenstand im Rahmen der Evaluation beleuchtet wird. Dieser Zugriff wird dabei vor allem durch theoretische Prämissen, durch die verfolgten Fragestellungen, durch methodologische Vorannahmen wie auch methodische Entscheidungen geprägt. All diese Entscheidungen bestimmen, mit welcher „Brille“ der Gegenstand erfasst und in welchem Licht er betrachtet wird. „Gegenstandsangemessenheit“ bezeichnet vor diesem Hintergrund den Anspruch, dass die eingenommene Perspektive dem Gegenstand gerecht wird. Vermieden werden soll, dass wesentliche Eigenschaften unberücksichtigt bleiben und damit ein verzerrtes Abbild des Gegenstands entsteht und damit letztlich verzerrte Erkenntnisse generiert werden. Grundlage für eine Einschätzung der Gegenstandsangemessenheit ist eine klare Gegenstandstheorie, also eine Vorstellung davon, was den Gegenstand ausmacht und welche wesentlichen Eigenschaften er aufweist.¹

Gegenstandsangemessenheit als Anspruch an Evaluationen speist sich vor allem aus zwei Quellen. Zum einen stellt sie ein Gütekriterium empirischer (qualitativer) Sozialforschung dar (Strübing et al. 2018, S. 88), zum anderen ist sie Bestandteil verschiedenster Standards der Evaluation, wie sie unter anderem von der Gesellschaft für Evaluation (DeGEval 2017) formuliert wurden. Diese zwei verschiedenen Quellen bilden den Ausgangspunkt für diesen Beitrag, und es wird im Folgenden (Abschnitte 3.1 und 3.2) zunächst näher beschrieben, welche Inhalte unter dem Thema Gegenstandsangemessenheit verhandelt werden. Auf dieser Basis wird dann spezifiziert, was Gegenstandsangemessenheit im

1 Das setzt voraus, dass eine entsprechende Wissensbasis zum Gegenstand bereits vorhanden ist, bevor theoretische und methodische Entscheidungen getroffen werden. Mitunter steht dieses Wissen jedoch nicht oder nicht ausreichend zur Verfügung, weil es sich um weitgehend unbekannte Gegenstände handelt. Hier gilt es, während der Arbeit zum Gegenstand fortwährend die Gegenstandsangemessenheit zu prüfen und gegebenenfalls theoretisch/methodisch oder auch mit Blick auf die verfolgten Fragestellungen Veränderungen vorzunehmen, die dann wieder Gegenstandsangemessenheit herstellen.

Kontext von Wirkungsevaluationen bedeuten kann und welche Implikationen insbesondere im Bereich der Evaluation von Wirkungen im pädagogischen Feld von Demokratieförderung, Vielfaltgestaltung und Extremismusprävention entstehen (Abschnitt 3.3). Die entsprechenden Überlegungen werden abschließend in einem Fazit (Abschnitt 3.4) gebündelt.

3.1 Gegenstandsangemessenheit als Gütekriterium empirischer (qualitativer) Sozialforschung

Wie bereits beschrieben, meint Gegenstandsangemessenheit eine Passung zwischen einerseits dem Gegenstand, auf den sich wissenschaftliche Forschung bezieht, und andererseits dem forschenden Zugriff, durch den der Gegenstand beleuchtet wird. In empirischen Untersuchungen stellt sich entlang der Anforderung an Gegenstandsangemessenheit das Problem der Entscheidung für ein geeignetes Forschungsdesign. Die Auswahl für einen bestimmten forschenden Zugriff muss nach gründlicher Auseinandersetzung mit dem spezifischen Forschungsgegenstand und der Forschungsfrage erfolgen und möglichst gut zu ihnen passen. Die Frage der Passung zum Gegenstand ist generell für jegliche empirische Forschung relevant, hat aber im Bereich der qualitativen Sozialforschung besonderen Stellenwert und ist als Gütekriterium anerkannt (Strübing et al. 2018, S. 88). Damit verbindet sich eine vergleichsweise starke Intensität der Auseinandersetzung mit Fragen der Gegenstandsangemessenheit (z. B. Kromrey 2006, S. 12; Nover 2022, S. 3 f.). Gegenstandsangemessenheit umfasst dabei höchst unterschiedliche Ebenen und bezieht sich auf konkrete Eigenschaften des Gegenstands wie auch der untersuchten Gruppe(n) bzw. Individuen bzw. Situation(en), der Fragestellung, aber auch der zur Verfügung stehenden Ressourcen. Entsprechend ist für Stefer (2013, S. 61 f.) Gegenstandsangemessenheit dann gegeben, wenn

- die gewählte(n) Methode(n) sämtliche für das jeweilige Forschungsvorhaben relevanten Daten und Informationen erfasst,
- „die Besonderheiten von Forschungsgegenstand (beispielsweise die Sprachkompetenz der Probanden) und Forschungsfrage (etwa die Einnahme eines speziellen Blickwinkels)“ (Stefer 2013, S. 61) berücksichtigt werden und außerdem
- „weitere beeinflussende Faktoren (z. B. Zeit- und Ressourcenbeschränkungen)“ (Stefer 2013, S. 62) beachtet werden.

Gegenstandsangemessenheit geht entsprechend über eine Passung der Methode zum untersuchten Gegenstand hinaus und kennzeichnet im Sinne „multipler Passungsverhältnisse“ „eine Abgestimmtheit von Theorie, Fragestellung, em-

pirischem Fall, Methode und Datentypen“ (Strübing et al. 2018, S. 86). Diese Passungsverhältnisse sind dabei nicht nur zu Beginn des Forschungsprozesses zu bestimmen und zu justieren, sondern fortwährend über den gesamten Prozess hinweg (ebd.). Die Justierung zwischen Empirie und Methode sollte dabei von der Empirie her erfolgen. Gemäß der „Vorfahrtsregel ‚Empirie vor Methode‘“ (Strübing et al. 2018, S. 87) sind also Methoden gegebenenfalls zu modifizieren. Gegenstandsangemessenheit steht also immer auch im potenziellen Widerspruch zu (auch im Bereich qualitativer Forschung durchaus verbreiteten) dogmatischen Vorstellungen von Methodenanwendungen (ebd. sowie schon Flick 2007, S. 478).

Damit steht die Idee der Gegenstandsangemessenheit nach Ansicht von Breuer und Reichertz in „grundsätzlichem Widerspruch zu (naturwissenschaftlich inspirierten) einheitswissenschaftlichen Postulaten für Sozial- bzw. Humanwissenschaften“ (Breuer/Reichertz 2001, o. S.), da diese davon ausgehen, dass der forschende Zugriff in erster Linie aus methodologischen Überlegungen erwächst und unabhängig vom Gegenstand erfolgen kann. Dieses Methodenverständnis², das behauptet, dass es Methoden der empirischen Sozialforschung gibt, die unabhängig vom Gegenstand anwendbar sind und dabei gehaltvolle Erkenntnisse hervorbringen, wird als unangemessen kritisiert. Vielmehr gilt es, methodische Entscheidungen jenseits von individuellen und institutionellen Routinen und Vorlieben zu treffen und dabei grundsätzlich sämtliche methodischen Möglichkeiten in Erwägung zu ziehen. Gegenstandsangemessenheit lässt sich insofern „auch umfassender und über die Grenzen der Methodenfamilien hinweg begreifen, so dass in die Betrachtung der individuellen Methodeneigenschaften das gesamte Spektrum der Methoden empirischer Sozialforschung mit einbezogen und der Weg zu ihrer Kombination geebnet wird“ (Stefer 2013, S. 62).

3.2 Gegenstandsangemessenheit in der Evaluation

Das Kriterium der Gegenstandsangemessenheit ist auch in Evaluationen relevant. Dies resultiert zunächst einmal daraus, dass Evaluation als Form der angewandten Forschung dieses Kriterium aus dem Kontext der empirischen Sozialforschung „mitgebracht“ und in den eigenen Anwendungskontext übertragen hat. Entsprechend ist der Anspruch hier, dem Evaluationsgegenstand sowie den evaluativen Fragestellungen, dem Erkenntnisziel und dem Zweck der Evaluation

2 Ein solches Methodenverständnis findet sich im Bereich der Evaluation von Wirkungen zum Beispiel in Vorstellungen zum Goldstandard experimenteller Designs, die als einziger Beweis für Evidenz angesehen werden. Andere Evaluationsdesigns werden hingegen als nicht hinreichend betrachtet (so z. B. bei Kober 2017; Armborst/Walsh 2019). Siehe hierzu auch die Ausführungen von Hense in diesem Sammelband.

gerecht zu werden und „zu fragen, welche Methoden für welche Aspekte eines zu evaluierenden Prozesses geeignet sind“ (Flick 2009, S. 13). Dieser Anspruch ist im Evaluationsdiskurs breit verankert und Gegenstand einiger Veröffentlichungen.

Caspari et al. (2019, S. 69) plädieren zum Beispiel für einen Methodengebrauch, der „Probleme, Schwächen, und Grenzen einzelner Methoden“ reflektiert und für die relevanten Gegenstandsbereiche genau prüft, ob die ausgewählten Methoden gut geeignet sind. Außerdem seien die Methoden „stets in Relation zu Fragestellungen, Projektzielen und besonderen Bedingungen, denen die Evaluation unterliegt“, auszuwählen und „der potentielle Zielkonflikt zwischen methodischer Strenge einerseits und dem Nutzen, der Fairness und vor allem der Durchführbarkeit“ (ebd.) zu berücksichtigen und angemessene Entscheidungen zu treffen. In diesen Überlegungen zeigt sich, dass Gegenstandsangemessenheit in Evaluationskontexten nicht auf die Funktion als Gütekriterium empirischer Sozialforschung reduziert bleibt, sondern darüber hinausreicht und elementare Bedeutung hat.

Auch die Fachstandards der Gesellschaft für Evaluation (DeGEval 2017; erste Fassung aus dem Jahr 2001: DeGEval 2002), die grundlegende Anforderungen an die Qualität von Evaluation formulieren, enthalten in unterschiedlichsten Facetten Verweise auf Gegenstandsangemessenheit, die weit mehr umfassen als „nur“ ein Gütekriterium empirischer Sozialforschung. In seiner Analyse der Evaluationsstandards der DeGEval in der ersten Fassung aus dem Jahr 2001 im Hinblick auf das Thema Gegenstandsangemessenheit kommt Stefer (2013) zu dem Schluss, dass neben der Thematisierung als Gütekriterium in der Forschung insbesondere Fragen der Einbindung von Stakeholdern und der Effizienz zusätzliche evaluationspezifische Aspekte darstellen, die zudem noch einen besonderen Stellenwert inne haben: „Die Besonderheiten der Stakeholder und ihre Interessen stellen zusammen mit der Berücksichtigung des Verhältnisses von Aufwand, Ressourcen und Nutzen Kernkriterien zur Wahl angemessener Datenerhebungsmethoden im Kontext von Evaluationen dar“ (Stefer 2013, S. 75).

Insgesamt zeigte die auch für die überarbeitete Version der Standards (DeGEval 2017) zutreffende Analyse von Stefer, dass „13 der 25 Standards der DeGEval Bezüge zur Gegenstandsangemessenheit der Datenerhebungsmethode herstellen“ (Stefer 2013, S. 66f.). Zum Beispiel ist als Anforderung formuliert, den Gegenstand durch die eingesetzten Datenerhebungsmethoden möglichst wenig zu belasten (D1) und inklusive seiner Zusammenhänge (G1) und seines Kontextes (G2) exakt zu erfassen, die Fragestellung auf Grundlage der erhobenen Daten möglichst adäquat zu behandeln (N4) und die Methoden anhand ihrer Passung zur Fragestellung auszuwählen (G3).³ Außerdem „sollen die Zwecke

3 Die Buchstabenkürzel beschreiben dabei grundlegende Eigenschaften von Evaluationen: N steht für Nützlichkeit, D für Durchführbarkeit, F für Fairness und G für Genauigkeit (DeGEval 2017, S. 18 ff.).

der Evaluation, aus denen die Fragestellung und damit auch die methodische Vorgehensweise hervorgeht, deutlich bestimmt sein (N2)“ (Stefer 2013, S. 72) und die Interessen der untersuchten Subjekte „so weit wie möglich in die Anlage der Evaluation einbezogen werden (N1)“ (ebd.). Abgesehen davon sollen die Datenerhebungsmethoden so offen sein, „dass sie unterschiedliche Sichtweisen gleichermaßen erfassen können und somit unverzerrte Daten und Ergebnisse sicherstellen (D2), sie sollen Stärken und Schwächen des Evaluationsgegenstands gleichermaßen erfassen und abbilden (F3) sowie, ebenfalls durch ausreichende methodische Offenheit, sicherstellen, dass keine spezifische Sichtweise einer Partei übernommen, sondern unterschiedliche Positionen gleichermaßen erfasst werden (F4)“ (Stefer 2013, S. 73).

Alles in allem weisen die Anforderungen, die in den Standards formuliert werden, also deutliche und konkrete Bezüge auf Gegenstandsangemessenheit als Gütekriterium empirischer Sozialforschung auf und werben dafür, sie zu berücksichtigen. Gleichzeitig werden zahlreiche zusätzliche evaluationsspezifische Aspekte mitgeführt, die zu Fragen der Durchführbarkeit, des Nutzens, der Genauigkeit und Fairness gegenstandsangemessene Antworten fordern.

3.3 Gegenstandsangemessene Wirkungsevaluation im Feld der Prävention von Extremismus und Förderung von Demokratie

An Wirkungsevaluationen wird in doppelter Hinsicht der Anspruch der Gegenstandsangemessenheit formuliert: Zum einen als Gütekriterium empirischer Sozialforschung, zum anderen als Evaluationsstandard, der dieses Gütekriterium aufnimmt, aber auch darüber hinausreicht. Auch Wirkungsevaluationen sind also so anzulegen, „dass der Gegenstand zu seinem Recht kommt“ (Brüsemeister 2008, S. 29). Im Folgenden wird diskutiert, was Gegenstandsangemessenheit im Zusammenhang mit Wirkungsuntersuchungen im Feld der pädagogischen Prävention von Extremismus und der Förderung von Demokratie bedeuten kann und welche Fragen sich dabei stellen. Eine Wirkung wird dabei verstanden als ein „kausal auf eine Intervention beliebiger Form (wie eine Maßnahme, ein Projekt oder ein Programm) rückführbarer Effekt“ (Widmer 2012, S. 42). Dabei kann es sich um Veränderungen und/oder Stabilisierungen handeln, die im Rahmen pädagogischer Arbeit entstehen. Wie können nun gegenstandsangemessene Wirkungsevaluationen aussehen und was ist dabei zu berücksichtigen?

3.3.1 Klärung des Gegenstands

Wird bei Wirkungsevaluationen Gegenstandsangemessenheit beansprucht, so gilt es zunächst zu klären, was genau in wirkungsbezogener Perspektive der Ge-

genstand ist und welche Eigenschaften er aufweist. Relevant ist hier zunächst vor allem, unter welchen Rahmenbedingungen und durch welche Aktivitäten welche Wirkungen entstehen sollen. Für einen Großteil der im Bundesprogramm „Demokratie leben!“ geförderten Angebote gilt, dass sie pädagogisch bzw. beraterisch arbeiten. Im Hinblick auf die Betrachtung von Wirkungen bei Individuen hat das bedeutende Implikationen: Wirkungen der entsprechenden Angebote entstehen in starker Abhängigkeit von den Kontexten in „Koproduktionsprozessen“ zwischen pädagogischen Fachkräften und Adressat:innen. Die Fachdiskussion spricht deshalb seit langem in diesem Zusammenhang von einem „Technologie-defizit“ (Luhmann/Schorr 1982) pädagogischer Praxis. Gegenstandsangemessene Wirkungsuntersuchungen müssten berücksichtigen und abbilden, dass die entstehenden Wirkungen stark von Kontexten und individuellen Voraussetzungen abhängen.

Darüber hinaus ist im Rahmen der Gegenstandsklärung auch zu eruieren, welche Wirkungen der betrachteten Aktivität bei den Adressat:innen erwartbar sind, und sicherzustellen, dass der evaluierende Blick dorthin gerichtet ist. Ist dies nicht der Fall, läuft eine Evaluation zum Beispiel Gefahr, andere Wirkungen in den Blick zu nehmen als die, die eigentlich interessieren, oder auch der Praxis Wirkungslosigkeit zu attestieren, da sie tatsächlich vorhandene Wirkungen empirisch übersieht, weil sie nicht erhoben werden. Entsprechend einer Stakeholderorientierung wäre vorab im Rahmen einer eigenen Erhebung und Analyse genau zu prüfen, was die evaluierten Angebotsumsetzenden an Wirkungen bei ihren Adressat:innen überhaupt beabsichtigen bzw. was aus Sicht der Mittelgeber an Wirkungen erreicht werden soll. Diese gegenstandsbezogenen Wirkannahmen sollen im nächsten Schritt mit bereits vorliegenden theoretischen/empirischen Erkenntnissen abgeglichen werden, um in den Blick zu nehmen, ob diese Wirkannahmen der Stakeholder tragfähig sind und dem Forschungsstand zu möglichen Wirkungen entsprechen.

Im Ergebnis der Analyse zu den verschiedenen Wirkhorizonten muss festgelegt und begründet werden, welcher „gegenstandsangemessene“ Wirkungsausschnitt empirisch beleuchtet wird. In den in diesem Band vorgestellten Praxisbeispielen wurden beispielsweise zumeist programmtheoriebasierte Ansätze genutzt, teilweise in Verbindung mit Wirkmodellen, um entsprechende Wirkausschnitte zu bestimmen, Wirklogiken zu plausibilisieren und auf dieser Basis einen gegenstandsangemessenen methodischen Zugriff zu entwickeln. Unter Wirkungsmodellen, die gewissermaßen eine Gegenstandsklärung in Modellform darstellen, werden dabei Beschreibungen der Verbindungen zwischen Aktivitäten einerseits und den (angestrebten) Wirkungen andererseits verstanden (für eine ausführliche Darstellung siehe den Beitrag von Hense in diesem Band).

Es empfiehlt sich, Wirkmodellierungen mit verschiedenen Projektumsetzenden durchzuführen, aber auch die Perspektive anderer Stakeholder einzubinden.

Hierdurch kann das Risiko verringert werden, dass Wirkungen „übersehen“ oder unrealistische Annahmen über zu erzielende Wirkungen oder deren Zustandekommen prägend werden. Dies erweist sich als äußerst wichtig, denn es gilt, im Rahmen einer zielorientierten Evaluation Angebote und Programme danach zu bewerten, was sie leisten sollen, was sie leisten wollen und was sie realistisch auch tun können.

Ein Bundesmodellprogramm wie „Demokratie leben!“ mit zugegebenermaßen vielen, aber mit Blick auf das gesamte Bundesgebiet dennoch eher punktuellen Aktivitäten beispielsweise auf seine Breitenwirksamkeit insgesamt hin zu prüfen, führte entsprechend in die Irre. Ähnliches gilt zum Beispiel, wenn bei einer Teilnahme von Jugendlichen an einem Workshop nachhaltige Effekte auf die individuellen Einstellungen erwartet werden, die innerhalb kurzzeitpädagogischer Angebote aber realistischerweise eher selten auftreten. Solche Wirkannahmen, zum Beispiel in Richtung einer durch Programmaktivitäten erwarteten sichtbaren Abnahme von Demokratieskepsis in der Allgemeinbevölkerung, wären derart überhöht, dass man im Prinzip schon vor der empirischen Untersuchung mit einiger Wahrscheinlichkeit davon ausgehen kann, dass man relative Wirkungslosigkeit feststellt. Entsprechende Wirkungserwartungen sind aber nicht selten, auch weil die einschlägigen Bundesprogramme immer auch über entsprechende „Wirkversprechen“ legitimiert wurden und werden.⁴ Solche Effekte wären aber wenn überhaupt nur dann erwartbar, wenn die im Bundesprogramm erprobten und als wirksam identifizierten Maßnahmen in der Breite umgesetzt würden. Das wiederum würde voraussetzen, dass sie in Regelstrukturen der Kinder- und Jugendhilfe oder auch im schulischen Bereich flächendeckend aufgenommen und umgesetzt werden.

3.3.2 Gegenstandsangemessene Fragestellungen

Ebenfalls im Rahmen von Wirkungsuntersuchungen zu prüfen ist, ob die verfolgten Fragestellungen gegenstandsangemessen sind. In den DeGEval-Standards wird dafür geworben, die Fragestellungen grundsätzlich auch von den Interessen der Stakeholder abhängig zu machen und sie durch diesen Einbezug der Perspektive von gegenstandskundigen Beteiligten letztlich auch gegenstandsangemessener zu gestalten. In Bezug auf Wirkungsfragen sind hier zunächst basale

4 So sprach die damalige zuständige Ministerin Manuela Schwesig 2015 in einem Interview davon, dass das Bundesprogramm „Demokratie leben!“ Angebote fördert, „damit sich die Jugendlichen gar nicht erst radikalisieren und in die Hände des IS geraten“ (Schwesig 2015). Auch die aktuell zuständige Ministerin Lisa Paus referiert Befunde des Bundeskriminalamtes, die die Wirksamkeit des Programms darüber nachgewiesen haben, dass es in „Regionen mit mehr Präventionsangeboten weniger Fälle von politisch motivierter Kriminalität“ (Paus 2024) gäbe.

Fragen zu klären, die von der Ausrichtung und dem Erkenntnisinteresse der Evaluation abhängen: Soll es innerhalb der Evaluation darum gehen, Wirkungen nachzuvollziehen (durch Nachweise oder Plausibilisierungen), um im Rahmen einer summativen Evaluation die geförderten Angebote zu bewerten? Oder soll im Rahmen formativer Evaluationen für fachliche Weiterentwicklungen nützliches Wissen dazu generiert werden, *warum* bestimmte Wirkungen einsetzen und in welchen Kontexten sie auftreten? Oder sollen unabhängig von summativer oder formativer Ausrichtung besonders die unintendierten Wirkungen in den Blick genommen werden, um bereits bestehendes Wissen zu den eintretenden erwünschten Wirkungen zu ergänzen?

Je nach Ausrichtung und Erkenntnisinteresse der Evaluation kommen unterschiedliche methodische Designs infrage, wobei festgehalten werden kann, dass qualitative Verfahren dann besonders geeignet erscheinen, solange über die Wirkung, die sie vermittelnde Wirkmechanismen und -kontexte wenig bekannt ist (vgl. die Beiträge von Rehse/Hemmann/Braun und Figlesthler/Haase/Schau in diesem Band). Quantitative Verfahren in Form von Randomized-Controlled-Trial-Designs (RCT)⁵, quasi-experimenteller Designs und Vorher-Nachher-Studien haben hingegen den Vorteil, im statistischen Sinne Wirkungen dann nachweisen zu können, wenn beispielsweise bereits klare Vorstellungen zur Wirkweise einer Maßnahme vorliegen oder wenn die Wirkungen einer Maßnahme im Vergleich zu anderen Maßnahmen oder Entwicklungen ohne begleitende Maßnahme untersucht werden sollen. Gleichzeitig haben diese Designs wie auch Vorher-Nachher-Befragungen jedoch den Nachteil, dass sie zwar Aussagen darüber treffen können, ob die erwünschten Wirkungen eingetreten sind, nicht jedoch, warum sie eintreten und welche Wirkbedingungen hierfür verantwortlich sind (siehe hierzu auch die Ausführungen von Hense zu Grenzen experimenteller Ansätze der Wirkungsevaluation in diesem Sammelband). Erkenntnisse hierzu können wiederum durch die Erhebung der Einschätzungen von eingetretenen Wirkungen bzw. des kontrafaktischen Zustands bei Adressat:innen oder relevanten Dritten gewonnen werden.

3.3.3 Gegenstandsangemessene Methoden

Im Zusammenspiel mit der Fragestellung muss außerdem geprüft werden, inwiefern bestimmte methodische Zugriffe zum Gegenstand passen. Um Wirkungen zu evaluieren, kann auf verschiedenste Methoden der empirischen Sozialfor-

5 In den Evaluationen in den Themenfeldern ist die im Rahmen von RCT nötige zufällige Zuordnung zu Gruppen schwierig und kaum praktikabel. Stattdessen wird im Rahmen sogenannter quasi-experimenteller Designs auf bereits vorhandene natürliche Gruppen, wie zum Beispiel Schulklassen, zurückgegriffen.

schung zurückgegriffen werden, sowohl quantitativ als auch qualitativ, und es ist genau zu prüfen, welche Methoden am besten zu den wirkungsbezogenen Eigenschaften des Gegenstandes passen, also geeignet sind, die (erwünschten) Wirkungen in den Blick zu bekommen. Um eine gegenstandsangemessene methodische Auswahl zu treffen, müssen zunächst die Voraussetzungen geprüft werden, die der Einsatz der jeweiligen Methode erfordert, bzw. sind die Implikationen zu berücksichtigen, die in der Praxis durch die Methode entstehen.

Grundsätzlich gilt für jegliche Wirkungsanalysen, dass sie nur dann umgesetzt werden können, wenn die Adressat:innen sowie die Umsetzenden eines Angebots überhaupt auskunftsbereit sind und Erhebungen offen gegenüber stehen. Erfahrungsgemäß ist diese *Offenheit* in bestimmten pädagogischen Settings eingeschränkt, z. B. wenn der pädagogische Raum als „geschützter Raum“ konzipiert ist und/oder intensivere Formen der Einzelfallarbeit stattfinden. In diesen Fällen wäre zunächst zu prüfen, ob und unter welchen Einschränkungen hier Wirkungsanalysen stattfinden können, da die Gefahr besteht, dass entsprechende empirische Erhebungen das pädagogische Setting empfindlich stören und im Grunde eine Intervention darstellen, die selbst Wirkungen hervorbringt. Insofern besteht das Risiko, dass solche Untersuchungen nicht mehr die Wirkung der eigentlich zu evaluierenden Aktivität in den Blick nehmen, sondern die Wirkung einer durch die Erhebung beeinflussten und gestörten Aktivität.

Eine weitere grundsätzliche Voraussetzung für Wirkungsanalysen ist ein Mindestmaß an *Routinehaftigkeit* der untersuchten Aktivität. Wenn nicht klar genug zu bestimmen ist, worin die Aktivität besteht, erübrigt es sich auch, die Wirkung einer solchen Aktivität zu ermitteln. Für den Gegenstandsbereich ist zu berücksichtigen, dass Standardisierung im Bereich pädagogischer und Sozialer Arbeit eher selten der Fall ist, weil diese Arbeit „nur sehr bedingt standardisierbar und manualisierbar [ist]. Ihr Setting ist vielfach – etwa in der Offenen und aufsuchenden Kinder- und Jugendarbeit – seitens der Fachkräfte faktisch wenig planbar oder bis in alle Einzelheiten strukturierbar. Ihre Konstellationen sind im Allgemeinen divers und dynamisch. Adressatinnen bzw. Adressaten und Fachkräfte konstruieren sie als Ko-Produzentinnen und -produzenten“ (Möller u. a. 2020, S. 411). Hinzu kommt, dass z. B. die geförderten Modellprojekte dazu aufgerufen sind, innovative Vorgehensweisen zu erproben. Daher ist davon auszugehen, dass sie mindestens in Teilen und mindestens am Anfang der Förderung (noch) keine hinreichenden Routinen entwickelt haben und sich in Suchbewegungen befinden (siehe die Ausführungen im Beitrag von Euteneuer/Rüger in diesem Band). Entsprechend wäre vor einer Wirkungsanalyse zu prüfen, ob eine gewisse „Reife“ der Umsetzung gegeben ist bzw. zumindest der Kern einer Aktivität feststeht und entsprechend analysiert werden kann.

Erfahrungsgemäß zeigt sich abseits dieses Kerns bezüglich der (inhaltlichen und methodischen) Flexibilität der Angebote eine große Bandbreite: Von Maßnahmen, die inhaltlich und methodisch stark manualisiert vorgehen, bis hin zu

Angeboten, die ihren Adressat:innen breite Partizipationsmöglichkeiten bieten und sie zur Mitgestaltung von Themen, Inhalten und/oder Methoden einladen. Bei Ersteren sind die umgesetzten Aktivitäten inhaltlich und methodisch stark standardisiert, während Letztere durch die aktive Mitgestaltung der Adressat:innen unterschiedlichste Gestalt annehmen können. Für solche wenig standardisierten Aktivitäten sind qualitative Verfahren besonders geeignet (Flick 2007), da sie offener gestaltet sind und dabei helfen können, überhaupt zu erfassen, was der jeweils wirksame Kern der Aktivitäten ist.

Neben den bislang angesprochenen generellen Voraussetzungen für Wirkungsanalysen gibt es insbesondere mit Blick auf quantitative Verfahren der Wirkungsanalyse noch weitere Bedingungen, die erfüllt sein müssen, damit die erhofften Erkenntnisse gewonnen werden können. So sollte idealerweise eine hinreichend große Grundgesamtheit (N) gegeben sein, damit die im sozialwissenschaftlichen Bereich üblichen Quantifizierungen und statistischen Operationen überhaupt sinnvoll sind. Dies trifft im Kontext des hier betrachteten Bundesprogramms nur für einen Teil der geförderten Angebote zu, und zwar vor allem jene, die im Rahmen von kurzzeitpädagogischen Angeboten regelmäßig mit unterschiedlichen Gruppen (z. B. Schulklassen) arbeiten. Im Rahmen langzeitpädagogischer Gruppen- oder Einzelfallarbeit dürften hinreichende Fallzahlen hingegen kaum zusammenkommen. Im Fall quasi-experimenteller Designs stellt sich zusätzlich noch die Schwierigkeit, dass eine mit der Experimentalgruppe vergleichbare Kontroll- bzw. Wartegruppe gefunden werden muss, die das Angebot nicht oder verzögert wahrnimmt. Bei Angeboten in schulischen Kontexten kann vergleichsweise einfach auf andere Schulklassen als Kontroll- oder Wartegruppen zurückgegriffen werden. In anderen Kontexten, z. B. bei außerschulischen Angeboten im Rahmen von Kinder- und Jugendarbeit, die im Unterschied zu schulischen Kontext zumeist auf einer freiwilligen Teilnahme basieren, stellt dies hingegen eine bedeutende Hürde dar.

Darüber hinaus spielt auch die Komplexität des Gegenstandes bei der Wahl von Methoden eine Rolle. Im hier behandelten Praxisfeld werden von den Angeboten häufiger mehrdimensionale Problemstellungen adressiert und entsprechend vielschichtige Aktivitäten umgesetzt, die häufig mehrere Ziele gleichzeitig verfolgen. Dies gilt nicht nur für die Evaluation von pädagogischen Maßnahmen und Projekten, sondern auch und erst recht für Maßnahmenbündel, wenn ein ganzes Bundesprogramm wirkungsevaluiert werden soll (siehe Beitrag von König/Stärck/Zierold in diesem Band). Um diesen Eigenschaften Rechnung zu tragen, empfehlen sich bestimmte methodische Designs: „Multimethodik und Multiperspektivität können [...] sowohl wegen der Eigenschaften des Evaluationsgegenstands selbst (beispielsweise aufgrund seiner Komplexität) gegenstandsangemessen sein als auch wegen entsprechender Anforderungen des Evaluationsauftrags (beispielsweise, wenn ein möglichst facettenreiches und vollständiges Bild der Arbeit gewonnen werden soll)“ (Klöckner et al. 2022, S. 2). Mit dem Begriff multi-

methodisch wird dabei der „Einsatz mehrerer unterschiedlicher Methoden“ (ebd.) und die Nutzung verschiedener Datenquellen bezeichnet, als Multiperspektivität die Erhebung der Perspektiven verschiedener Personen, die an der Umsetzung des Programms oder Projekts beteiligt sind.

Der methodische Anspruch an Gegenstandsangemessenheit umfasst auch Fragen des angemessenen Zugangs zu den untersuchten Subjekten und einer angemessenen Erfassung ihrer Perspektive. Hierzu gehört auch die Berücksichtigung der auf ihrer Seite vorhandenen Kompetenzen. Berücksichtigt werden müssen hier nach Stefer z. B. die „Sprachkompetenzen der untersuchten Subjekte oder Vorbehalte gegenüber den Forschenden“ (2013, S. 65), wie sie gerade in Evaluationskontexten sicher durchaus verbreitet sind. Gerade im Feld der Beratung bestehen z. B. Vorbehalte gegenüber Wirkungsuntersuchungen, da zwischen Beratungsnehmer:innen und Berater:innen ein sensibles Vertrauensverhältnis herrscht, das durch die Einbindung Dritter gefährdet wird. Hinzu kommt, dass bei einem Teil der Adressat:innen der Angebote nicht immer ausreichend deutsche Sprachkenntnisse vorhanden sind, um an Erhebungen teilzunehmen. Hier braucht es also Anpassungen, z. B. in Form mehrsprachiger Fragebögen, oder mit Blick auf Vorbehalte gegenüber den Evaluierenden intensivere Phasen des Vertrauensaufbaus.

3.3.4 Gegenstandsangemessene Ressourcen

Ein weiterer wesentlicher Aspekt von Gegenstandsangemessenheit besteht darin, das evaluative Vorgehen ins Verhältnis zu den vorhandenen Ressourcen zu setzen. Wirkungsuntersuchungen generell und insbesondere multiperspektivische, qualitative Untersuchungen sind enorm ressourcenintensiv, sodass sich die Frage aufdrängt, ob die in Aussicht stehenden Erkenntnisse den Ressourceneinsatz rechtfertigen. Dies gilt einerseits in Bezug auf die Gruppe der Forschenden, andererseits in Bezug auf den Forschungsgegenstand und die zu Untersuchenden. Für Letztere bedeutet eine Datenerhebung in der Regel Zusatzaufwand. Der Aufwand muss mit den zur Verfügung stehenden Ressourcen – maßgeblich Geld und Zeit – und dem erwarteten Nutzen in einem sinnvollen Verhältnis stehen. Ein im Rahmen der „Programmevaluation Demokratie leben!“ genutzter Ansatz war, eine gezielte Auswahl an Einzelfällen im Sinne einer Samplebildung vorzunehmen, um exemplarisch zentralen Interventions- und Wirklogiken empirisch nachzuspüren. In den Beiträgen dieses Sammelbandes findet sich eine solche Auswahl in unterschiedlichen Varianten, beispielsweise als Regionenauswahl bei König, Stärck und Zierold oder als Auswahl von Beratungsfällen bei Faglestahler, Haase und Schau. So können trotz begrenzter Ressourcen zumindest Aussagen zur prinzipiellen Wirkfähigkeit und zu zentralen Wirkmechanismen und Einflussfaktoren gewonnen werden.

3.4 Fazit

Gegenstandsangemessenheit stellt für Wirkungsevaluationen einen wesentlichen Anspruch dar, der nicht nur ein wichtiges Gütekriterium empirischer Sozialforschung ist, sondern ebenfalls seitens der Fachgesellschaft (DeGEval 2017) als Richtschnur professionellen Handelns in der Evaluation formuliert wurde. Entsprechend gilt es, eine Passung zwischen Theorie, Fragestellung, Methode und dem Gegenstand der Forschung bzw. Evaluation sicherzustellen. Gleichzeitig sind Merkmale und Interessen der Stakeholder der Evaluation zu berücksichtigen und ein ausgewogenes Verhältnis zwischen Aufwand, Ressourcen und Nutzen herzustellen. Bereits die schlichte Anzahl der hier genannten und zu berücksichtigenden Aspekte legt nahe, dass es im Kern bei der Gegenstandsangemessenheit darum geht, die verschiedenen Ebenen auszubalancieren und Einseitigkeiten zu vermeiden. Insbesondere im Fall von widersprüchlichen Anforderungen (z. B., wenn Interessen der Stakeholder auseinanderfallen) ist es dabei auch eine Option, bestimmte Prioritäten zu setzen und begründete Entscheidungen zu treffen.

Dies gilt auch und gerade für Wirkungsevaluationen, denn die Evaluation von Wirkungen ist einerseits mit besonders starken Erwartungen in Bezug auf Erkenntnisse und die erhoffte Legitimation durch den Beleg von Wirksamkeit verbunden, andererseits aber fachlich höchst herausforderungsvoll. Insofern ist hier eine gegenstandsangemessene Evaluationspraxis in besonderem Maße geboten. Bei allen legitimen Abwägungsprozessen gibt es jedoch eine grundsätzliche und nicht zu verhandelnde Grundanforderung: Sicherzustellen ist, dass der methodische Zugriff geeignet ist, die Wirkebene zu beleuchten, auf der die von den Maßnahmen beabsichtigten und realistisch zu erwartenden Wirkungen liegen, zum Beispiel demokratiebezogene Haltungen. Die Grundlage hierfür ist zunächst eine genaue Klärung, wodurch sich der Gegenstand in einer Wirkungsperspektive überhaupt auszeichnet, zum Beispiel in Form von Wirkungsmodellierungen. Darüber hinaus wäre zu prüfen, ob und inwiefern die Voraussetzungen für die Anwendung wirkungsbezogener Evaluationsdesigns gegeben sind. Nur auf dieser Grundlage können Erkenntnisse entstehen, die dem Gegenstand gerecht werden.

Literatur

- Armbrorst, A./Walsh, M. (2019): WESPE – Wissenstransfer für Ergebnisse aus Evaluationsstudien. In: Zeitschrift für Evaluation, 18. Jg., H. 1, S. 140–149.
- Breuer, F./Reichert, J. (2001): Standards of Social Research. In: Forum Qualitative Sozialforschung, 2. Jg., H. 3. <https://doi.org/10.17169/fqs-2.3.919>.
- Brüsemeister, T. (2008): Qualitative Forschung. Ein Überblick. 2. Auflage. Wiesbaden: Westdeutscher.

- Caspari, A./Erzberger, C./Kelle, U./Polak, J. T. (2019): Die „Standards für Evaluation“ aus methodischer Perspektive. In: Hense, J. U./Böttcher, W./Kalman, M./Meyer, W. (Hrsg.): Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis? Münster: Waxmann, S. 53–71.
- DeGEval – Gesellschaft für Evaluation (2002): Standards für Evaluation. Mainz: DeGEval – Gesellschaft für Evaluation.
- DeGEval – Gesellschaft für Evaluation (2017): Standards für Evaluation. Erste Revision 2016. https://www.degeval.org/fileadmin/content/Z03_Publikationen/DeGEval-Standards/DeGEval-Standards_fuer_Evaluation.pdf (Abruf 27.06.2024).
- Flick, U. (2007): Qualitative Sozialforschung. Eine Einführung. Reinbek bei Hamburg: Rowohlt.
- Flick, U. (2009): Qualitative Methoden in der Evaluationsforschung. In: Zeitschrift für Qualitative Forschung, 10. Jg., H. 1, S. 9–18.
- Klößner, M./Verhovnik-Heinze, M./Schlicht-Schmälzle, R./Nakamura, R./Strunk, J. (2022): Multimethodische Evaluationsdesigns: eine Erhebung der sozialwissenschaftlichen Praxis. (PRIF Reports 1). Frankfurt am Main: Hessische Stiftung Friedens- und Konfliktforschung.
- Kober, M. (2017): Zur Evaluation von Maßnahmen der Prävention von religiöser Radikalisierung in Deutschland. In: Journal for Deradicalization, H. 11, S. 219–257.
- Kromrey, H. (2006): Empirische Sozialforschung, 11. Auflage. Stuttgart: Lucius & Lucius.
- Luhmann, N./Schorr, K.-E. (1982): Das Technologiedefizit der Erziehung und die Pädagogik. In: Luhmann, N./Schorr, K.-E. (Hrsg.): Zwischen Technologie und Selbstreferenz: Fragen an die Pädagogik. Frankfurt am Main: Suhrkamp, S. 11–41.
- Möller, K./Buschbom, J./Pfeiffer, T. (2020): Zur Evaluation von Praxisansätzen der Extremismusprävention. In: Ben Slama, B./Kemmensies, U. E. (Hrsg.): Handbuch Extremismusprävention. Gesamtgesellschaftlich. Phänomenübergreifend. Wiesbaden: Bundeskriminalamt, S. 389–424. https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/PolizeiUndForschung/1_54_HandbuchExtremismuspraevention.html?nn=125568 (Abruf 10.06.2024).
- Nover, S. U. (2022): Im Fokus: Gegenstandsangemessenheit. In: Nover, S. U. (Hrsg.): Forschungswege in der Pflegewissenschaft 2. Die Rolle des Designs. Wiesbaden: Springer, S. 1–17.
- Paus, L. (2024): Wir haben eine lebendige und wehrhafte Demokratie. (27.05.2024). Interview Neue Osnabrücker Zeitung. BMFSFJ. <https://www.bmfsfj.de/bmfsfj/aktuelles/reden-und-interviews/lisa-paus-wir-haben-eine-lebendige-und-wehrhafte-demokratie-240346> (Abruf 24.06.2024).
- Schwesig, M. (2015): Über Terrorgefahr und Extremismusprävention. (23.11.2015). Interview mit der Bild-Zeitung. BMFSFJ. <https://www.bmfsfj.de/bmfsfj/aktuelles/reden-und-interviews/manuela-schwesig-ueber-terrorgefahr-und-extremismuspraevention-101556> (Abruf 24.06.2024).
- Stefer, C. (2013): Die Gegenstandsangemessenheit empirischer Datenerhebungsmethoden im Kontext von Lehrevaluationen an Hochschulen. Dissertation. Marburg: Philipps-Universität Marburg/Lahn. archiv.ub.uni-marburg.de/diss/z2013/0727/pdf/dcs.pdf (Abruf 10.06.2024).
- Strübing, J./Hirschauer, S./Ayaß, R./Kränke, U./Scheffer, T. (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. In: Zeitschrift für Soziologie, 47. Jg., H. 2, S. 83–100.
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./Lobermeier, O./Heitmeyer, W. (Hrsg.): Evaluation von Programmen und Projekten für eine demokratische Kultur. Wiesbaden: Springer VS, S. 41–68.

III Evaluation von Wirkungen im Bundesprogramm „Demokratie leben!“

4 Wirkungsorientierte Evaluation von Beratung nach der Realist Evaluation

Zur Erschließung von Wirkmechanismen in der Beratung von Betroffenen rechter, rassistischer und antisemitischer Gewalt und in der Ausstiegs- und Distanzierungsberatung

Carmen Figlestahler, Katrin Haase und Katja Schau

„What Is This Thing Called a Mechanism?“ (Lemire et al. 2020) ist eine Frage, die sich wohl viele Forschende stellen (müssen), die sich in ihrer Forschungsarbeit am Ansatz der Realist Evaluation orientieren. Die Realist Evaluation stellt ein „heuristisches Rahmenwerk“ (Haunberger/Baumgartner 2017, S. 139) in der wirkungsorientierten Evaluation dar. Der Wirkmechanismus ist darin ein zentrales Konzept: Mit ihm soll beschrieben werden, wie Programme oder Maßnahmen Wirkungen in verschiedenen Kontexten hervorbringen. Im Diskurs werden die Mehrdeutigkeit des Begriffs und die unterschiedlichen Verständnisse in Forschungsarbeiten mitunter problematisiert (Lemire et al. 2020; Werthern 2020). So ist die Kernargumentation, es brauche für glaubwürdige Wirkungsuntersuchungen nach der Realist Evaluation ein geteiltes konzeptionelles Verständnis vom Mechanismus (Lemire et al. 2020, S. 74). Diese Positionierung und die inhaltliche Klärungsbedürftigkeit von Mechanismen sind der Ausgangspunkt des vorliegenden Beitrags. Wir werden eine eigene, professionstheoretisch inspirierte Konzeptualisierung von Wirkmechanismen vorstellen und kritisch diskutieren. In diesem Zusammenhang positionieren wir uns abschließend und werben dafür, die konzeptionelle Offenheit produktiv für gegenstandsangemessene, theoriebasierte Klärungen zu nutzen.

Empirische Basis des Beitrags sind die in den Jahren 2021 bis 2023 umgesetzten Wirkungsuntersuchungen der wissenschaftlichen Begleitung im Handlungsbereich Land des Bundesprogramms „Demokratie leben!“ (Haase/Rottinghaus 2024; Figlestahler/Schau 2023). In diesem Handlungsbereich werden in den Bundesländern unter anderem Angebote der Beratung von Betroffenen rechter, rassistischer und antisemitischer Gewalt sowie der Ausstiegs- und Distanzierungsberatung im Bereich Rechtsextremismus vorgehalten. Die

zivilgesellschaftlichen Opferberatungsstellen¹ beraten und unterstützen Betroffene bei der Bewältigung der (im-)materiellen Folgen von rechten, rassistischen und antisemitischen Gewalttaten und bei der Wahrnehmung und Durchsetzung ihrer Rechte. Die zivilgesellschaftlichen Angebote der Ausstiegs- und Distanzierungsberatung unterstützen Personen bei der Distanzierung von extrem rechten Szenen und Ideologiebezügen sowie bei der Hinwendung zu einer Lebensweise, die mit demokratischen und pluralistischen Grundwerten vereinbar ist. Die Beratungsfelder lassen sich der Sozialen Arbeit zuordnen.

Im vorliegenden Beitrag erläutern wir zunächst – ausgehend von dem heuristischen Rahmenwerk und Kausalitätsverständnis der Realist Evaluation – das Konzept von Wirkmechanismen (Abschnitt 4.1). Daran anschließend skizzieren wir einen Vorschlag, wie ein gegenstandsangemessenes, professionstheoretisch inspiriertes Verständnis von Wirkmechanismen aussehen kann: Wirkungsermöglichende Mechanismen fassen wir für den vorliegenden Gegenstand von Beratung und Unterstützung im Kontext Sozialer Arbeit als haltungsbasierte Handlungen (Abschnitt 4.2.1). Am empirischen Material beschreiben wir je einen ausgewählten Mechanismus, um unsere abstrakte Definition und unser methodisches Vorgehen nachvollziehbar zu machen (Abschnitt 4.2.2). Wir kontrastieren die eingesetzten Haltungen in den jeweiligen Arbeitsfeldern und arbeiten dadurch die Eigenlogik des jeweils professionellen Handelns heraus (Abschnitt 4.2.3). Abschließend diskutieren wir Potenziale und Grenzen dieses spezifischen Verständnisses von Wirkmechanismen (Abschnitt 4.3). Anliegen des Beitrags ist es, die empirische Erforschung bzw. Erforschbarkeit von Wirkmechanismen anwendungsorientiert zu reflektieren und damit die Diskussion um Wirkungsevaluationen in der Sozialen Arbeit, insbesondere im Bereich Demokratieförderung und Extremismusprävention, anzureichern.

4.1 Realist Evaluation als kontextsensible Wirkungsanalyse

Bei der Realist Evaluation handelt es sich um einen Ansatz, der sich der theoriebasierten Evaluation zuordnen lässt (Werthern 2020, S. 173). Realist Evaluation betont, dass die Programmtheorien in ihren realen Kontexten und komplexen Wirkweisen betrachtet werden müssen (Pawson/Tilley 2004, S. 4). Empirisch untersucht werden Wirkungszusammenhänge, die an den tatsächlichen sozialen Lebenszusammenhängen ansetzen. Mit der Realist Evaluation wird so auch beschreibbar, wie und warum Maßnahmen bei verschiedenen Adressat:innen unterschiedlich wirken (Haunberger/Baumgartner 2017, S. 125 ff.) und weshalb Veränderungen jeweils unterschiedlich nachhaltig sind (Pawson/Tilley 2004, S. 4).

1 In diesem Beitrag wird sowohl der Begriff Opferberatung als auch Betroffenenberatung für die Beratung von Betroffenen rechter, rassistischer und antisemitischer Gewalt genutzt.

Mit dem Erkunden und Erschließen von Wirkzusammenhängen kreist die Forschung um kausale Verständnisse. Wirkungen sind Thomas Widmer zufolge ein „kausal auf eine Intervention beliebiger Form (wie eine Maßnahme, ein Projekt oder ein Programm) rückführbarer Effekt“ (2012, S. 42). Diese kausale Zuschreibung wird als zentral angesehen, um intendierte Entwicklungen von zufälligen und parallel entstandenen Veränderungen zu unterscheiden. Bernd Dollinger (2024, S. 13) stellt heraus, dass Kausalität, die immer auf Wahrnehmungen und Zuschreibungen basiert, perspektivenabhängig ist. In der Realist Evaluation werden die komplexen Bedingungen anerkannt und untersucht, unter denen spezifische Maßnahmen kontextabhängige Wirkung entfalten (Dollinger 2018, S. 252; Pawson/Tilley 2004, S. 10). Diese Evaluationsrichtung zielt auf die „theoretisch informierte Klärung generativer Mechanismen“ (Dollinger 2018, S. 252), das heißt, sie stellt wirkungsermöglichende Prozesse heraus. Die generative Erklärungslogik sieht vor, den Weg von der Ursache zu einer Wirkung aufzuschlüsseln. Damit ist das Anliegen verbunden, komplexe und multiple Ursache-Wirkungs-Beziehungen zu verstehen. Wirkungsnachweise im engeren Sinne können mit dem Ansatz nicht erbracht werden.²

Der Ansatz der Realist Evaluation wurde für die Studien gewählt, weil er für wirkungsorientierte Analysen Sozialer Arbeit, die sich durch ihren Koproduktionscharakter auszeichnet, besonders geeignet ist (Nielsen/Lemire/Tangsig 2022; Haunberger/Baumgartner 2017). Mit dem Ansatz kann die zentrale Relevanz von Kontext- und Subjektbezüge in der Praxis Sozialer Arbeit angemessen berücksichtigt werden. Wir wollten gerade das Zusammenwirken jener Faktoren plausibel beschreiben, welche die Berater:innen und Adressat:innen in die spezialisierte Betroffenenberatung sowie die Ausstiegs- und Distanzierungsberatung einbringen, und die in komplexer Weise Wirkungen hervorbringen. Wir analysierten hierfür die standortgebundenen Wirkungszuschreibungen und fragten danach, was sich aus der Perspektive der Berater:innen und Adressat:innen wie und in welchem Kontext als Wirkung der Beratung darstellt. So konnten wir ausgewählte Wirkungszusammenhänge empirisch erschließen und Gelingensbedingungen für wirksame Beratung und Unterstützung ableiten.

Das analytische Gerüst der Untersuchungen bildeten entsprechend der Realist Evaluation Kontext-Mechanismus-Outcome (KMO)-Konfigurationen, mit denen Wirkungszusammenhänge erklärbar werden sollen (Pawson/Tilley

2 In Wirkungsuntersuchungen sind grundlegende methodische und theoretische Entscheidungen folgenreich: In den Diskursen um Wirksamkeit (vgl. Albus/Micheel/Polutta 2018, S. 1566 ff.; Armbrorst u. a. 2018, S. 6; Stockmann 2006) gelten (randomisierte) kontrollgruppenbasierte Untersuchungen mit Vorher- und Nachermessungen zur Wirkungsfeststellung teilweise als Goldstandard. Dies wurde im Rahmen von Debatten über Möglichkeiten und Grenzen von Wirksamkeitsuntersuchungen in der Extremismusprävention kontrovers diskutiert (vgl. Möller/Buschbom/Pfeiffer 2020, S. 407 ff.).

2004, S. 9). Outcomes umfassen bei Ray Pawson und Nick Tilley intendierte und unintendierte Effekte. Wir verstehen sie im Beratungszusammenhang als Veränderungen und Stabilisierungen von Handlungsfähigkeiten, Einstellungen und Wissensbeständen bei den Adressat:innen, die als Ergebnis der sozialen Beratungsinteraktionen zugeschrieben werden. Auf die Konzipierung von Wirkmechanismen als haltungsbasierte Handlungen wird im folgenden Abschnitt näher eingegangen, in dem wir unser Verständnis auch in Professionalitätsdiskurse der Sozialen Arbeit und Evaluationsdebatten einbetten. Bei den Kontexten nahmen wir insbesondere die Ressourcen der Handelnden in den Blick. Wirkungen bei Adressat:innen entstehen in der Beratungsarbeit grundsätzlich durch die Interaktionen zwischen Fachkräften und Adressat:innen in tragfähigen Arbeitsbündnissen. Dies hat zur Folge, dass die Ressourcen der Handelnden (Wissensbestände, Handlungsfähigkeiten), die sie in den Prozess einbringen, wichtige Kontextfaktoren darstellen. Außerdem berücksichtigten wir sozialräumliche, rechtliche und gesellschaftliche Kontexte, in die die Fallkonstellationen jeweils spezifisch eingelassen sind.³

4.2 Professionstheoretisch inspirierte Konzeption und Rekonstruktion von Wirkmechanismen

Im Folgenden wird der konzeptionelle Zuschnitt der Wirkmechanismen in den Untersuchungen zur Opferberatung sowie zur Ausstiegs- und Distanzierungsberatung erläutert und begründet. Anschließend konkretisieren wir anhand unserer empirischen Befunde zunächst separat für die Ausstiegs- und Distanzierungsberatung sowie für die Opferberatung ausgewählte Wirkmechanismen und stellen dar, wie wir diese erschlossen haben. Anschließend verschränken wir die Betrachtung beider Arbeitsfelder, um die spezifischen Potenziale unseres Konzeptvorschlags mit Blick auf die Reflexion professioneller Kernprobleme zu illustrieren.

3 In den Begriffen des Logischen Modells ausgedrückt handelt es sich hier um Incomes und Teile der Inputs. Incomes sind jene Ressourcen, die die Adressat:innen in den Beratungsprozess einbringen. Dies umfasst Vorwissen, bisherige Kompetenzen und Einstellungen sowie soziale Merkmale wie Bildungsgrad und Alter. Aus den vielfältigen Inputs im Logischen Modell werden hier vor allem die personellen und fachlichen Ressourcen als Kontexte berücksichtigt, die die Berater:innen einbringen. Diese stellen die Voraussetzung für professionelles haltungsbasiertes Handeln und damit für Wirkungsermöglichung dar und beziehen sich zum Beispiel auf Qualifikationen und Erfahrungen der Berater:innen.

4.2.1 Konzeption von Wirkmechanismen als haltungsbasierte Handlungen

Folgt man den Prämissen der Realist Evaluation, dann ist das Konzept der Wirkmechanismen grundlegend (Kita et al. 2022, S. 74; Haunberger/Baumgartner 2017, S. 125). Ray Pawson und Nick Tilley definieren Wirkmechanismen als (unsichtbare) Prozesse, die ermöglichen, dass die Beteiligten Ressourcen für die Arbeit an der Veränderung einbringen. Sie sind die Basis für mögliche Veränderungen und Stabilisierungen (Outcomes) (2004, S. 6). In dieser Definition wird deutlich, dass den Wirkmechanismen eine generative Kausalkraft zugeschrieben wird: Sie sind also wirkungsermöglichend, nicht jedoch zwingend wirksam.⁴ Zuzanna Kita et al. konkretisieren Mechanismen als „soziale oder psychologische Triebkräfte“ (2022, S. 75), die Veränderungen verursachen. Aber auf was diese wirkungsermöglichenden Prozesse oder Triebkräfte fokussieren und wie sie qualifiziert werden können, wird in den Grundlagentexten von Ray Pawson und Nick Tilley und bei Anwendungsversuchen wie von Zuzanna Kita et al. nicht näher bestimmt.⁵ Eine konzeptionelle Annäherung nehmen Brad Astbury und Frans Leeuw (2010) und übersetzt auch Anna von Werthern (2020) vor, indem sie zwei typische Fehler in der Konzeptualisierung von Mechanismen anführen und so eine Negativdefinition anbieten: Wirkmechanismen seien nicht mit den Programminputs oder den Programmaktivitäten gleichzusetzen und zudem seien sie keine einfach beobachtbaren Attribute (Werthern 2020, S. 222). Gleichzeitig wird an einem Beispiel von Brad Astbury und Frans Leeuw zu intendierter Schwangerschaftsreduktion deutlich, dass sie in den Beispielen zwar die Verwechslung von Mechanismus und Aktivitäten umgehen. Aus unserer Sicht geschieht dies aber zu dem Preis, dass die stringente Unterscheidung von Mechanismus zu Outcomes aufweicht. So seien unter anderem das veränderte Wissen und die Überwindung kultureller Tabus bei den Adressat:innen Mechanismen, die als affektive und kognitive Reaktion zu den erwünschten Schwangerschaftsrückgängen führen können (Astbury/Leeuw 2010, S. 367). Mechanismen werden hier mit vorgelagerten Outcomes in komplexen Wirkketten gleichgesetzt. Dies löst die Definitionsprobleme rund um die sehr abstrakte

4 Erkenntnistheoretisch orientiert sich der Wirkmechanismus damit am Ansatz des „Critical Realism“. In Abgrenzung zu positivistischen Perspektiven geht es hier nicht um die Formulierung abstrakter Gesetzmäßigkeiten, sondern um die Analyse von Strukturen und Mechanismen, die in spezifischen Kontexten Phänomene hervorbringen, den sogenannten „causal powers“ (Ziegler 2023, S. 27).

5 Auch eine Metaanalyse von Sebastian Lemire und seinen Kolleg:innen verdeutlicht, dass selbst bei Ray Pawson und Nick Tilley verschiedene Zuschnitte angelegt sind (Lemire et al. 2020, S. 75 ff.). So verwundert es nicht, dass in den 195 untersuchten Anwendungsstudien Mechanismen uneinheitlich, als Programmkomponente oder als Reaktionen der Teilnehmenden auf Programmaktivitäten, gefasst werden.

Bestimmung von Wirkmechanismen in der Evaluationsdebatte aus unserer Sicht nicht zufriedenstellend. Gerade formative Evaluationen, die Angebote durch die Reflexion ihres sozialpädagogischen Handelns weiterbringen wollen, interessieren sich für die „Ursachen“ in der Gestaltung von Beratungsinteraktionen. Und die eingangs zitierte Frage „What Is This Thing Called a Mechanism?“ bleibt daher weiterhin virulent und für gegenstandsbezogene Bestimmungen offen.

Aus unserer Sicht war es demgegenüber gewinnbringend, Wirkmechanismen als spezifische Kombination von Inputs und Aktivitäten zu fassen. Inputs und Aktivitäten wurden nicht mehr isoliert als einzelne Bestandteile der Wirkmodellierung gedacht, sondern in ihrer Verwobenheit qualifiziert. Anschließend an Debatten der Sozialen Arbeit konnten wir so das professionelle Handeln in Beratungs- und Unterstützungsprozessen in den Mittelpunkt der Wirkungsevaluation setzen. Wir verstehen Wirkmechanismen als haltungsbasierte Handlungen der Berater:innen, die im Zusammenspiel mit den Adressat:innen Stabilisierungen und Veränderungen ermöglichen. Dieses Verständnis haben wir in der Arbeit mit dem empirischen Material entwickelt. Hier zeigte sich, dass die Adressat:innen ihre eigenen Veränderungen vor allem auf die Art (das Wie) des beraterischen Handelns, Kommunizierens und Begegnens zurückführen.

Mit dieser Idee von Wirkmechanismus schließen wir an einen zentralen Begriff im Professionalitätsdiskurs der Sozialen Arbeit an: professionelle Haltungen. Der Begriff der Haltungen ist zwar in der Praxis und Forschung sehr gebräuchlich, aber er bleibt in theoretischer Hinsicht vielschichtig und teilweise unspezifisch (Domes/Wagner 2020). Diese Diffusität liegt möglicherweise auch am Charakter von Haltungen selbst, die sich schwerlich offensichtlich erfassen und direkt aus Handlungen ableiten lassen. Professionelle Haltungen können als verinnerlichte Einstellungen oder Überzeugungen gefasst werden (Albrecht 2017; Schmitt 2011), die sich auf Menschen, Situationen und Arbeitsprozesse beziehen und denen im Fachdiskurs eine spezifische Qualität zugewiesen wird.⁶ Als Beispiele nennt Ralf Albrecht für die Beratungsarbeit unter anderem Ressourcenorientierung, Kontextsensibilität, Wertschätzung, Echtheit, Zirkularität und Allparteilichkeit als relevante Haltungen (2017, S. 50 ff.). Professionelle Haltungen strukturieren Wahrnehmungen und Handlungsoptionen der Berater:innen vor und ermöglichen flexible Handlungsentscheidungen „jenseits von biografisch tradierten Einbahnstraßen und starren Handlungsrezepten“ (Schmitt 2011, S. 29). Sie sind schwer formal erlernbar und verdichten sich i. d. R. in langjährigen Erfahrungsaufschichtungen und kontinuierlichen Reflexionsprozessen. Sie basieren auf der Bereitschaft, sich mit biografischen Hintergründen und eigenen Arbeitsmotivationen auseinanderzusetzen, orientieren sich an sozialarbeiterischen und feldspezifischen Standards und sind insbesondere mit Fragen nach

6 In dieser Konzeptionalisierung sind Haltungen schwer abgrenzbar gegenüber Konzepten wie Gesinnung, Ethik und Habitus (Domes/Wagner 2020).

den Verhältnissen von normativen Ansprüchen, theoretischem Wissen und persönlichen Zielstellungen reflektierbar. Haltungen sind demnach nur vermittelt sichtbar und qualifizieren so die Aktivitäten der Berater:innen im Sinne der Mechanismusdefinition von Ray Pawson und Nick Tilley in wirkungsermöglichender Weise. Es ist die Haltungsbasierung, die begründet, dass die Handlungen der Berater:innen über simple, beobachtbare Aktivitäten hinausgehen.

Wieso eignet sich aus unserer Sicht der konzeptionelle Anschluss an professionelle Haltungen? Erstens setzten wir mit dem Blick auf haltungsbasiertes Handeln die Person der Berater:innen in den Mittelpunkt der Analyse. Dies birgt die Chance, individuell gestaltete Praxis professionsbezogen zu rekonstruieren und professionelles Beratungshandeln kleinteilig – abseits von groben Beratungsansätzen – zu analysieren. So sind fachliche Herausforderungen, zum Beispiel das konkrete Nähe- und Distanzverhältnis zu den Adressat:innen, fallspezifisch reflektierbar. Im Sinne einer formativen Evaluation lässt sich mit dieser vorgeschlagenen Konzeptionierung von Wirkmechanismen die Praxis der Berater:innen zielorientiert unterstützen. Zweitens belegen Befunde aus der Beratungsforschung, dass die Beratungsbeziehung ein zentraler Wirkfaktor ist (Schubert/Rohr/Zwicker-Pelzer 2019, S. 181). Wir wollten diesen übergreifenden Befund für die konkreten Wirkzusammenhänge unserer Beratungsfelder untersuchen und die haltungsorientierten Handlungen als Basis für das entstehende Vertrauensverhältnis zwischen Adressat:innen und Berater:innen fassen. Die Fachdebatte zur Beratung im Kontext Rechtsextremismus weist Haltungen in diesem Zusammenhang eine besonders hohe Bedeutung zu (Köbberling 2021).

Als Folge des Zuschnitts unserer Analysen werden die Adressat:innen als Handelnde mit ihren Mitwirkungen an den Beratungen weniger sichtbar als die Berater:innen. Dies kann mit Blick auf die koproduktiven Prozesse der Wirkungsherstellung in der Sozialen Arbeit kritisiert werden. Die Adressat:innen haben selbstverständlich mit ihren Lebenslagen und ihrer Veränderungsmotivation, aber auch mit ihrem Handeln und der eigensinnigen Aneignung von Veränderungsimpulsen, zentralen Einfluss auf die Richtung und Intensität von Wirkungen. Wir haben uns dazu entschieden, Koproduktion eher als Grundprämisse in den KMO-Konfigurationen mitzuführen und nicht zum zentralen Gegenstand des Mechanismus zu machen. Für uns überwiegt so der Vorteil, dass mit diesem Zuschnitt für die Beratungsangebote der Nutzen verbunden ist, ihre Handlungsoptionen professionell reflektieren zu können. Es kann auf diese Weise außerdem dargestellt und diskutiert werden, welchen konkreten Beitrag die Berater:innen zu Veränderungen und Stabilisierungen bei den Adressat:innen leisten (können).

4.2.2 Empirische Rekonstruktion von Wirkmechanismen

Was man mit dieser theoretischen Brille in den Blick bekommt und welchen Nutzen dieser Zuschnitt in der Betrachtung der beiden Beratungsfelder hat, beschreiben wir im folgenden Abschnitt. Dafür haben wir exemplarisch zwei zentrale Wirkmechanismen beider Beratungsfelder herausgegriffen: die kritisch-zugewandte Differenzgestaltung in der Ausstiegs- und Distanzierungsberatung und das parteiliche Beiseitestehen in der Opferberatung. Dabei ist zu berücksichtigen, dass es sich um zwei getrennte Studien handelt, in deren Rahmen unterschiedliche Daten erhoben und genutzt wurden. In der Analyse der Ausstiegs- und Distanzierungsberatung konnten mehrere Fälle fallvergleichend ausgewertet werden. In der Studie zur Opferberatung hingegen analysierten wir einzelfallbezogen Schlüsselsequenzen. Entsprechend unterscheiden sich auch die folgenden Darstellungen beider Studien.

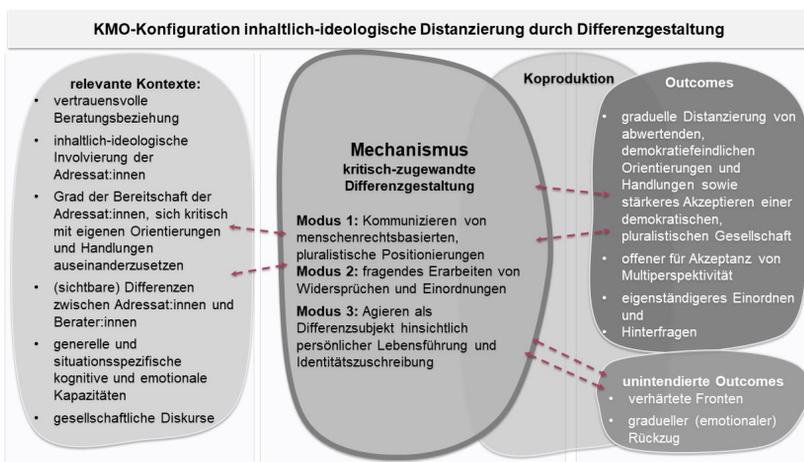
4.2.2.1 Ausstiegs- und Distanzierungsberatung

In der Untersuchung der Ausstiegs- und Distanzierungsberatung wurden relevante Wirkzusammenhänge der einzelfallorientierten Beratung von Personen rekonstruiert, die bei der Distanzierung von extrem rechten Szenen und Ideologiebezügen unterstützt werden (Figlesthler/Schau 2023). Auf Basis von qualitativen Interviews mit Berater:innen (acht Einzelfälle) sowie mit Adressat:innen (drei Einzelfälle) wurde ein Modell von kontextabhängigen Wirkzusammenhängen für dieses Arbeitsfeld erarbeitet. Der Fokus der Studie lag vorrangig auf der Identifikation von Mechanismen und damit auf der Frage, wie Berater:innen durch haltungsbasiertes Handeln zu Veränderungen der Adressat:innen beitragen. Für die kontextsensible Analyse unter Berücksichtigung subjektiver Perspektiven der Berater:innen und Adressat:innen wurde ein qualitatives Design gewählt, orientiert an den Grundprämissen der Grounded Theory (Breuer/Muckel/Dieris 2019). Die Auswertung der Daten wurde durch sensibilisierende Konzepte (u. a. Kategorien der KMO-Konfigurationen, Wirkverständnis der Interviewten) und Forschungsbefunde zu Distanzierungsprozessen strukturiert. Zugleich erfolgte die Codierung offen hinsichtlich subjektiver Bedeutungszuschreibungen. Zusätzlich wurden Schlüsselstellen interpretiert, zum Beispiel mit konkreten Schilderungen von Vorgehensweisen. So wurden die Mechanismen in einem iterativen Prozess kontinuierlich im Wechselspiel von theoretischen Zuspitzungen und empirisch angeleiteten Verdichtungen geschärft und konkretisiert. Im gesamten Auswertungsprozess war eine fallvergleichende Perspektive zentral für die Erarbeitung der KMO-Konfigurationen. Zudem nutzten wir in einer gemeinsamen Interpretationsgruppe den Vergleich zur Opferberatung, um die Spezifik der KMO-Konfigurationen und insbesondere der Mechanismen zu schärfen. Neben der Auswertung der Daten im Forschungsteam spielte auch die Diskussion und Validierung

mit der Fachpraxis eine relevante Rolle für die Konkretisierung und Präzisierung der Mechanismen.

Wir haben für die Ausstiegs- und Distanzierungsberatung ein Modell von fünf verschränkten KMO-Konfigurationen mit je einem zentralen Mechanismus herausgearbeitet. Diese Aufschlüsselung in verschiedene, miteinander verwobene Wirkzusammenhänge ermöglicht eine analytische Darstellung des komplexen Gesamtprozesses einer professionell begleiteten Distanzierung. Damit wird das konkrete haltungsbasierte Handeln der Berater:innen, das in spezifischen Kontexten in Koproduktionsprozessen mit den Adressat:innen Outcomes ermöglicht, versteh- und reflektierbar. Im Folgenden konzentrieren wir uns auf den Mechanismus der kritisch-zugewandten Differenzgestaltung, der für die KMO-Konfiguration zur inhaltlich-ideologischen Distanzierung zentral ist.⁷

Abbildung 1: KMO-Konfiguration inhaltlich-ideologische Distanzierung durch Differenzgestaltung



Quelle: Figlesthler/Schau 2023, S. 36 (angepasste Darstellung)

In der Abbildung 1 sind die relevanten Kontexte, der Mechanismus sowie intendierte Outcomes und mögliche unintendierte Outcomes dieser KMO-Konfiguration dargestellt. Der Mechanismus wird durch unterschiedliche Handlungs-

7 Die anderen vier Mechanismen sind: adressat:innenorientierte Gestaltung der Beratungsinteraktion zur Etablierung einer vertrauensvollen Beratungsbeziehung, veränderungsorientierte narrative Kohärenzarbeit zur Anregung identitätsbezogener Veränderungen, autonomieerhaltende Veränderungsbegleitung zur sozialen Distanzierung sowie autonomieermöglichende Unterstützung zur Alltagsbewältigung und Stabilisierung der Lebenssituation der Adressat:innen.

modi konkretisiert, um die vielfältigen Handlungspraxen der Beratungsinteraktionen abzubilden.

Von Beginn an lag ein besonderer Fokus der Auswertung darauf, zu rekonstruieren, wie die Berater:innen inhaltlich-ideologische Distanzierungen bei den Adressat:innen unterstützen. Dieses Anliegen ist eine grundlegende Zielstellung zivilgesellschaftlicher Ausstiegs- und Distanzierungsberatung und zugleich ein sensibles Thema, das oft vage bleibt. Im Laufe der Auswertung erarbeiteten wir eine kritisch-zugewandte Differenzgestaltung als zentralen Mechanismus, der unter anderem eine graduelle Distanzierung von menschen- und demokratiefeindlichen Orientierungen und Handlungen sowie ein stärkeres Akzeptieren einer demokratischen, pluralistischen Gesellschaft aufseiten der Adressat:innen anregen kann.

Die Adressat:innen sollen durch konstruktive Differenzverfahren im Rahmen der Beratungsbeziehung unter anderem darin unterstützt werden, die Koexistenz verschiedener Perspektiven und Ansichten (wieder) anzuerkennen und zum Beispiel pauschalisierende Feindbilder aufzubrechen. Eine kritisch-zugewandte Haltung ist mit spezifischen professionellen Anforderungen verbunden, die es auszubalancieren gilt: Einerseits markieren die Berater:innen gegebenenfalls deutlich Differenzen zu ihrer eigenen politischen Haltung und positionieren sich mitunter kritisch zu Äußerungen und (vergangenen) Taten der Adressat:innen. Andererseits agieren sie im Beratungsprozess zugewandt und begegnen ihrem Gegenüber offen und wertschätzend. Dieses haltungsbasierte Agieren zeigt sich zum Beispiel im konkreten Handlungsmodus des Kommunizierens von menschenrechtsbasierten, pluralistischen Positionierungen. Es ist eine wichtige Funktion dieses Mechanismus, durch kritisch-zugewandtes Agieren Differenz für die Adressat:innen in der Beratungsinteraktion in konstruktiver Weise erfahrbar zu machen. Dazu gehört unter anderem, zu erleben, dass verschiedene Perspektiven und Haltungen (innerhalb eines demokratischen, nicht menschenfeindlichen Spektrums) selbstverständlich nebeneinanderstehen können, ohne die Beratungsbeziehung zu belasten. In den Interviews zeigt sich deutlich, dass die konstruktive Erfahrung von Differenz aus Sicht der Adressat:innen auf einer vertrauensvollen Beziehung zu den Berater:innen basiert. Sie ist die notwendige Grundlage, um Differenzen aushalten und gezielt für Veränderungsprozesse sowie Neuorientierungen nutzen zu können.

Daneben ist das fragende Erarbeiten von Widersprüchen und Einordnungen eine wichtige Vorgehensweise, um Ablehnungshaltungen zu irritieren und zum Beispiel Öffnungen für die Akzeptanz von Multiperspektivität anzuregen. Dabei fragen Berater:innen im Rahmen des kritisch-zugewandten Agierens beispielsweise nach, wieso Adressat:innen eine bestimmte Meinung vertreten. Durch das fragende Erarbeiten von Positionierungen im Rahmen des längerfristigen Beratungsprozesses sollen Adressat:innen dazu angeregt werden, eigenständige Einordnungen von Sachverhalten vorzunehmen, die im Einklang mit ihren

eigenen Werten und Glaubenssätzen stehen, die ebenfalls in der Beratung besprochen werden. In diesem Modus werden durch Fragen der Berater:innen auch gezielt Widersprüchlichkeiten bei pauschalisierenden, abwertenden Äußerungen thematisiert, etwa gegenüber Geflüchteten oder Homosexualität. Ziel ist es, unter anderem auf diese Weise Stereotype aufzubrechen und pauschalisierende Ablehnungskonstruktionen⁸ zu irritieren. Durch das kritisch-zugewandte Agieren der Berater:innen sollen auf diese Weise Distanzierungen von abwertenden Orientierungen oder von Hass auf vermeintlich Fremdes angestoßen werden. In den Interviews mit den Adressat:innen wird deutlich, dass das grundlegende Infragestellen jahrelanger Überzeugungen und Deutungsmuster als herausforderungsvoll und zum Teil krisenhaft erlebt wird. Daher braucht es für eine nachhaltige inhaltlich-ideologische Distanzierung die Bereitschaft der Adressat:innen, offen über inhaltliche Fragen und Widerstände zu sprechen, die im Beratungsprozess aufkommen können. Die Analyse unterstreicht die Relevanz der kritisch-zugewandten Haltung der Berater:innen, um diese Bereitschaft kontextabhängig herzustellen, zu verfestigen und für Veränderungsprozesse zugänglich zu machen.

4.2.2.2 Beratung und Unterstützung von Betroffenen rechter, rassistischer und antisemitischer Gewalt

Auftrag der Beratungsangebote für Betroffene rechter, rassistischer und antisemitischer Gewalt ist es, Betroffene bei der Bewältigung der (im-)materiellen Folgen der Gewalttaten und bei der Wahrnehmung und Durchsetzung ihrer Rechte zu beraten und zu unterstützen. In einer multiperspektivischen Fallstudie wurden die Perspektiven einer Beratungsnehmerin, der Berater:innen und einer Engagierten aus einer lokalen erinnerungspolitischen Initiative qualitativ erhoben und ins Verhältnis gesetzt. Diese Fallstudie, auf die in diesem Abschnitt Bezug genommen wird, wurde in der Untersuchung mit den Befunden einer quantitativen Beratungsnehmenden-Befragung kombiniert (Haase/Rottinghaus 2024).

Die Beratungsnehmerin ist eine Überlebende eines rassistisch motivierten Anschlags, die von den Berater:innen bei der Bewältigung der psychischen, physischen und materiellen Folgen des Anschlags unterstützt wird. Ein besonderer Fokus lag auf der Interpretation der Eigensichtweisen der Adressatin, die sich auf das subjektive Erleben von Veränderung und die Wiedererlangung von Handlungsfähigkeit beziehen. Wirkmechanismen stellen sich in diesem Zusammenhang als jene haltungsbasierte Handlungen dar, die – sich in unterschiedlicher Weise äussernde – Handlungsfähigkeit der Adressatin unter bestimmten Bedingungen (Kontexte) hervorbringen.

8 Das Konzept der pauschalisierenden Ablehnungskonstruktionen (PAKOs) wird von Kurt Möller (2017) ausführlich beschrieben.

Aufgrund der Parallelen zum bereits vorgestellten methodischen Vorgehen in der Studie zur Ausstiegs- und Distanzierungsberatung wird im Folgenden ausschließlich auf die tiefergehende Analyse von Schlüsselsequenzen Bezug genommen. Dafür wurden Sequenzen mit evaluierenden Anteilen zum Handeln und zur Wirksamkeit der Opferberatung sowie szenisch-episodische Erzählungen⁹ ausgewählt. Die Analyse ermöglichte es, die erzählten Veränderungen interpretativ zu erschließen. Ausgewählte Sequenzen wurden, orientiert an der Narrationsanalyse, segmental gegliedert und die Sprachmodi der Darstellung (Erzählung, Beschreibung, Argumentation) berücksichtigt. Sprechereinheiten (Satzteile, Sätze) wurden Zug um Zug interpretiert. Da (Nach-)Fragen im Interview häufig auf Bewertungen und Einschätzungen zur Opferberatung abzielten, nahmen Argumentationen (Bewertungen, Begründungen, Erklärungen, evaluative Stellungnahmen) in den Interviews insgesamt viel Raum ein. Szenisch-episodische Erzählungen erwiesen sich aber in besonderer Weise als geeignet, Wirkmechanismen zu erschließen. Denn sie ermöglichten einen tieferen Einblick in die konkreten Interaktionen von Berater:innen und Adressat:innen und die situierte Praxis. In diesem Abschnitt wird deshalb die Auswertung einer solchen Schlüsselsequenz in gekürzter Fassung vorgestellt. Damit soll exemplarisch illustriert werden, wie sich auf der Basis des Interviewmaterials das parteiliche Beiseitestehen als ein Wirkmechanismus rekonstruieren ließ.¹⁰ In der Sequenz geht die Beratungsnehmerin näher auf die Art und Weise der Begleitung durch ihren Berater ein, um die besondere Qualität dieser zu plausibilisieren:

„Ja. Also wir hatten unsere erste Gedenkveranstaltung gehabt. Der Hannes (Pseudonym) war dabei. Der saß neben mir, und dann hab' ich gesagt: ‚Hannes, ich hab' keine Kraft, ich weiß es nicht, wie ich reagieren soll‘. Und da musste ich das Knie von Hannes so festgehalten haben, ich glaub', der hat einen blauen Fleck gekriegt, meine Meinung. Ich hab' ihm aber das nicht gesagt. Und (...) also die Kraft, wie er neben uns saß oder wir das erste Gespräch mit der Stadt – der hat sich viel gekümmert, sag' ich mal. Also Hannes hat mir viel Kraft gegeben. Also [Initiative] war jetzt, die hat uns unterstützt, bis heute noch, aber von der Kraft von Hannes hab' ich – die Kraft war anders gewesen. Und der fehlt mir wirklich“ (BN 2022, Z. 168–175).

Die Sequenz setzt sich aus einer kurzen szenischen Episode, die die Interaktion von Berater und Adressatin auf der ersten Gedenkveranstaltung umfasst, und einer über dieses Setting hinausweisenden evaluativen Passage zusammen. Die Interviewte berichtet in der Sequenz davon, wie sie am ersten Gedenktag des rassisi-

9 Bei der Erzählung kann es sich um szenisch-episodische Erzählungen (Erzählen im engeren Sinne), die berichtende Darstellung und die chronikartige Darstellung handeln.

10 Insgesamt wurden in dieser Fallstudie drei ineinandergreifende KMO-Konfigurationen rekonstruiert.

tisch motivierten Anschlags von ihrem Berater unterstützt wurde. Mit dem ersten Gedenktag wählt sie ein besonderes Ereignis aus. Ihre persönliche Situation ist unter anderem durch die traumatisierenden Wirklichkeitserfahrungen und die langwierigen Folgen der Traumatisierung durch den rassistisch motivierten Anschlag geprägt. Dies ist zwar nicht Gegenstand dieser Sequenz, ihre Schilderung ist aber nur vor dem Hintergrund ihrer hochbelasteten Situation (u. a. Posttraumatische Belastungsstörung, fehlendes Unterstützungsnetzwerk) und ihrer biografischen Erfahrungen zu verstehen.

In der kurzen szenischen Episode im ersten Subsegment fokussiert die Interviewte ausschließlich auf die Wechselbeziehung mit ihrem Berater. In der Szene sitzt er neben ihr – direkt an ihrer Seite. Die Überlebende stellt die Interaktion dar, indem sie ihre direkte Ansprache des Beraters in der Situation wiedergibt. Auf diese Weise versetzt sie sich in eine zu diesem Zeitpunkt überforderte, handlungsunfähige und hilfsbedürftige Verfassung. Die Interviewte stellt dar, ihrem Berater ihr Befinden selbst aktiv übermittelt zu haben. In der Situation der Kraftlosigkeit und im dargestellten Zustand enormer Handlungsunsicherheit beschreibt sie, das Knie des Beraters vehement festgehalten zu haben. Bemerkenswert ist in der Sequenz der „blaue Fleck“, den sie als Resultat ihres Handelns vermutet. Der kraftvolle Zugriff einerseits und die Erfahrung von Kraftlosigkeit andererseits werden widerspruchsfrei dargestellt. Schließlich ermöglicht der kraftvolle Zugriff auf die Person neben ihr erst die helfende Erfahrung von Kraft.

Im zweiten Subsegment nimmt die Interviewte eine Bewertung der unterstützenden Beratungsbeziehung vor, die über die exemplarisch angeführte Interaktion hinausgeht. In ihrer Wahrnehmung vermittelt ihr Hannes als Kraft-Subjekt unmittelbar durch seine Positionierung und Präsenz, die eine Nähe in der Interaktion hervorbringt, Kraft. In der evaluativen Passage weitet sie das Setting aus. Einerseits wird die Bewertung über den ersten Gedenktag hinaus auch für das erste Gespräch mit der Stadtverwaltung formuliert. Andererseits wird die Bedeutungszuschreibung auf die Familie („uns“) bezogen. Die Adressatin bettet damit ihre Bedeutungszuschreibung in einen größeren Rahmen ein und hebt auf diese Weise die Unterstützungsleistung des Beraters hervor.

Im Interview fokussiert sie – vor dem Hintergrund des Interviewkontexts – ausschließlich auf Hannes als kraftgebendes Subjekt, die Interaktion und ihre Resultate. Dabei wird der Berater in anderen Schlüsselsequenzen durchaus auch stärker in seinem Agieren dargestellt. Das Setting wird nicht näher beschrieben. Es ist aber anzunehmen, dass für ihre Wahrnehmungen die sozialräumlichen Kontexte – die soziale Aushandlung mit Stadtvertreter:innen und das kollektive Erinnern – hoch relevant sind. Die Interaktion ist in den öffentlichen (Erinnerungs-)Raum und in den machtvollen Raum der politischen Aushandlung mit Vertreter:innen der Stadt, die Träger von Macht sind, eingelassen. Die Interviewte nimmt in ihre Einschätzung die erinnerungspolitische Initiative auf, denn ohne deren Wirken wären diese Räume aus ihrer Sicht nicht existent. Mit der

Herstellung einer Differenz zwischen Initiative und Opferberatung hebt sie die Qualität des Beraters als Kraft-Ressource hervor. Sie schließt mit der Markierung ihres emotionalen Gefühls der Sehnsucht, die durch den Weggang ihres Beraters ausgelöst wird und die – sie spricht im Präsens – zum Interviewzeitpunkt nach wie vor besteht.

Handlungsfähigkeit entsteht in dieser Perspektive, indem der Berater durch parteiliches, kraftspendendes Beiseitestehen Erfahrungen von „Kraft“ ermöglicht. Kraft kann als Spüren von Regie über den eigenen Körper und das eigene Sprechen, als Heraustreten aus der Position der wahrgenommenen Ohnmacht und gesellschaftlich produzierten Unsichtbarkeit der von rassistischer Gewalt Betroffenen gedeutet werden. Durch das parteiliche Beiseitestehen wurde die Überlebende offenbar in dem mehrjährigen empowernden und zugleich entlastenden Beratungs- und Unterstützungsprozess in die Lage versetzt, für sich selbst zu sprechen und ihre Anliegen zu vertreten.

Im kraftspendenden Beiseitestehen spiegelt sich Parteilichkeit als uneingeschränktes und solidarisches Engagement der Beratenden für die Interessen der Adressatin wider (Bitzan 2011, S. 307). Die Relevanz der parteilichen Haltung resultiert daraus, dass sie – in Differenz zu der von Betroffenen wahrgenommenen gesellschaftlichen Verharmlosung rechter, rassistischer und antisemitischer Gewalt – Anerkennung vermittelt und darüber Veränderungen bei Adressat:innen ermöglicht. Die Adressatin fühlt sich von ihrem Berater mit ihren Leidenserfahrungen und Forderungen verstanden, anerkannt und unterstützt.

In diesem Zusammenhang ist es wichtig, zu berücksichtigen, dass das von der Betroffenen wahrgenommene (Nicht-)Handeln, insbesondere der Stadt, in der Vergangenheit und Gegenwart hochrelevant für ihren Veränderungsprozess ist. Ihre Wahrnehmung und ihr Veränderungsprozess sind eingelassen in den gesellschaftlichen Kontext. Gesellschaftliche Verharmlosung und Verleugnung rassistischer Gewalt lassen sich, wie Gesa Köbberling (2022) darstellt, als Prozesse der Unsichtbarmachung der Betroffenen und ihrer Gewalterfahrung lesen und als wesentliche Dimension rassistischer Gewalt beschreiben. Die Kämpfe Betroffener um selbstbestimmte Sichtbarkeit können als relevanter Teil der Bewältigung der Gewalt verstanden werden (Köbberling 2022, S. 268). In den Interviews mit der Adressatin, den Berater:innen und einer Engagierten einer erinnerungspolitischen Initiative wurde deutlich, dass das Erinnern an die Todesopfer des Anschlags umkämpft und in hegemoniale Diskurse um Erinnern eingebettet ist. Aushandlungen des öffentlichen Erinnerns unterliegen gesellschaftlichen Verhältnissen der Betrauerbarkeit (Fischer 2021, S. 153).

Vor dem Hintergrund des gesamten Interviewmaterials zeigt sich, dass aus der Perspektive der Adressatin parteiliche Nähe Veränderungen in ihrer Handlungsfähigkeit hervorbringt. Der Wunsch nach Nähe in dem dargestellten Fall lässt sich plausibel anhand ihrer hochbelasteten Lebenssituation, ihrer Biografie und ihrer Bedürfnisse nach parteilichem Beistand und Unterstützung in

den herausfordernden Situationen und anhand des gesellschaftlichen Kontexts erklären. Die jeweils ermöglichenden und einschränkenden Perspektiven von Nähe und Distanz in der Gestaltung der Beratungsbeziehung wurden von den Berater:innen im Interview diskutiert. In der multiperspektivischen Betrachtung des Falls und in der Triangulation von fachlicher Perspektive und der Eigensicht der Adressatin konnte die Bedeutung einer Balance von Nähe und Distanz im Veränderungsprozess der Adressatin geschärft werden.

4.2.3 Die Balance von Nähe und Distanz im wirkungsermöglichenden Handeln

Mit Blick auf beide qualitativen Studien ermöglichte die Analyse von Wirkmechanismen als haltungsbasierte Handlungen in besonderer Weise, die Binnenlogiken der unterschiedlichen Beratungsfelder zu erschließen und Kernprobleme professioneller Fallbearbeitung zu diskutieren. In die empirisch erschlossenen Wirkmechanismen ist das Spannungsfeld von Nähe und Distanz als professionelles Kernproblem eingelassen. Im Professionalitätsdiskurs in der Sozialen Arbeit wird an Sozialarbeitende die Anforderung formuliert, das Spannungsfeld von Nähe und Distanz in der professionellen Fallbearbeitung zu bewältigen (Königter 2017, S. 94). Dabei ist zu berücksichtigen, dass die fachliche und ethische Spannung zwischen solidarischer Akzeptanz und kritischer Distanz gegenüber Adressat:innen und ihren Lebenswelten prinzipiell nicht gelöst werden kann (Dörr/Müller 2012, S. 15).

Die relevanten Haltungen, kritisch-zugewandt in der Ausstiegs- und Distanzierungsberatung und parteilich in der Opferberatung, auf denen die Differenzgestaltung und das Beiseitestehen basieren, lassen sich auf den ersten Blick als Kontrastfolie lesen. Dieser Kontrast resultiert aus den unterschiedlichen Adressat:innen und Anliegen der Beratungsarbeit. Dementsprechend stellt sich auch das konkrete Spannungsfeld von Nähe und Distanz in unterschiedlicher Weise dar: In der Ausstiegs- und Distanzierungsberatung ist das kritische Moment, das unter anderem in der bewussten Arbeit mit der Differenz zwischen den Einstellungen und dem Verhalten der Adressat:innen einerseits und der Berater:innen andererseits liegt, mit der offenen, zugewandten und wertschätzenden Haltung der Berater:innen in Balance zu bringen. In der Opferberatung ist das Spannungsmoment auszugestalten zwischen einer parteilichen Haltung, in der sich eine solidarische Nähe zu den Betroffenen rechter, rassistischer und antisemitischer Gewalt ausdrückt, und distanzierenden Momenten, die einen positiven, Autonomie ermöglichenden Veränderungsprozess befördern.

Auf einen zweiten Blick wird deutlich, dass Differenz in der Veränderungsarbeit in beiden Beratungsfeldern unterschiedlich eingesetzt wird. Die Analysen haben gezeigt, dass Differenz zwischen den Haltungen von Berater:innen und

Adressat:innen in der Ausstiegs- und Distanzierungsberatung in erster Linie eingesetzt wird, um inhaltlich-ideologische Distanzierungen bei den Adressat:innen anzuregen und Differenz als nicht beziehungsgefährdend zu verinnerlichen. Die Berater:innen der Opferberatung dagegen vermitteln – in Differenz zu der von Betroffenen wahrgenommenen gesellschaftlichen Verharmlosung rechter, rassistischer und antisemitischer Gewalt – Anerkennung und können darüber Stabilisierungen und Veränderungen bei den Adressat:innen ermöglichen. Die Erfahrung glaubwürdiger Anerkennung des Erlebten und des rechten, rassistischen und antisemitischen Tatmotivs durch die Berater:innen und die Erfahrung, nicht allein zu sein und emotional zugewandt gehört zu werden, erweist sich als eine Differenz Erfahrung für die Adressat:innen, die Stabilisierung und Veränderung ermöglicht.

Gemeinsam ist den Beratungsfeldern die adressat:innenorientierte Gestaltung des Beratungsprozesses, die auf zentralen übergreifenden Prämissen, wie unter anderem Wertschätzung, Transparenz und Ressourcenorientierung (Albrecht 2017) basiert. Eine solche Ausgestaltung von Beratungsarbeit kann und soll eine vertrauensvolle Beratungsbeziehung als Basis für Stabilisierungs- und Veränderungsprozesse bei Adressat:innen ermöglichen und aufrechterhalten.

4.3 Fazit

Die in diesem Beitrag dargestellten analytischen Zugänge zeigen, dass die Realist Evaluation geeignet ist, die Wirkungsweise von Beratung und Unterstützung zu erklären. In der Logik der Realist Evaluation können die empirisch erschlossenen Wirkmechanismen je nach Kontext anders wirken und zu anderen Ergebnissen führen. Dabei ist zu berücksichtigen, dass auch Wirkmechanismen ihr Erklärungspotenzial für Wirkungszusammenhänge erst in der Rekonstruktion von KMO-Konfigurationen entfalten. In den Untersuchungen bildet das Ineinandergreifen einzelner Konfigurationen aus Kontexten, Wirkmechanismen und graduellen Veränderungen und Stabilisierungen (Outcomes) die komplexen Beratungs- und Unterstützungsprozesse in den Beratungsangeboten ab.

Mit unserer Konzipierung von Wirkmechanismen haben wir den Input (Haltungen) und die Aktivitäten (konkrete Handlungen) in ihrer Verwobenheit professionstheoretisch begründet qualifiziert. Wir rückten die Berater:innen als Handelnde in den Fokus und betteten deren Handlungspraxis professionstheoretisch ein. Es konnte gezeigt werden, dass sich durch diese Konzipierung die Handlungslogiken und professionellen Kernprobleme der unterschiedlichen Beratungsfelder erschließen und diskutieren lassen. Dabei war die Kontrastierung der kritisch-zugewandten Differenzgestaltung in der Ausstiegs- und Distanzierungsberatung und des parteilichen Beiseitestehens in der Opferberatung erkenntnisreich. Sie ermöglichte es, empirisch fundiert sowohl die Spezifika der

untersuchten Beratungsfelder als auch beratungsfeldübergreifende Charakteristika wirkungsermöglichenden beraterischen Handelns herauszustellen. Solche in der Praxis verwertbaren Erkenntnisse können für die Reflexion und zum Beispiel für die Strukturierung von Beratungsprozessen nützlich sein. Bewertungen der Prozessqualität können, wie hier illustriert, an Fachdiskurse – etwa zu Nähe und Distanz oder zu Parteilichkeit – zurückgebunden werden.

Die Konzipierung des Wirkmechanismus als haltungsbasierte Handlungen erwies sich vor diesem Hintergrund für uns als gegenstandsangemessen. Gleichzeitig kann allein der Begriff des Wirkmechanismus, der unterschiedlich verwendet wird, ein technokratisches Verständnis begünstigen und dadurch Missverständnisse hervorrufen. Dies beobachten wir v. a. in interdisziplinären Kontexten. Um dieser Verwechslung entgegenzuwirken, haben wir wiederholt die komplexen Beratungsinteraktionen betont. Berechtigt ist aber die Kritik, dass durch unseren Zuschnitt die koproduktiven Prozesse der Wirkungsherstellung zwar kontinuierlich mitgedacht, aber nicht fokussiert analysiert werden. Kritisiert wird im wissenschaftlichen Diskurs außerdem die mangelnde Abgrenzung von Kontext und Wirkmechanismen oder die Gleichsetzung von Mechanismus und Programmaktivität in Untersuchungen nach der Realist Evaluation (Astbury/Leeuw 2010, S. 367). Auch unsere Konzipierung ging auf Kosten einer klaren Trennung einzelner Wirkkomponenten. Inwiefern der hier diskutierte Ansatz einen tragfähigen konzeptionellen Beitrag zur Weiterentwicklung der Debatte um Wirkmechanismen darstellt und insbesondere in der Sozialen Arbeit nutzbar ist, muss weiter untersucht werden.

Sebastian Lemire et al. (2020) plädieren vor dem Hintergrund der aus ihrer Sicht problematischen Mehrdeutigkeit von Wirkmechanismen dafür, ein geteiltes konzeptionell klares Verständnis des Mechanismus zu entwickeln, um die Glaubwürdigkeit von Wirkungsuntersuchungen nach der Realist Evaluation zu erhöhen. Eine solche Bedeutung konzeptioneller Klarheit ist sicher nicht auszuschließen. Sie suggeriert aber auch eine klare, gegenstandsübergreifende Anwendbarkeit des analytischen Gerüsts KMO. Für uns bot die konzeptionelle Offenheit des Wirkmechanismus in den grundlegenden Ausführungen zur Realist Evaluation die Chance, ein gegenstandsangemessenes Verständnis von Mechanismen im Zusammenspiel von Empirie und Theorie zu entwickeln. Dies bedeutet nicht, dass wir eine derartige Konzipierung übergreifend für die Vielfalt an Forschungsgegenständen und -interessen als sinnvoll erachten. Wir plädieren vielmehr im Hinblick auf die Anwendung von Realist Evaluation dafür, diese gegenstandsbezogenen Klärungsprozesse anzuerkennen und die eigenen konzeptionellen Entscheidungen im Forschungsprozess zu explizieren.

Literatur

- Albrecht, R. (2017): Beratungskompetenz in der Sozialen Arbeit. Auf die Haltung kommt es an! In: Kontext. Zeitschrift für Systemische Therapie und Familientherapie, 48. Jg., H. 1, S. 45–64.
- Albus, S./Micheel, H.-G./Polutta, A. (2018): Evaluation und Wirkungsorientierung. In: Böllert, K. (Hrsg.): Kompendium Kinder- und Jugendhilfe. Wiesbaden/Heidelberg, S. 1563–1579.
- Armborst, A./Biene, J./Coester, M./Greuel, F./Milbradt, B./Nehlsen, I. (2018): Evaluation in der Radikalisierungsprävention. Ansätze und Kontroversen. http://gesellschaftextrem.hsfk.de/fileadmin/HFSK/hsfk_publicationen/prifl118.pdf (13.09.2018)
- Astbury, B./Leeuw, F. L. (2010): Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. In: American Journal of Evaluation, 38. Jg., H. 3, S. 363–381.
- Bitzan, M. (2011): Parteilichkeit. Stichwort. In: Ehler, G./Funk, H./Stecklina, G. (Hrsg.): Wörterbuch Soziale Arbeit und Geschlecht. Weinheim und Basel: Beltz Juventa, S. 307–309.
- Breuer, F./Muckel, P./Dieris, B. (2019): Reflexive Grounded Theory. Eine Einführung für die Forschungspraxis. 4. Auflage. Wiesbaden: Springer VS.
- Dollinger, B. (2018): Paradigmen sozial- und erziehungswissenschaftlicher Wirkungsforschung: Eine Analyse kausaltheoretischer Annahmen und ihrer Folgen für die Soziale Arbeit. In: Soziale Passagen, 10. Jg., H. 2, S. 245–262.
- Dollinger, B. (2024): Kausalität als Konflikt und narrative Legitimierung sozialpädagogischer Praxis. In: neue praxis Sonderheft „Kausalität und die Re-/Konstruktion von Folgen Sozialer Hilfen“, S. 11–23.
- Domes, M./Wagner, L. (2020): Haltung (Gesinnung). <https://www.socialnet.de/lexikon/Haltung-Gesinnung> (Abruf 05.12.2023).
- Dörr, M./Müller, B. (2012): Einleitung: Nähe und Distanz als Strukturen der Professionalität pädagogischer Arbeitsfelder. In: Dörr, M./Müller, B. (Hrsg.): Nähe und Distanz. Ein Spannungsfeld pädagogischer Professionalität. 3., aktualisierte Auflage. Weinheim und Basel: Beltz Juventa, S. 7–31.
- Figlesthaller, C./Schau, K. (2023): Was wirkt wie und warum? Analyse zur Wirksamkeit der direkten Distanzierungsberatung im Bereich Rechtsextremismus und islamistischer Extremismus. Wissenschaftliche Begleitung Handlungsbereich Land im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024, Halle (Saale): DJI.
- Fischer, G. (2021): (Un)doing memory – fehlendes Erinnern an Todesopfer rechter Gewalt. Rekonstruktionen anhand eines Mordes im Landkreis Esslingen 1992. In: Soziale Probleme, 32. Jg., H. 2, S. 151–166.
- Haase, K./Rottinghaus, B. (2024): Was wirkt wie und warum? Analysen zur Wirksamkeit der Beratungsangebote für Betroffene rechter, rassistischer und antisemitischer Gewalt. Wissenschaftliche Begleitung Handlungsbereich Land im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von Michèle Rosenkranz. Halle (Saale): DJI.
- Hauemberger, S./Baumgartner, E. (2017): Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview. In: Zeitschrift für Evaluation, 16. Jg., H. 1, S. 121–145.
- Kita, Z./Hauemberger, S./Wirz, D./Gisler, F. (2022): Wirkungsannahmen sichtbar machen: Konstruktion eines kontext-mechanismus-basierten Wirkungsmodells in der Arbeitsintegration. In: Schweizerische Zeitschrift für Soziale Arbeit, H. 29, S. 72–100.
- Köbberling, G. (2021): Parteilichkeit als Grundlage einer gesellschafts- und machtkritischen Sozialen Arbeit. In: Kaur Cholia, H./Jänicke, C. (Hrsg.): Unentbehrlich. Solidarität mit Betroffenen rechter, rassistischer und antisemitischer Gewalt. Münster: edition assemblage, S. 159–167.
- Köbberling, G. (2022): Rassistische Gewalt als Erfahrung der Markierung und Unsichtbarmachung. In: Zeitschrift für Rechtsextremismusforschung (ZReX), 2. Jg., H. 2, S. 268–286.
- Köngeter, S. (2017): Professionalität. In: Kessl, F./Kruse, E./Stövesand, S./Thole, W. (Hrsg.): Soziale Arbeit – Kernthemen und Problemfelder. Opladen und Toronto: Barbara Budrich, S. 87–105.

- Lemire, S./Kwako, A./Nielsen, S. B./Christie, C. A./Donaldson, S. I./Leeuw, F. I. (2020): What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations. In: *New Directions for Evaluation*, H. 167, S. 73–86.
- Möller, K. (2017): Das Konzept „Pauschalisierende Ablehnungskonstruktionen“ (PAKOs) und die KIS-SeS-Strategie – Theoretische Grundlagen, empirische Befunde und zentrale Schlussfolgerungen. In: Möller, K./Neuscheler, F. (Hrsg.): „Wer will die hier schon haben?“ – Ablehnungshaltungen und Diskriminierung in Deutschland. Stuttgart: Kohlhammer, S. 91–110.
- Möller, K./Buschbom, J./Pfeiffer, T. (2020): Zur Evaluation von Praxisansätzen der Extremismusprävention. In: Slama, B. B./Kemmesies, U. (Hrsg.): *Handbuch Extremismusprävention. Gesamtgesellschaftlich. Phänomenübergreifend*. Wiesbaden: BKA, S. 389–424.
- Nielsen, S. B./Lemire, S./Tangsig, S. (2022): Unpacking context in realist evaluations: Findings from a comprehensive review. In: *Evaluation*, 28. Jg., H. 1, S. 91–112.
- Pawson, R./Tilley, N. (2004): *Realist Evaluation*. https://www.communitymatters.com.au/RE_chapter.pdf (Abruf 11.03.2019).
- Schmitt, J. (2011): Die Haltung – professionell. Repräsentanz und Wirkung: zur Bewältigung pädagogischer Alltagssituationen. In: *Sozialmagazin – Die Zeitschrift für Soziale Arbeit*, 36. Jg., H. 12, S. 28–33.
- Schubert, F.-C./Rohr, D./Zwicker-Pelzer, R. (2019): *Beratung. Grundlagen – Konzepte – Anwendungsfelder*. Wiesbaden: Springer VS.
- Stockmann, R. (2006): *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Münster [u. a.]: Waxmann Münster.
- Werthern, A. v. (2020): *Theoriebasierte Evaluation. Entwicklung und Anwendung eines Verfahrensmodells zur Programmtheoriekonstruktion*. Wiesbaden: Springer VS.
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./Lobermeier, O./Heitmeyer, W. (Hrsg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden: Springer VS, S. 41–68.
- Ziegler, H. (2023): Neo-Aristotelische Perspektiven auf Kausalität: Der meta-theoretische „Critical Realism“. In: *Kausalität und die Re-/Konstruktion von Folgen sozialer Hilfen. Neue Praxis Sonderheft 18*, S. 26–57.

5 Wirkungsevaluation im Bereich Demokratieförderung

Forschungsdesign, Umsetzungserfahrungen und Reflexion von Herausforderungen beim Erfassen von Wirkungen

Aline Rehse, Max Hemmann und Marina Braun

5.1 Einleitung

Wirken demokratiefördernde Projekte – und wenn ja, wie genau und unter welchen Bedingungen? Mit dieser Leitfrage greift der Beitrag die in der Evaluationsforschung seit gut 20 Jahren geführte Debatte um die Wirksamkeit bzw. Wirkungen (sozial-)pädagogischer Praxis auf (u. a. Hense/Taut 2021; Lüders/Haubrich 2006). Die aktuelle Diskussion befasst sich unter anderem mit Fragen gegenstandsangemessener Wirkungsevaluation, geeigneter Methoden der Erfassung von Wirkungen sowie ihrem Nachweis bzw. ihrer Plausibilisierung (Ottmann 2023; Ackermann 2016). Dabei liegt ein zentrales Erkenntnisinteresse auf der Beschreibung und Identifikation von Mechanismen, die Wirkungen generieren, und ihrer Kontextabhängigkeit (Haunberger/Baumgartner 2017, S. 124; Haunberger et al. i. V.).

Innerhalb der Kinder- und Jugendpolitik in Deutschland ist das Praxisfeld Demokratieförderung¹ vergleichsweise jung und noch wenig konturiert. Für dieses liegen bisher nur wenige Wirkungsstudien und forschungspraktische Erfahrungen mit Wirkungsuntersuchungen vor (Bischoff/Zimmermann/König 2021, S. 246, Bischoff/König/Zimmermann 2018).

Im Bundesprogramm „Demokratie leben!“ (BMFSFJ, 2020–2024) sind im Handlungsfeld Demokratieförderung über den Programmzeitraum 34 Modellprojekte gefördert worden, die innovative Angebote mit dem Ziel erprobt haben, in unterschiedlichen gesellschaftlichen Bereichen die politischen Teilhabechancen junger Menschen zu stärken bzw. sie zu demokratischem Handeln zu

1 Für die geförderte Fachpraxis und in der deutschen Fachdebatte gilt der Begriff „Demokratieförderung“ auch als „Containerbegriff“ für verschiedene Demokratie vermittelnde Ansätze wie Demokratiebildung und -pädagogik (Sturzenhecker/Wohnig 2019; weiterführend Rehse/Johann 2022). Für die Evaluation geförderter Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ haben Ehnert et al. (2021) eine Konzeption für Demokratieförderung entwickelt.

befähigen. Die Autor:innen dieses Beitrages haben diese Modellprojekte in ihrem Innovations- und Transferhandeln wissenschaftlich begleitet (Braun et al. 2023).

Vor diesem Hintergrund stellt der Beitrag ein auf Modellprojekte im Handlungsfeld „Demokratieförderung“ angepasstes und erprobtes, kontextsensibles, multiperspektivisches und multimethodisches Wirkungsevaluationskonzept vor (Abschnitt 5.2). Anhand von Anwendungserfahrungen werden Herausforderungen im Umgang mit dem komplexen Evaluationsgegenstand sowie mit der Identifikation und Plausibilisierung von Wirkungen reflektiert, die aktuell in der Fachliteratur diskutiert werden (u. a. Langner/Schott/Grunow 2021; Lemire et al. 2020) (Abschnitt 5.3). Das Fazit bilanziert das vorgestellte wirkungsevaluierende Vorgehen, verweist auf Forschungsdesiderate sowie Grenzen von Wirkungsevaluation (Abschnitt 5.4).

5.2 Wirkungsevaluationskonzept mit Anwendungsbezügen zu demokratiefördernder Praxis

Wirkungsevaluationen erfordern ein theoretisch fundiertes Forschungsdesign, das den jeweiligen Evaluationsgegenständen Rechnung trägt, in den Untersuchungsschritten nachvollziehbar sowie für Evaluierende und Evaluierte ressourcenschonend umsetzbar ist. Ihr Erkenntnisinteresse richtet sich in der Regel einerseits auf die Wirksamkeit von Maßnahmen und andererseits auf deren Wirklogik (Albus/Micheel/Polutta 2011), um Wirkprozesse zwischen Intervention und (un-)intendierten Wirkeffekten nachvollziehbar zu beschreiben und möglichst plausibel zu erklären (Haunberger/Baumgartner 2017, S. 124). All diese Anforderungen greift das im Zentrum des Beitrags stehende Wirkungsevaluationskonzept auf. Erläutert wird das Vorgehen entlang der Untersuchungsphasen (Konzeption, Sampling, Erhebung, Auswertung, Darstellung), verbunden mit Anwendungsbezügen aus dem Feld der Demokratieförderung.

5.2.1 Konzeptionelle Grundlagen

5.2.1.1 Wirkverständnis und Plausibilisierungsanliegen

Zielführend für Wirkungsevaluationen ist es, ihnen ein explizites, theoretisch fundiertes Wirkverständnis sowie „gut begründete Wirkannahmen“ voranzustellen (Hense/Taut 2021, S. 288). Im Beitrag werden Wirkungen im Projektkontext als Veränderungen bei Zielgruppen bzw. in Zielinstitutionen infolge einer Aktivität und damit temporal gefasst. Mit der Einbettung von Projektmaßnahmen in (sozial-)pädagogischen Settings können Wirkungen als das Produkt sozialer Interaktionen definiert werden, die aus Koproduktionsprozessen der daran

Beteiligten hervorgehen.² Für die Zielgruppen zeigen sich koproduzierte Wirkungen als „subjektiv wahrgenommene Änderung lebensweltlicher Faktoren, Ressourcen und Handlungsoptionen“ (Klawe 2006, S. 126). Die soziale Einbettung der Projektaktivitäten wird als Kontextgebundenheit von Wirkungen (Pawson/Tilley 1997) zu einem zentralen definitorischen Bestandteil des Wirkverständnisses im Beitrag. Die Wirksamkeit von Maßnahmen kann folglich nicht losgelöst von ihrer kontextuellen Einbettung betrachtet werden. Situativ variierende Lern- und Beratungssettings, eigensinnige Aneignungsweisen der Adressierten sowie weitere Kontextfaktoren, die mit Mechanismen interagieren, legen ein multifaktorielles Wirkgefüge als Erklärung für Veränderungen nahe. Dem Gegenstand sozialer Praxis werden unidirektionale, lineare, monokausale Erklärungen hingegen nicht gerecht (Braun et al. 2023, S. 52 f.).

Vor diesem Hintergrund sind kausale Wirkungsnachweise sowie insbesondere dafür erforderliche (Quasi-)Experimentaldesigns schwer zu erbringen (Ottmann 2023, S. 10; Ackermann 2016). Naheliegender erscheint uns die Plausibilisierung von Wirkzusammenhängen als „Annäherungen an kausale Aussagen“ (Lüders/Haubrich 2006). Hierfür können Wirkungen auf Basis interindividueller Deutungsmuster von möglichst allen an einer Intervention Beteiligten (Klawe 2006, S. 125 ff.) rekonstruiert werden. Über die Triangulation der verschiedenen Perspektiven und anhand analytischer Einordnungen der Evaluierenden anhand von Fachliteratur sowie etablierten theoretischen wie methodischen Konzepten sollen Wirkungen plausibel geschlussfolgert werden. Rekonstruierte Wirkungszusammenhänge sind dementsprechend als modellhafte Annäherungen an komplexe „tatsächliche“ Wirkungszusammenhänge zu verstehen, die im Rahmen von Evaluationen kaum vollständig empirisch aufgeklärt werden können (Bischoff/Zimmermann/König 2021, S. 245).

5.2.1.2 Wirkprozesse in Wirkungsevaluationen identifizieren

Die soziale Einbettung von Maßnahmen bzw. die Kontextgebundenheit von Wirkungen bildet auch eine Kernannahme der Realist Evaluation (Pawson/Tilley 2004). Ihr Erkenntnisinteresse zielt auf die Frage, ob, wie, für wen und unter welchen Kontextbedingungen eine Intervention funktioniert (Haunberger/Baumgartner 2017). In der Realist Evaluation geht es neben dem Aufspüren von Wirkeffekten vor allem um die Wirklogik einer Intervention als Zusammenspiel von Kontextfaktoren (C), den die Wirkungen generierenden Mechanismen (M) und den Outcomes bzw. Wirkeffekten (O). Sogenannte CMO-Konfigurationen

2 Zum Ausgangspunkt für die Erklärung pädagogischen Handelns bedient sich Klawe (2006) der konstruktivistischen Kernannahme des symbolischen Interaktionismus, wonach Wirkungen soziale Konstruktionen der Beteiligten sind, die eine soziale Situation als wahr deuten und danach handeln, sodass auf diese Weise die Konsequenzen ihres Handelns wahr werden.

werden den Wirkprozessen in Interventionen unterstellt (Pawson/Tilley 2004, S. 9). Mechanismen, die das „Herzstück“ der Realist Evaluation bilden, beschreiben, wie, und erklären, warum eine Maßnahme zu einer Veränderung führt bzw. führen kann.³ Unter Mechanismen werden im Beitrag in Anlehnung an Anthony Lacouture et al. (2015, S. 8) reale, nicht direkt beobachtbare Prozesse verstanden, die das Verarbeiten, Denken sowie Reagieren von Personen(-gruppen) auf verfügbare Ressourcen im gegebenen Kontext, in dem eine Intervention stattfindet, betreffen. Im dynamischen Zusammenspiel mit einem oder mehreren Kontextfaktoren und weiteren Mechanismen sind sie am Zustandekommen von Veränderungen mit beteiligt (ebd.). Innerhalb von Maßnahmen können Mechanismen bei verschiedenen Personen unterschiedlich wirken (Pawson/Tilley 2004, S. 6 f.).

5.2.1.3 Zugang zum Gegenstand der Wirkungsevaluation

Aus dem Grundsatz des Gegenstandsbezugs von und aus dem Genauigkeitsanspruch an Evaluationen (DeGEval – Gesellschaft für Evaluation 2017) leitet sich für die Konzeption von Wirkungsevaluationen der Anspruch ab, den Evaluationsgegenstand genau und umfassend zu bestimmen – idealerweise unter Einbezug der an der Evaluation Beteiligten. Im hier vorgestellten Fall bestimmt sich der Gegenstand aus der Gesamtheit der im Handlungsfeld „Demokratieförderung“ geförderten Modellprojekte und umfasst sowohl die Ebene des Handlungsfeldes als auch die Ebene der Einzelprojekte.⁴ Angesichts vielfältiger Projekte und ihrer Einzelmaßnahmen in unterschiedlichen Erprobungssettings ist die Wirkungsevaluation im Handlungsfeld mit einer hohen Komplexität des Gegenstands konfrontiert. Im Rahmen einer prozessbegleitenden Evaluation ist dieser Gegenstand zugänglich zu machen und „handhabbar“ einzugrenzen (Durchführbarkeitsstandard, DeGEval – Gesellschaft für Evaluation 2017).

Ausgangspunkt der Gegenstandsbestimmung in der wissenschaftlichen Begleitung zu Beginn der Programmförderung bilden das Konzept der Demokratieförderung (Ehnert et al. 2021) sowie die Ausarbeitung von Logischen Modellen

3 In der Literatur herrscht kaum Einigkeit darüber, was einen Mechanismus ausmacht, sie zeugt von zahlreichen unterschiedlichen Definitionen sowie Anwendungen des Begriffs (Haunberger et al. i. V.; Lemire et al. 2020 sowie Figlestahler/Haase/Schau in diesem Band).

4 Die 34 Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ sind zwei Themenfeldern zugeordnet. Sie unterbreiten ihre Angebote in verschiedenen institutionellen bzw. sozialräumlichen Kontexten. Ihre Maßnahmen beziehen sich auf demokratische Kompetenzen, beteiligungsfördernde Verfahren bzw. Strukturen sowie auf die demokratische Beteiligung und Aushandlung fördernde Interaktions- und Kommunikationsprozesse in Gruppen (weiterführend Ehnert et al. 2021).

zum einen für das Handlungsfeld und zum anderen für alle untersuchten Projekte in einem partizipativen Prozess mit den Projektverantwortlichen.⁵

Als geeigneter Zugang zum Evaluationsgegenstand erweist sich für die wissenschaftliche Begleitung – neben der Realist Evaluation – die Cluster-Evaluation (Haubrich 2006, S. 110; nach Worthen/Schmitz 1997).⁶

Das Handlungsfeld der Modellprojekte der Demokratieförderung kann nach Haubrich (2006, S. 109) verstanden werden als Programmbereich mit „einzelne[n] Projekte[n], die unter einer gemeinsamen Initiative bzw. einem Auftrag“ agieren und an ihren „Standorten relativ autonom [...] eigene kontextbezogene Strategien entwickeln, um die übergeordneten Programmziele zu erreichen“. Ziel der Cluster-Evaluation ist es, die Kontextbedingungen der Erprobung zu rekonstruieren, nachvollziehbar zu beschreiben, beobachtete Wirkungszusammenhänge zu analysieren und darüber die Voraussetzungen und Hürden für eine Übertragung der Ergebnisse auf andere Kontexte zu identifizieren (Haubrich 2006, S. 120).

Den Kern des Ansatzes bilden inhaltlich bzw. thematisch identifizierbare Cluster (Haubrich 2006, S. 110). Diese bestehen aus mindestens zwei Projekten, die vergleichbare Maßnahmen verfolgen, so aggregierbare Wirkbeobachtungen ermöglichen (1) und auf diese Weise die Anonymität der Einzelprojekte wahren können (2). Für die einzelnen Cluster gilt es zu beschreiben, was, und zu erklären, warum etwas passiert (3). Eine vergleichende Betrachtung der Projekte eines Clusters erlaubt dann fundierte Rückschlüsse auf Erfolge und Kontextbedingungen von Projektansätzen (4) (ebd.). Diese eher vagen vier Prämissen bieten für das weitere Vorgehen der wissenschaftlichen Begleitung einen hilfreichen Orientierungsrahmen. Ergänzend dazu empfehlen Worthen/Schmitz (1997, S. 304) für die forschungspraktische Anwendung, empirisch überprüfbare Arbeitshypothesen über die Wirkungen der Projekte zu formulieren.

5.2.2 Wirkungsevaluationskonzept für Demokratieförderung und Anwendungsbezüge

Aus den vorherigen Betrachtungen ergibt sich als Ziel der hier vorgestellten Wirkungsevaluation die Formulierung von (übertragbaren) kontextsensiblen sowie projektübergreifenden Schlussfolgerungen für die jeweiligen Cluster. Die leitende Fragestellung der wissenschaftlichen Begleitung lautet: Inwieweit und warum,

-
- 5 Die im Vorfeld der Wirkungsuntersuchungen erstellten Logischen Modelle erfassen die Handlungs- bzw. Wirklogiken auf Einzelprojektebene. Sie stellen ein vereinfachtes Abbild der deutlich komplexeren Projektarbeit dar.
 - 6 Die beiden Ansätze, Realist Evaluation und Cluster-Evaluation, sind aus Sicht der Autor:innen miteinander kompatibel. Sie gehen beide davon aus, dass Kontexte für die Wirkung von Projektangeboten virulent sind, und beide versuchen im Ansatz zu beschreiben und zu erklären, was die Wirkungen hervorbringt.

in welchen Kontexten und in Bezug auf welche Zielgruppen sind bestimmte Angebote der Demokratieförderung hinsichtlich des Erreichens intendierter Wirkungen (nicht) erfolgreich bzw. wirksam? Wie die Umsetzung sowie die Beantwortung der Frage erfolgen kann, wird in den nächsten Abschnitten unter Rückgriff auf die vorgestellten Konzepte erläutert und anhand von Anwendungsbeispielen veranschaulicht.⁷

5.2.2.1 Clusterbezogener Samplingprozess

Für ein gegenstandsangemessenes Vorgehen und in der konkreten Anwendung der Cluster-Evaluation werden mithilfe Logischer Modelle auf Projektebene relevante Inhalte bzw. Themen als demokratiefördernde Strategien (z. B. demokratische Organisationsentwicklung, Empowerment im Kontext von Diskriminierung) sowie relevante Umsetzungskontexte (z. B. Hort, Kommunalverwaltungen) im Handlungsfeld bestimmt. Mittels des theoretisch begründeten Samplings werden je zwei Modellprojekte mit vergleichbaren Zielsetzungen, Zielgruppen und vergleichbarer Wirkannahme innerhalb eines Kontextes identifiziert und zu Clustern zusammengefasst. Anschließend erfolgt die Auswahl vergleichbarer Einzelmaßnahmen innerhalb der Cluster, die die „Wirkausschnitte“ bilden und so die notwendige Verengung der zu beobachtenden Untersuchungsgegenstände für die Wirkungsanalyse leisten (Lüders/Haubrich 2006).⁸ Die Erhebungs- und Auswertungsmethodik wird jeweils auf die beiden Modellprojekte innerhalb eines Clusters abgestimmt und soll eine vergleichbare Methoden- und Datengrundlage gewährleisten. Diese Modellprojekte werden einem kontrastierenden Paarvergleich entlang von Wirkungsdimensionen unterzogen. Im Anschluss daran werden auf Basis der untersuchten Einzelmaßnahmen projektübergreifende Aussagen zu den Wirkweisen und Kontextbedingungen für das jeweilige Cluster abgeleitet.

Das clusterspezifische Vorgehen unserer Wirkungsevaluation wird nachfolgend anhand eines Beispiel-Clusters von Angeboten der Demokratieförderung in Bildungs- und Betreuungseinrichtungen im Grundschulalter, dem Hort, näher erläutert. Als gegenstandsbezogene, theoretische Begründung der Wirkannahme im Cluster lässt sich aus der Literatur unter anderem das Konzept der Demokratiekompetenz von Himmelmann und Lange (2005) heranziehen, das sich

7 Siehe auch den Beitrag von Euteneuer und Rüger in diesem Band.

8 Aufgrund der Restriktionen in begleitenden Evaluationsforschungen (Bohn/Saßmannshausen 2021, S. 290) ist davon auszugehen, dass die Komplexität der Wirkungen von Modellprojekten nicht in Gänze zu erfassen ist (Lüders/Haubrich 2006).

für die Identifikation von relevanten Mechanismen der Demokratieförderung im Bildungskontext heuristisch als hilfreich erweist.⁹

Das Cluster setzt sich aus den Modellprojekten Corgi und Dachs zusammen. Aus Platzgründen kann an dieser Stelle nur das Projekt Corgi exemplarisch vorgestellt werden. Beide Modellprojekte bieten unter anderem langfristig angelegte Beratungen für pädagogisches Personal in Horten sowie spiel- und erfahrungsbasierte Maßnahmen für Hortkinder an. Die Einzelmaßnahmen, die hier als Wirkausschnitt untersucht wurden, streben an, dass die pädagogischen Fachkräfte im Rahmen der Prozessbegleitung der Einrichtung beteiligungsorientierte Haltungen und Kompetenzen weiterentwickeln und in der pädagogischen Arbeit mit Kindern anwenden.

Für die Identifikation von Wirkungen wird ein multimethodisches, multiperspektivisches Erhebungsdesign umgesetzt. Im Hort-Cluster umfasst es teilnehmende Beobachtungen, leitfadengestützte Gruppeninterviews mit dem jeweiligen Hortpersonal bzw. den Hortkindern sowie Einzelinterviews mit dem Leitungspersonal bzw. mit den Projektverantwortlichen (Braun et al. 2023).

5.2.2.2 Empirische Erfassung von Wirkungen und die Rolle von Multiperspektivität

Das Plausibilisierungskonzept zur Gewinnung von plausibel abgeleiteten Aussagen über Wirkungszusammenhänge sieht vor, die relevanten Perspektiven von zentral an der Intervention Beteiligten unter Einsatz verschiedener qualitativer Erhebungs- und Auswertungsverfahren zu erfassen (Perspektivensampling). Mittels qualitativer, teilstrukturierter Leitfadeninterviews (Loosen 2016) werden die Perspektiven der Projektverantwortlichen und -umsetzenden, der Adressat:innen der Angebote (Hortlerzieher:innen und Hortkinder) sowie eine semi-externe Perspektive im Umsetzungskontext der Projektaktivitäten (u. a. Leitungskräfte im Hort) erfasst. Zusätzlich wurde über teilnehmende Beobachtungen (Lüders 2015) ethnografisches Material in die Wirkungsevaluation einbezogen. Mithilfe der Verschränkung der verschiedenen Perspektiven können plausible, empirisch gesättigte Wirkaussagen über die Intervention getroffen werden.

Die wissenschaftliche Begleitung hat neben der Erhebungs- und Auswertungsfunktion zu den einzelnen Perspektiven zugleich eine übergeordnete, alle Perspektiven berücksichtigende, bewertende Funktion. In diesem Zusammenhang werden im Rahmen der hier vorgestellten Wirkungsevaluation multiperspektivisch abgesicherte Aussagen als Konvergenzen zwischen den einzelnen

9 „Demokratie lernen“ findet nach Himmelmann/Lange (2006) auf verschiedenen Ebenen des Lernens statt. Analog dazu verwenden wir diese Ebenen des Lernens als Wirkebenen und unterscheiden die kognitiv-reflexive, die affektiv-motivationale und die praktisch-instrumentelle Ebene (ebd.).

Perspektiven höher gewichtet und als plausibler eingestuft als Aussagen auf Basis von Einzelperspektiven, die von Divergenzen zwischen den jeweiligen Perspektiven zeugen.

5.2.2.3 Identifikation von Wirkungskonstellationen

Für die Auswertungsphase werden die transkribierten Interviews sowie die erstellten ethnografischen Beobachtungsprotokolle (Przyborski/Wohlrab-Sahr 2021) mittels zusammenfassender und strukturierender qualitativer Inhaltsanalyse nach Mayring (2015) ausgewertet. Die Auswertung des Datenmaterials orientiert sich am Wirkschema der Realist Evaluation (Pawson/Tilley 2004). Sie basiert auf einem deduktiv entwickelten, induktiv weiter ausdifferenziertem Auswertungsschema, das der CMO-Logik folgt. Das heißt, im Material werden empirische Hinweise auf Kontextfaktoren, Outcomes und Mechanismen der untersuchten Intervention gesucht (Haunberger/Baumgartner 2017, S. 125).

Zur besseren Abgrenzung von Kontextfaktoren und dem Handeln der Projektverantwortlichen hat die wissenschaftliche Begleitung zudem die Kategorie „Aktivität“ in das CMO-Schema integriert. Für das in der Inhaltsanalyse verwendete Kategorienschema ist „Aktivität“ definiert als Maßnahme der Projekte und ihrer Merkmale. Unter „Kontext“ fassen wir verschiedene individuelle, institutionelle und gesellschaftliche Rahmenbedingungen (u. a. Covid-19-Pandemie) mit Einfluss auf die Wirkweise der Aktivität. Unter „Wirkmechanismen“ verstehen wir – unter Berücksichtigung der Mechanismen-Definition oben – vermittelnde Wirkweisen, die die Outcomes hervorbringen, und unter „Outcome“ plausibel auf die Maßnahme rückführbare (un-)intendierte Effekte.

Um Wirkmechanismen zu identifizieren, werden zusätzlich dichte Schlüsselstellen in den Gruppeninterviews codiert und anschließend auszugsweise mit der Dokumentarischen Methode (Przyborski/Wohlrab-Sahr 2021) ausgewertet. Hierbei wird das Ziel verfolgt, orientierungsstiftende Sinngehalte herauszuarbeiten (ebd., S. 369f.) und die Handlungsorientierungen der Interviewten zu explizieren (ebd., S. 380).

Unter Anwendung des oben beschriebenen Plausibilisierungskonzeptes werden die Befunde dann mittels Daten- und Methodentriangulation (Flick 2011) auf der Einzelprojektebene zusammengeführt und anhand eines kontrastierenden Paarvergleichs für die Ebene des jeweiligen Clusters aggregiert (Braun et al. 2023). Im Folgenden werden beispielhaft Befunde zum Modellprojekt Corgi mit multiperspektivisch rekonstruierten Aussagen über Aktivitäten, Kontextfaktoren, Mechanismen und Outcomes dargestellt (weiterführend Braun et al. 2023).

5.2.2.4 Ergebnisdarstellung – Beispiele für Befunde auf Einzelprojektebene

Das Modellprojekt Corgi bietet im Rahmen seiner Prozessbegleitung kurzzeitpädagogische Beratungen für das Hortpersonal an. Kennzeichnend für die Aktivität ist aus Sicht aller Beteiligten die Kontinuität und Bedarfsorientierung sowie die aktivierende, hierarchiefreie wie partizipative Gestaltung der Sitzungen.

Mechanismen und Outcomes zeigen sich beispielhaft bei der Betrachtung der Top-Flop-Methode, die im Rahmen einer einrichtungsspezifischen Bedarfsanalyse zu Beginn der Prozessbegleitung stattgefunden hat. Die beteiligten Kinder und Erwachsenen sollten mithilfe von Tablets ihre Lieblings- sowie unbeliebte Orte in ihrem Hort (Kontext) dokumentieren. Anschließend haben die Fachkräfte die Fotos der Kinder gesichtet und gemeinsam mit dem Projektmitarbeitenden besprochen (Aktivität). Es zeigte sich, dass die Kinder andere als von den Fachkräften vermutete Orte im Hort fotografierten bzw. als Top- oder Flop-Ort einordneten (Ressource und neues Kontextmerkmal). Das Kennenlernen der Wahrnehmungen der Kinder irritierte die subjektiven Wissensstände der Fachkräfte (Mechanismus auf kognitiv-reflexiver Ebene). Aus dieser Irritation ging neben einem neuen Wissensstand, ein gestiegenes Verständnis und eine größere Empathie für die Perspektive der Kinder hervor (Outcome): „Man versucht sich jetzt, bevor man sich selber ein Bild macht, [...] gleichzeitig auch in die Lage des Kindes automatisch mit reinzusetzen“ (Corgi-pFK, Abs. 91) (Mechanismus auf affektiv-motivationaler Ebene).

Um darauf aufbauend die Veränderungswünsche der Kinder umzusetzen, erarbeiteten die Projektmitarbeiter:innen in der Prozessbegleitung mit den Fachkräften gemeinsam verschiedene Handlungsoptionen (Aktivität). Die Fachkräfte integrieren diese als neue Handlungsmöglichkeiten (Outcome) in ihr bisheriges Handlungsrepertoire (Mechanismus auf kognitiv-reflexiver Ebene). Mit Blick auf die Frage nach sichtbaren Veränderungen bzw. Wirkungen probierten sich die Fachkräfte in neuen Handlungsweisen (Mechanismus auf praktisch-instrumenteller Ebene) und verstärkten sowohl aus eigener Sicht als auch aus Sicht der Hortleitung und des Modellprojekts ihre Bemühungen, Kinder aktiv in Entscheidungen einzubeziehen und an Gestaltungsprozessen zu beteiligen (Outcome): „Dahingehend haben wir [...] mit den Bedingungen, die wir haben, [...] die Gruppenräume ein bisschen umstrukturiert, auch nach den Wünschen der Kinder“ (Corgi_Hortleitung, Abs. 70) (weiterführend Braun et al. 2023). Zudem begannen die Fachkräfte mithilfe des Modellprojekts, die routinierten Abläufe im Hortalltag zu hinterfragen und fortan stärker an kindlichen Bedürfnissen zu orientieren (Outcome).

In den Termini der Realist Evaluation gefasst, bedeutet das für das Hort-Cluster: Durch die Top-Flop-Methode (Aktivität) konnten die Annahmen bzw. Wahrnehmungen der Fachkräfte bezüglich unbeliebter Orte der Kinder im Hort (Kontext) irritiert werden (Mechanismus). Sie reflektierten und revidierten ihre

Annahmen darüber, erwarben neues Wissen zu den Sichtweisen der Kinder und entwickelten so ein sensibleres Bewusstsein für die Perspektive der Kinder sowie neue Handlungsoptionen (Outcome).

5.2.2.5 Exemplarische Ergebnisdarstellung – Beispiel für die Aggregation von Befunden auf Clusterebene

Im Ergebnis der multiperspektivischen Wirkungsanalysen werden über kontrastierende Paarvergleiche projektübergreifend Schlussfolgerungen in vereinfachender Form über die Wirkungsweisen der demokratiefördernden Strategie sowie über zentrale Scheiterns- bzw. Gelingensbedingungen innerhalb der verschiedenen Cluster für das Handlungsfeld abgeleitet.

Der Paarvergleich im Hort-Cluster ergibt, dass beide Modellprojekte unter mannigfachen, teils schwierigen Bedingungen (vor allem Personalmangel, Ressourcenknappheit) arbeiteten und insbesondere eine Unterstützungs- und Kompensationsfunktion für die überlasteten Fachkräfte im Hort erfüllten.

Beide Modellprojekte fokussierten auf kognitiv-reflexive Wirkmechanismen (Irritieren bzw. Fundieren von Wissensbeständen, Perspektivwechseln) für Lern- und Veränderungsprozesse auf der Individualebene bei ihren Zielgruppen, die mit der affektiv-motivationalen Ebene (Entlastungserleben, Motivieren, Bestärken) sowie der praktisch-instrumentellen Ebene (Aktivieren und Stärken neuer Handlungsoptionen) verbunden sind. Ein Modellprojekt bewirkte eher Veränderungen in der Haltung und im Handeln der Fachkräfte, das andere eine Stabilisierung und Bestärkung von vorhandenen Haltungen und Handlungsabläufen.

Bei beiden Modellprojekten war es nicht möglich, umfangreiche strukturelle Veränderungen in den beobachteten Einrichtungen festzustellen, die auf die Projektinterventionen zurückführbar gewesen wären. Das lag zum Teil an fehlenden Zeit- und Personalressourcen (Kontextfaktoren) in den untersuchten Einrichtungen im Untersuchungszeitraum. Durch den begrenzten Erhebungszeitraum blieben zudem weitere potenzielle Wirkungsdimensionen (z. B. gestärkte Position des Hortteams beim Lehrkollegium, Etablierung von Mitbestimmungsgremien für Kinder) außerhalb des Blickfeldes der wissenschaftlichen Begleitung.

5.3 Herausforderungen im Vorgehen und Umgangsweisen

In Wirkungsevaluationen treten in den verschiedenen Phasen des Forschungsprozesses verschiedene Herausforderungen auf, die vor allem aus der Komplexität des Gegenstands resultieren und die Identifikation von Wirkungen und deren Plausibilisierung sowie die Generalisierbarkeit von projektspezifischen Wirkausagen für das Handlungsfeld betreffen. Diese Herausforderungen und ihre Konsequenzen für den Forschungs- bzw. Evaluationsprozess sowie einen angemess-

senen Umgang mit ihnen sollen nachfolgend vor dem Hintergrund der Umsetzungserfahrungen in der hier vorgestellten Wirkungsevaluation kritisch diskutiert werden.

5.3.1 Allgemeine Herausforderungen

Als eine immerwährende Herausforderung erweist sich die *Komplexität des Gegenstands*¹⁰ (Haunberger/Baumgartner 2017). Evaluierende bewegen sich beim Versuch, Gegenstandsangemessenheit zu berücksichtigen, notwendigerweise im Spannungsfeld zwischen Strategien der Annäherung an Komplexität (u. a. über Kontextsensibilität¹¹, Multiperspektivität, Multimethodik im Feldzugang) und notwendigen Reduktionsstrategien, um komplexe Gegenstände handhabbar zu machen (u. a. durch die Auswahl von Clustern, Festlegung und Verengung von Wirkausschnitten, vereinfachte Abbildung von Wirkzusammenhängen).

Die größte Herausforderung in der Wirkungsevaluation im hier betrachteten Handlungsfeld ist die *Identifikation von Wirkungen sowie deren Plausibilisierung*. Diese resultiert zum einen aus der Realität komplexer Wirkungszusammenhänge in (sozial-)pädagogischen Interventionen, zum anderen aus dem Prozess der Plausibilisierung von Wirkaussagen anhand des empirischen Materials auf Basis qualitativer Methoden selbst, jenseits von *Kausalitätsnachweisen*. Um sich Wirkungen theoretisch und empirisch anzunähern, eignet sich aus Sicht der Autor:innen dieses Beitrages die Verwendung der Heuristik der Realist Evaluation, die relevante Kontextfaktoren, Mechanismen, Outcomes sowie Aktivitäten in einen Zusammenhang setzt (siehe 5.2.2.4 und 5.2.2.5, ausführlicher dazu Braun et al. 2023). Gestützt auf Multiperspektivität und multimethodische Zugänge zum Feld können begründete Aussagen über die Wirksamkeit einer Maßnahme getroffen und deren Aussagekraft anhand von Plausibilitätsgraden abgestuft beurteilt werden.

Eine weitere zentrale Herausforderung besteht im *Anspruch der Generalisierung* von Wirkbefunden für das Handlungsfeld, also die Übertragung auf andere, nicht untersuchte Modellprojekte. Über die Anwendung des Ansatzes der Cluster-Evaluation kann zwar eine begründete Auswahl und Clusterung von Projekten erfolgen, die für verschiedene Demokratieförderstrategien in zentralen Umsetzungskontexten stehen. Dennoch bleibt es aus unserer Sicht ein ungelöstes Problem,

10 Sie betrifft nicht nur die Ebene des Handlungsfeldes und die Projektebene, sondern auch die Ebene der Einzelmaßnahmen. Auch in der jeweils beobachteten sozialen Situation der pädagogischen Praxis stellen komplexe Vorgänge und beeinflussende Faktoren eine Herausforderung dar.

11 Für die Umsetzung einer kontextsensiblen Wirkungsevaluation ist die angemessene Berücksichtigung der Umsetzungskontexte als komplexes soziales Gefüge für Projektmaßnahmen wichtig. Im vorliegenden Fall konnte eine Kontextanalyse für institutionelle Settings im Grundschulalter (Braun et al. 2023) im Vorfeld der Wirkungsevaluation durchgeführt werden.

die kontext- und inhaltspezifischen Wirkaussagen auf Modellprojekt- und Clusterebene auf nicht untersuchte Modellprojekte im Handlungsfeld mit ähnlichen inhaltlichen Schwerpunkten in möglicherweise vergleichbaren, aber nie identischen Umsetzungskontexten zu beziehen.¹²

5.3.2 Umgang mit Herausforderungen in den Phasen des Forschungsprozesses

Die *Phase der Konzeption* der Wirkungsevaluation und des Erhebungsdesigns beinhaltete ein theoretisch begründetes Vorgehen, als Ausgangspunkt für eine gegenstandsangemessene Untersuchung von Wirkungen. Die Prämissen der Cluster-Evaluation und das Konzept der Realist Evaluation (siehe Abschnitt 5.2) dienten hier zur Orientierung, um die Komplexität und Mehrdimensionalität des Untersuchungsgegenstands (Haunberger/Baumgartner 2017, S. 123) zu meistern. Auslegungsspielraum bezüglich der Cluster-/Gruppierungslogik, als Alternative zu der hier vorgestellten Variante, sehen die Autor:innen darin, die Modellprojektarbeit entlang von weiteren Vergleichsdimensionen zu gruppieren (z. B. Innovations- bzw. Transferhandeln).

Die wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld Demokratieförderung hat sich sehr streng an die Vergleichbarkeitsprämisse der Cluster-Evaluation für die Projekte innerhalb der Cluster gehalten, mit in der Folge restriktiven Vorgaben für die Auswahl der Modellprojekte. Das ging teils auf Kosten der Beschreibung von besonders innovativen Projektansätzen, teils war es im hier vorgestellten Fall mit Folgen für die Aussagekraft für das Handlungsfeld verbunden.

Weiterhin ist für den Umgang mit der Komplexität des Gegenstands in der Konzeptionsphase die Festlegung bzw. Auswahl eines geeigneten Wirkausschnitts notwendig (Lüders/Haubrich 2006). Damit ist der Anspruch verbunden, dass dieser Wirkausschnitt einerseits den Kriterien der Cluster-Evaluation genügt und andererseits in der Erhebungsphase einen ressourcenschonenden Personal- und Zeiteinsatz ermöglicht. Über die Engführung des Wirkausschnitts blieben im Rahmen der Erhebung jedoch jene potenziellen Wirkungen unberücksichtigt, die nicht zum festgelegten Wirkausschnitt gehörten. In dem hier betrachteten Projekte-Cluster waren das zum Beispiel – wie bereits erwähnt

12 Wirkgefüge können sich unter anderen kontextuellen Umständen, etwa wenn Projekte mit anderen Projektansätzen und Akteurskonstellationen gearbeitet haben, anders darstellen. Streng genommen gelten daher Wirkbefunde, vor allem auf Clusterebene, für die untersuchten Modellprojekte und nur bedingt darüber hinaus. Dennoch liefern sie für andere Modellprojekte Ergebnisse zu Gelingens- und Scheiternsbedingungen, die von anderen Trägern bei der Umsetzung ihrer Angebote berücksichtigt werden können.

– mögliche strukturelle Veränderungen in den Einrichtungen, in denen die Modellprojekte Einzelmaßnahmen umsetzen.

Für das Vorhaben der Plausibilisierung von Wirkungen über die Berücksichtigung von Multiperspektivität ist ein Perspektivensampling sinnvoll. Alle gewünschten bzw. erforderlichen Perspektiven einzuholen, kann sich im *Erhebungsprozess* jedoch als schwierig herausstellen. Für die Erhebungsphase können beispielsweise durch begrenzte Erhebungszeitfenster Schwierigkeiten beim Feldzugang und in der Erhebungsdurchführung resultieren. Auch kann der Projektumsetzungsstand beispielsweise nicht ausreichend fortgeschritten sein. Zudem müssen die Beteiligten in dieser Phase erreichbar und zum Interview bereit sein. Um all das sicherzustellen, hat sich neben der Festlegung des Wirkausschnittes und der gemeinsamen Erhebungsplanung mit den Projektverantwortlichen als nützlich erwiesen, die Erhebung selbst am Projektfortschritt zu orientieren.

Mit Blick auf die gewählten Erhebungsinstrumente können wir aufgrund unserer Erfahrungen in der wissenschaftlichen Begleitung qualitative Experteninterviews und teilnehmende Beobachtungen als sinnvoll und passfähig für die Identifikation von Wirkzusammenhängen einstufen. Gruppeninterviews als Erhebungsmethode haben sich hingegen nur teilweise bewährt. Im hier vorgestellten Fall nahmen sie häufig einen konsensorientierten Verlauf und ermöglichten dadurch seltener tiefergehende Einblicke in Wirkzusammenhänge.

Was die *Auswertungsphase* angeht, hat sich die Anwendung der Realist Evaluation, genauer des (A)CMO-Schemas, als heuristischer Rahmen zur Orientierung bewährt. Die Auswertung mit verschiedenen Interpretationsverfahren sowie die Triangulation der Ergebnisse konnten die Aussagen zu Wirkungen absichern. Für plausible Wirkaussagen auf Einzelprojektebene können, wie schon angedeutet, Konvergenzen – übereinstimmende Aussagen aus verschiedenen Perspektiven zum Wirkausschnitt – herangezogen werden. Divergenzen zwischen den erhobenen Perspektiven deuten auf ebenfalls spannende Befunde hin, die jedoch weniger abgesichert sind und daher als weniger plausibel eingestuft werden sollten. Gleiches gilt unseres Erachtens für Befunde, die nur über ein Auswertungsverfahren abgesichert werden konnten. Die Befunde zur Verschränkung der drei Ebenen, auf denen Wirkungen erfasst werden konnten, „kognitiv-reflexiv“, „affektiv-motivational“ sowie „praktisch-instrumentell“ beim Projekt Corgi, mit einem Wirk-Schwerpunkt auf der kognitiv-reflexiven Ebene, konnten mit diesem Vorgehen hinreichend plausibel begründet werden.

Eine Herausforderung für die Anwendung der Realist Evaluation stellten darüber hinaus nicht nur die nicht eindeutig bestimmten Kategorien, sondern auch deren unterschiedliche Verwendung in der Literatur dar. Das betrifft insbesondere die vage und vielfältige Auslegung des Mechanismus-Begriffs (u. a. Lemire et al. 2020). Auch die Identifikation und Abgrenzung der Komponenten voneinander gilt als Dilemma (Lacouture et al. 2015; Haunberger et al. i. V.). Im Um-

gang damit hat sich zunächst die Festlegung von Arbeitsdefinitionen für die Begriffe „Kontext“, „Mechanismen“, „Outcome“ und „Aktivität“ clusterübergreifend bewährt. Diese Begriffe wurden – wie bereits geschildert – in einem deduktiv-induktiv strukturierten Inhaltsanalyseschema weiterverarbeitet. Außerdem haben eine gemeinsame Reflexion der Anwendung des Codier-Schemas sowie Ankerbeispiele für Mechanismen und Outcomes dabei geholfen, Wirkungen und Mechanismen zu identifizieren sowie Intercoder-Reliabilität herzustellen. Dennoch blieb vor allem die analytische Abgrenzung des Begriffs „Mechanismus“ von den anderen Modellkomponenten und das Auffinden von Mechanismen im Material herausfordernd (so auch Lacouture et al. 2015).

Dennoch ist es uns durch die Verwendung verschiedener Auswertungsmethoden und die Triangulation der jeweiligen Befunde im nächsten Schritt gelungen, Mechanismen herauszuarbeiten. Mit der Dokumentarischen Methode war es möglich, die im Rahmen der Inhaltsanalyse identifizierten Schlüsselstellen, die potenziell CMO-Konfiguration enthielten, sequenziell zu interpretieren und CMO-Konfigurationen zu rekonstruieren (zum Vorgehen Nohl 2017). Die Dokumentarische Methode für die Identifikation von Mechanismen auf Gruppeninterviews anzuwenden, erwies sich nur für einige Modellprojekte als sinnvoll und ist stark abhängig vom Fragestimulus sowie von der Gruppenzusammensetzung. Empirische Befunde, auch in anderen als dem in diesem Beitrag vorgestellten Beispiel-Cluster, zeugen darüber hinaus von komplexen CMO-Zusammenhängen. Dabei können Komponenten mehrmals auftreten (z. B. CCMOO) oder ihre Abfolge variieren (z. B. CMCMO; auch Bergeron/Gaboury 2020, S. 3). Diesbezüglich ist von multiplen, miteinander verschränkten Wirkprozessen auszugehen.

In der *Phase der Ergebnisdarstellung und Veröffentlichung* bestand die Herausforderung darin, das umfangreiche Befundmaterial zu verdichten und die zentralen Befunde nachvollziehbar darzulegen. Hierfür eignen sich vereinfachende Darstellungen von Wirkmodellen mit ergänzenden Erläuterungen der Komponenten und Wirkprozesse (Braun et al. 2023, 2024).

5.4 Fazit und Diskussion

Bilanziert man das Wirkungsstudiendesign für Modellprojekte der Demokratieförderung anhand der Umsetzungserfahrungen in der wissenschaftlichen Begleitung des Handlungsfeldes und der geschilderten Herausforderungen, kann man zu folgendem Schluss gelangen: Konzeptionell sollte eine Wirkungsevaluation demokratiefördernder Praxis theoretisch begründet hergeleitet und gegenstandsbezogen sowie kontextsensibel ausgerichtet sein. Mittels Perspektivensampling und einem multimethodischen Herangehen kann gezielt der Fragestellung nachgegangen werden, in welchem Kontext, unter welchen Bedingungen, bei wem und warum eine Aktivität zu (un-)beabsichtigten Veränderungen

gen (Wirkungen) führt. Für die Entwicklung eines gegenstandsangemessenen Untersuchungsdesigns hierfür bieten die Prämissen der Cluster-Evaluation (Haubrich 2006, Worthen/Schmitz 1997) sowie die begriffliche Heuristik der Realist Evaluation (Pawson/Tilley 2004) wichtige Anhaltspunkte. Erforderlich für ihre Anwendung in vergleichbaren Gegenstandsbereichen wie dem unseren sind aus Sicht der Autor:innen neben begrifflich-konzeptionellen Weiterentwicklungen des Ansatzes der Realist Evaluation (u. a. durch Ergänzung des CMO-Schemas um „Aktivitäten“ im Sinne einer A-CMO-Konfiguration), die Ausdeutung von Prämissen sowie die methodische Erprobung verschiedener Anwendungsvarianten der Cluster-Evaluation.

In der praktischen Umsetzung hat sich bei der Auswahl der Untersuchungsfälle das systematische, clusterevaluative Sampling auf Basis der Projektmodelle bewährt. Darüber gelang ein aus unserer Sicht für den Evaluationszweck ausreichend tiefgehender Einblick in die Projektarbeit und -erträge im Handlungsfeld. Mithilfe von Wirkmodellierungen ist es möglich, die Wirkerwartungen der Modellprojekte gemeinsam zu reflektieren, sie zu konkretisieren und anschließend in Wirkuntersuchungen Erkenntnisse über tatsächliche Wirkzusammenhänge zu gewinnen.

Damit das aufwendige, ressourcenintensive Vorhaben von Wirkungsevaluationen (auch Bohn/Saßmannshausen 2021) für alle Seiten transparent und trotz des Aufwands als gewinnbringend eingestuft wird, gilt es, die Projektpraxis frühzeitig in das Vorhaben einzubinden sowie Wirkausschnitte und Feldzugänge eng miteinander abzustimmen. Dieses Vorgehen hat im Fall der von uns untersuchten Modellprojekte für Offenheit gesorgt und Abwehrhaltungen gegenüber dem Evaluationsanliegen vermieden. Die Projektpraxis und ihre Adressat:innen ermöglichen tiefe Einblicke in ihre Arbeit. Zugleich wurden so Reflexionsanlässe im Sinne einer formativen, die Projekte zum Lernen anregenden Evaluation geschaffen.

Als sinnvoll hat sich zudem das Vorgehen erwiesen, Projektpaare innerhalb eines Kontextes bzw. in ähnlichen Kontexten kontrastierend zu vergleichen. Durch den Paarvergleich sind projektübergreifende, plausible Wirkaussagen sowie Aussagen zu relevanten Kontextfaktoren, zu erprobten Inhalten, Aktivitäten und zu pädagogischen Standards gut rekonstruierbar. Dabei ermöglichte das gewählte Forschungsdesign eine abgestufte Plausibilisierung von Befunden auf der Ebene der Modellprojekte und Cluster: Eine hohe Plausibilität leitet sich im betrachteten Fall aus der Übereinstimmung mehrerer Perspektiven ab und ist abhängig davon, wer sich zu einem Sachverhalt äußert. Widersprüche in der Wahrnehmung und Zurechnung von Veränderungen deuten auf eine geringere Plausibilität hin.

Eine Übertragung bzw. Verallgemeinerung der in einem Cluster gewonnenen Befunde auf andere Modellprojekte im Handlungsfeld ist hingegen ohne weitere Daten und Erhebungszeitpunkte nicht möglich. Das hängt mit dem starken Gegenstandsbezug der durchgeführten Untersuchungen zusammen. Die Wirk-

aussagen beziehen sich zum einen auf einzelne, spezielle Befunde zu Einzelmaßnahmen. Zum anderen resultiert daraus eine zwar umfangreiche, jedoch auf den Wirkausschnitt begrenzte Datengrundlage für die Plausibilisierung. Die auf den engen Wirkausschnitt bezogenen Aussagen aus dem Perspektivenabgleich gelten somit in erster Linie als plausibel für die jeweils betrachteten Modellprojekte bzw. für das untersuchte Cluster.

Schwächen weist das gewählte Evaluationsdesign mit Blick auf die produzierte Datenmenge, die Identifikation von Mechanismen bzw. auf das Zusammenspiel der ACMO-Komponenten auf. Um Multiperspektivität für die Plausibilisierung und zur Vermeidung von Fehlschlüssen zu gewährleisten, wurden Datenmengen produziert, die zwar notwendig, aber zugleich umfangreich sind und einen hohen Auswertungsaufwand bedeuten. Trotz dieser Datengrundlage blieben insbesondere die Verbindungen und Wechselwirkungen zwischen den einzelnen Komponenten der ACMO-Konfigurationen mitunter unkonkret.

Um die Datengrundlage für die Rekonstruktion dieser Konfigurationen zu verbessern, bietet es sich zukünftig beispielsweise an, den methodischen Vorschlag von Ana Manzano (2016) aufzugreifen und Befragungen entlang der Projekttheorie (Wirkmodell) zu planen. Dadurch kann gewährleistet werden, dass sich alle relevanten Akteur:innen mit ihren Perspektiven zu den wesentlichen Sachverhalten äußern. Auch könnte man alternative Verfahren der Datenerhebung und Auswertung erproben, um dem Zusammenspiel von Wirkkomponenten auf den Grund zu gehen (z. B. Orientierung an Grounded Theory, vgl. Figlestahler/Haase/Schau in diesem Band).

Darüber hinaus könnte es für die Verbesserung des Plausibilisierungsprozesses hilfreich sein, dafür eine eigene Strategie zu entwerfen, die mehrere Plausibilisierungsschritte umfasst und Bestandteil des Evaluationskonzepts ist (Ackermann 2016). So ließe sich die Absicherung bzw. Plausibilisierung von Befunden durch ein iteratives Forschungsdesign bei der Identifikation von Wirkzusammenhängen verbessern, das heißt, unter anderem durch weitere Erhebungs- und Auswertungsmethoden, durch mehrere Erhebungszeitpunkte und durch weitere Konkretisierung von identifizierten ACMO-Konfigurationen im Auswertungsprozess. Um plausible Aussagen für ein Handlungsfeld als Ganzes zu erhalten, sind zudem alternative Auslegungen der Prämissen der Cluster-Evaluation zu überlegen.

Abschließend bleibt auf Basis unserer Erfahrungen festzuhalten, dass Wirkungsevaluationen als besondere Art der Erforschung von komplexen sozialen Realitäten in vielfältiger Weise herausfordernd sind. Sie unterliegen nicht nur materiellen, sondern auch zeitlichen Restriktionen. Was Letztere angeht, geraten in Untersuchungen, die sich auf einen Erhebungszeitpunkt während eines laufenden Projekts beschränken (müssen), nur kurzfristige Erträge in den Blick. Langfristige Wirkungen bleiben außen vor. Will man diese erfassen, benötigen Wirkungsevaluationen einen flexibleren, deutlich größeren zeitlichen und

entsprechend personell wie finanziell angepassten Rahmen als im vorgestellten Fall zur Verfügung stand. Als notwendig und sinnvoll erachten wir ein Untersuchungsdesign, das es ermöglicht, Wirkungen kontinuierlich über mehrere Jahre hinweg auch auf anderen als der individuellen Ebene zu erfassen und beispielsweise Aussagen unter anderem zu Veränderungen von Strukturen, Verfahrensweisen bzw. Interaktionskulturen im Bereich der Demokratieförderung zu untersuchen.

Literatur

- Ackermann, G. (2016): Evaluation und Komplexität. Wirkungskonstruktion in der Evaluation von Gesundheitsförderung und Prävention. Dissertation. Rheinfelden: Universität Basel. edoc.unibas.ch/41047 (Abruf 03.06.2024).
- Albus, S./Micheel, H.-G./Polutta, A. (2011): Der Wirkungsdiskurs in der Sozialen Arbeit und seine Implikationen für die empirische Sozialforschung. In: Oelerich, G./Otte, H.-U. (Hrsg.): Empirische Forschung und Soziale Arbeit. Ein Studienbuch. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 243–251.
- Bergeron, D. A./Gaboury, I. (2020): Challenges Related to the Analytical Process in Realist Evaluation and Latest Developments on the Use of NVivo from a Realist Perspective. In: International Journal of Social Research Methodology, 23. Jg., H. 3, S. 355–365.
- Bischoff, U./König, F./Zimmermann, E. (2018): Pädagogik wirkt?! Wirkungsevaluationen in der pädagogischen Präventionsarbeit. In: Glaser, M./Frank, A./Herding, M. (Hrsg.): Gewaltorientierter Islamismus im Jugendalter. Perspektiven aus Jugendforschung und Jugendhilfe. 2. Sonderband Sozialmagazin. Weinheim und Basel: Beltz Juventa, S. 147–157.
- Bischoff, U./Zimmermann, E./König, F. (2021): Erkennen was wirkt. Die Erprobung von Ansätzen der Wirkungsuntersuchung in der Evaluation von Bundesmodellprogrammen der Demokratieförderung und Extremismusprävention und die damit gemachten Erfahrungen. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluation von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 244–268.
- Bohn, I./Saßmannshausen, J. (2021): Wirkungen komplexer Programm-Möglichkeiten und Grenzen evidenzbasierter Evaluation am Beispiel der Partnerschaften für Demokratie. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 269–295.
- Braun, M./Ehnert, K./Euteneuer, J./Hemmann, M./Johann, T./Rehse, A./Rüger, S./Walter, E. (2023): Demokratieförderung im Kindesalter: Kontext- und Wirkungsanalysen. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von H. Bilgili. Halle (Saale): DJI.
- Braun, M./Ehnert, K./Euteneuer, J./Hemmann, M./Johann, T./Rehse, A./Rüger, S./Walter, E. (2024): Demokratieförderung im Jugend- und jungen Erwachsenenalter – Wirkungsanalysen. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von H. Bilgili. Halle (Saale): DJI.
- DeGEval – Gesellschaft für Evaluation (2017): Standards für Evaluation. Erste Revision 2016. https://www.degeval.org/fileadmin/content/Z03_Publikationen/DeGEval-Standards/DeGEval-Standards_fuer_Evaluation.pdf (Abruf 27.06.2024).

- Ehnert, K./Johann, T./Mielke, A./Rehse, A./Walter, E./Zimmermann, E. (2021): Handlungsfeld und Handlungslogiken. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld Demokratieförderung im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von C. Wagner. Halle (Saale): DJI.
- Flick, U. (2011): Triangulation. Eine Einführung. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Haubrich, K. (2006): Wirkungsannahmen sichtbar machen: Cluster-Evaluation innovativer multi-zentrischer Programme. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 101–122.
- Haunberger, S./Baumgartner, E. (2017): Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview. In: Zeitschrift für Evaluation, 16. Jg., H. 1, S. 121–145.
- Haunberger, S./Braun, M./Hemmann, M./Rehse, A. (i. V.): Wirkmechanismen auf der Spur. Zwei Anwendungsbeispiele der Realist Evaluation zur qualitativen Identifikation von Wirkmechanismen.
- Hense, J./Taut, S. (2021): Wie Wirkungsmodelle zur Wirkung kommen. Nutzungsvarianten und Kosten ihrer Verwendung in der Evaluationspraxis. In: Zeitschrift für Evaluation, 20. Jg., H. 2, S. 267–292.
- Himmelman, G./Lange, D. (Hrsg.) (2005): Demokratiekompetenz. Beiträge aus Politikwissenschaft, Pädagogik und politischer Bildung. Wiesbaden: Springer VS.
- Klawe, W. (2006): Wie wirken pädagogische Interventionen? Zur Wirkungsforschung bei individualpädagogischen Maßnahmen. In: erleben und lernen, H. 3/4, S. 6–11.
- Lacouture, A./Breton, E./Guichard, A./Ridde, V. (2015): The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. In: Implementation Science, 10 (Art. 153). <https://doi.org/10.1186/s13012-015-0345-7> (Abruf 03.06.2024).
- Langner, J./Schott, M./Grunow, D. (2021): Wirkmechanismen pädagogischer Praxis der Radikalisierungsprävention. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluationen von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 222–243.
- Lemire, S./Kwako, A./Nielsen, S. B./Christie, C. A./Donaldson, S. I./Leeuw, F. L. (2020): What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations. In: New Directions for Evaluation, H. 167, S. 73–86.
- Loosen, W. (2016): Das Leitfadeninterview – eine unterschätzte Methode. In: Averbek-Lietz, S./Meyen, M. (Hrsg.): Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft. Wiesbaden: Springer VS, S. 139–155.
- Lüders, C. (2015): Beobachten im Feld und Ethnographie. In: Flick, U. (Hrsg.): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt, S. 384–401.
- Lüders, C./Haubrich, K. (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe. Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 5–23.
- Manzano, A. (2016): The craft of interviewing in realist evaluation. In: Evaluation, 22. Jg., H. 3, S. 342–360.
- Mayring, P. (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12., überarbeitete Auflage. Weinheim und Basel: Beltz.
- Nohl, A.-M. (2017): Interview und Dokumentarische Methode. Anleitungen für die Forschungspraxis. Wiesbaden: Springer VS.
- Ottmann, S. (2023): Wirkungsorientierung in der Sozialen Arbeit – Herausforderung und Chance für das Arbeitsfeld. In: Ottmann, S./König, J. (Hrsg.): Wirkungen im Blick – Wirkungsorientie-

- rung in der Sozialen Arbeit und Sozialwirtschaft. Tagungsdokumentation der Transferkonferenz. Nürnberger Hochschulschriften Nr. 53. Nürnberg: Evangelische Hochschule, S. 6–15.
- Pawson, R./Tilley, N. (1997): Realistic evaluation. Thousand Oaks, CA: Sage.
- Pawson, R./Tilley, N. (2004): Realist Evaluation. https://www.communitymatters.com.au/RE_chapter.pdf (Abruf 11.03.2019).
- Przyborski, A./Wohlrab-Sahr, M. (2021): Qualitative Sozialforschung. Ein Arbeitsbuch. 5., überarbeitete und erweiterte Auflage. Berlin/München/Boston: de Gruyter.
- Rehse, A./Johann, T. (2022): Demokratieförderung vs. Politische Bildung – Versuch einer empiriebasierten Verhältnisbestimmung. In: Journal für Politische Bildung, 12. Jg., H. 3, S. 4–7.
- Sturzenhecker, B./Wohnig, A. (2019): Begriffsvielfalt, Entgrenzung, Aufmerksamkeitskultur. Kommentare zur neuen Unübersichtlichkeit auf dem Arbeitsfeld der politischen Bildung. In: Journal für Politische Bildung, 9. Jg., H. 2, S. 10–15.
- Worthen, B. R./Schmitz, C. C. (1997): Conceptual Challenges Confronting Cluster Evaluation. In: Evaluation, H. 3, S. 300–319.

6 „Das hätte ich, dann jetzt rückblickend betrachtet, noch anders gestaltet“

Zur Auseinandersetzung mit Lernen und Scheitern in wirkungsorientierter Evaluation von Modellprojekten zur Demokratieförderung

Jonas Euteneuer und Stella Rürger

6.1 Einleitung

Als wissenschaftliche Begleitung (wB) der Modellprojekte im Handlungsfeld Demokratieförderung des Bundesprogramms „Demokratie leben!“ haben wir im Rahmen unseres Evaluationsauftrags in den Förderjahren 2022 und 2023 multiperspektivische Wirkungsuntersuchungen durchgeführt. Primär haben wir uns dabei mit Wirkungen auf der Ebene der Adressat:innen der Projekte auseinandergesetzt. Ausgehend von der Realist Evaluation (Pawson/Tilley 1997) und dem Verfahren der Clusterevaluation (Haubrich 2006, 2009) standen dabei ausgewählte Modellprojekte des Handlungsfeldes Demokratieförderung im Fokus. In den Interviews, die wir im Rahmen der Wirkungsstudien mit verschiedenen beteiligten Stakeholdern geführt haben, bekamen wir auch Einblicke in Modifikationen in der Modellprojektkonzeption und -umsetzung. An verschiedenen Stellen im Material berichten insbesondere Mitarbeitende von Lernprozessen, die sich auf die Umsetzung der Modellprojektarbeit beziehen. Hierbei handelt es sich also um Prozesse, in denen die Modellprojektmitarbeitenden selbst dazulernen, wodurch sich ihre fachliche Praxis weiterentwickelt oder verändert. Sie machen dabei auch Erfahrungen von (graduellem) Scheitern transparent, welche als Grundlage für solche Lernprozesse gefasst werden können (Kessler/König 2024, S. 10f.).

Im Vorgehen sowie in der Ergebnisdarstellung unserer Wirkungsuntersuchungen standen diese Prozesse des Lernens und Scheiterns weniger im Mittelpunkt. Wir halten es aber für lohnenswert, Lern- und Scheiternserfahrungen systematischer in den Blick zu nehmen und im Sinne einer „[f]ortgesetzte[n] Justierung“ (Strübing et al. 2018, S. 86) der Gegenstandsangemessenheit unseres methodischen Vorgehens reflexiv danach zu fragen, inwiefern solche Erfahrungen für eine wissenschaftliche Begleitung zugänglich gemacht werden können. Am Beispiel der Interviews zeigt sich: Ob wir als wB entsprechende Einblicke bekommen, hängt davon ab, wie ausführlich die Interviewten von sich aus über Scheiternserfahrungen und daraus resultierende Lernprozesse in der Modell-

projektarbeit sprechen – und wie eine wirkungsorientierte Evaluation einen geeigneten Rahmen dafür schaffen kann.

Die Auseinandersetzung mit Lernen und Scheitern in der Modellprojektarbeit erscheint uns nicht nur vor dem Hintergrund des empirischen Materials bedeutsam, sondern insbesondere aufgrund des Gegenstands, mit dem wir uns im Rahmen der Wirkungsstudien empirisch beschäftigt haben. Der Auftrag der von uns wissenschaftlich begleiteten Modellprojekte im Handlungsfeld Demokratieförderung ist es, „neue, innovative Ansätze für die präventiv-pädagogische Arbeit mit Kindern, Jugendlichen, jungen Erwachsenen und deren Eltern sowie für die Arbeit von Akteurinnen und Akteuren der Kinder- und Jugendhilfe [zu entwickeln]“ (BMFSFJ 2023, o. S.). Die Methoden und Materialien der Modellprojekte sollen „nach erfolgreicher Erprobung in der pädagogischen Praxis der Kinder- und Jugendhilfe sowie in anderen Regelstrukturen eingesetzt werden“ (ebd.). Modellprojekte agieren dementsprechend „in einem Experimentierraum“ (Ehnert et al. 2021, S. 7), in dem „die Möglichkeit des Scheiterns inhärent und gewollt [ist]“ (ebd.).

Der Überlegung folgend, dass diese Charakterisierung des Gegenstands auch in der Auseinandersetzung mit Wirkungen von Modellprojekten bedeutsam ist, möchten wir im Beitrag zwei Fragestellungen miteinander in Bezug setzen: Es soll erstens im Fokus stehen, *wie Lern- und Scheiternerfahrungen der Modellprojekte im Rahmen eines an der Realist Evaluation orientierten Vorgehens empirisch systematischer erfasst bzw. berücksichtigt werden können*. Daran anschließend möchten wir zweitens die Frage diskutieren, *welches Potenzial ein systematischerer Einbezug von Lern- und Scheiternerfahrungen und darauffolgender Modifikationen für eine wirkungsorientierte Evaluation von Modellprojekten und für deren Auftrag zur Entwicklung von (erfolgreicher) Praxis haben kann*.

Im Folgenden (Abschnitt 6.2) stellen wir zunächst jene Punkte vor, die unsere Auseinandersetzung rahmen. Neben Informationen zum Handlungsfeld Demokratieförderung im Bundesprogramm „Demokratie leben!“ und der wB von Modellprojekten rücken wir hier insbesondere die von uns durchgeführten Wirkungsstudien in den Blick. Anschließend (Abschnitt 6.3) befassen wir uns mit den Möglichkeiten der Auseinandersetzung mit Lernen und Scheitern in Wirkungsstudien. Wir legen zunächst unser Verständnis von Lernen und Scheitern dar und bieten Einblicke in zwei empirische Beispiele dafür im Rahmen der Arbeit der von uns begleiteten Modellprojekte. Dann diskutieren wir mögliche Anschlusspunkte in unserem Vorgehen, die eine stärkere Berücksichtigung von Lernen und Scheitern ermöglichen, und gehen auf methodische Erfordernisse für diese stärkere Berücksichtigung ein, setzen uns aber auch mit den Grenzen des Vorgehens auseinander. Abschnitt 6.4 umfasst Überlegungen zur Frage des Potenzials des systematischeren Einbezugs von Lernen und Scheitern, die sich von dem engen Bezug auf unser Vorgehen in den Wirkungsstudien wegbewegen und den Blick auf den Transfer von Lern- und Scheiternerfahrungen lenken. Wir schließen un-

seren Beitrag dann mit einigen abschließenden Überlegungen zur Relevanz der Auseinandersetzung mit Lernen und Scheitern in Wirkungsstudien ab (Abschnitt 6.5).

6.2 Gegenstand Modellprojekte, Handlungsfeldevaluation und Vorgehen in den Wirkungsstudien

Als Rahmung unserer Überlegungen geben wir nachfolgend einen kurzen Einblick in das Handlungsfeld „Demokratieförderung“ des Bundesprogramms „Demokratie leben!“, gehen auf einige Charakteristika des Gegenstands „Modellprojekte“ ein und skizzieren die Anforderungen, die sich an die wB von Modellprojekten stellen. Im Anschluss legen wir unser Verständnis von Wirkungen dar und stellen das empirische Vorgehen unserer Wirkungsstudien vor.

6.2.1 Modellprojekte im Handlungsfeld „Demokratieförderung“ und Aufgabe der wB

Das Handlungsfeld „Demokratieförderung“ wurde mit der zweiten Förderphase des Bundesprogramms „Demokratie leben!“ (BMFSFJ) als eigenständiges Handlungsfeld für (pädagogische) Modellprojekte neben dem der Vielfaltgestaltung und Extremismusprävention im Jahr 2020 neu eingerichtet. In diesem Handlungsfeld wurden zwischen 2020 und 2024 insgesamt 34 Projekte mit einer Förderlaufzeit von fünf (n=29) bzw. drei Jahren (n=5) mit dem Auftrag gefördert, (übertragbare) „Methoden und Konzepte der Demokratieförderung (weiter) [zu] entwickeln, [zu] erproben und [zu] überprüfen“ (Ehnert et al. 2021, S. 7; BMFSFJ 2019).

Der wB des Handlungsfeldes kam dabei die Aufgabe zu, formativ und summativ Erkenntnisse über das Handlungsfeld und seine inhaltliche wie organisatorische Entwicklung zu gewinnen, eine fachliche Bewertung der Projekt-konzeptionen und -praxis vorzunehmen sowie durch Reflexionsanstöße und wissenschaftlich fundierte Empfehlungen zur Weiterentwicklung demokratiefördernder Fachpraxis beizutragen (Ehnert et al. 2021, S. 10). Der im letzten Teilsatz angedeutete formative Evaluationsansatz hat – Flick folgend – „auch die Modifikation oder Verbesserung der laufenden Intervention“ zum Ziel (2009, S. 11). Es geht also neben erfolgreichen Projektumsetzungen immer auch um die Identifikation von Herausforderungen und dem möglichen Misslingen erprobender Ansätze, aus denen Modifikationen abgeleitet werden. Die wB kann in diesem Rahmen auf Einzelprojektebene Reflexionsräume schaffen, eine wissenschaftliche *Draufsicht* anbieten und auf Handlungsfeldebene gesammelte und systematisierte Erfahrungen für den programminternen wie -externen Transfer

(in Workshops, Publikationen und auf Fachveranstaltungen) zur Verfügung und Diskussion stellen.

Neben den inhaltlichen Aspekten und Bezugsrahmen des Handlungsfeldes Demokratieförderung (Begriffsentwicklung, Bezugskonzepte, Identifikation von Strategien und Ansätzen) spielt für die wB als Handlungsfeldevaluation der Gegenstand *Modellprojekte* eine besondere und zentrale Rolle. (Pädagogische) Modellprojekte zeichnen sich vor allem durch ihren Innovationsauftrag¹ sowie dadurch aus, dass sie durch ein erprobendes Vorgehen auf spezifische fachliche Herausforderungen reagieren sollen. Ihr Ziel ist es, aus der eigenen Praxis modellhafte, das heißt übertragbare Ansätze und Konzepte, zu entwickeln. Dass die Möglichkeit des Scheiterns innerhalb dieser Erprobung in diesem Auftrag ebenfalls eingeschrieben ist, ist in der Einleitung bereits formuliert.

Ob und in welchem Maße Modellprojekte erfolgreich bzw. wirksam sind und wie eine solche Einordnung und Bewertung aus evaluatorischer Sicht zustande kommt, hängt dabei nicht (nur) davon ab, inwiefern Projekte antragstreu oder im Rahmen eines fest abgesteckten Korridors beabsichtigter Wirkungen und Wirkungspfade handeln. Vielmehr ist eine Bewertung eines Modellprojekts oder einer Projektmaßnahme durch die wB vom gewählten Zugang und der Gegenstandskonstruktion selbst abhängig (Bischoff et al. 2019). Die Charakterisierung und Bewertung von Wirkung und Wirksamkeit ist dementsprechend immer auch abhängig vom Blick auf den Gegenstand sowie auf die Ebene, wo die Wirkung eines Programmbereichs verortet wird.

6.2.2 Unser Wirkungsverständnis und Vorgehen in den Wirkungsstudien

Eine der Leitfragen, welche während des fünfjährigen Begleitzeitraumes durch die wB bearbeitet wurde, betraf die (vornehmlich auf Zielgruppenebene beobachtbaren) Wirkungen der geförderten Modellprojekte in ihren jeweilig spezifischen Umsetzungskontexten. In der Auseinandersetzung mit Wirkungen haben wir diese grundlegend als Veränderungen infolge einer Leistung (Brand et al. 2021, S. 175) gefasst und die Plausibilisierung von Wirkungszusammenhängen in den Blick genommen, welche als „Annäherungen an kausale Aussagen“ (Lü-

1 Auf eine detaillierte und begriffliche Auseinandersetzung mit (sozialer) Innovation wird an dieser Stelle aus Platzgründen verzichtet. Grundlegend liegt unseren Überlegungen ein prozess- und handlungsbezogenes Innovationsverständnis zugrunde (Schulz-Schaeffer 2021, S. 30). Innovation kann nur aus Handeln entstehen, welches wiederum immer sozial eingebettet ist (Kessler/König 2024, S. 10). Es geht also weniger darum, ein Konzept oder eine Idee als genuin innovativ und neu einzuordnen, sondern den Prozess ihrer Anwendung zu betrachten und zu bewerten. Innovationshandeln ist somit immer Gegenstand von kommunikativer Konstruktion und – weil es im Feld multiplen sozialen Aushandlungen unterworfen ist – letztendlich von Koproduktion (Schulz-Schaeffer 2021, S. 40).

ders/Haubrich 2006) zu verstehen sind. Wir nehmen dabei bewusst Abstand vom Begriff von Wirkung „als ein kausal auf eine Maßnahme rückführbaren Effekt“ (Widmer 2012, S. 42), welcher durch „kausale Attribution“ und den systematischen Ausschluss „rivalisierender Erklärungen“ (ebd.) eindeutig identifiziert und nachgewiesen werden kann. Ein solcher Wirkungsbegriff wird unseres Erachtens den „komplexen (pädagogischen) Settings mit sich ständig verändernden Kontexten“ (Braun et al. 2023, S. 31) nicht gerecht. Deshalb haben wir uns auf die Beschreibung von empirisch rekonstruierten und multiperspektivisch plausibilisierten Wirkungszusammenhängen konzentriert.²

In Bezug auf Wirkungsfragen standen neben Fragen nach dem *Ob* von Wirkung, zentral auch jene nach dem *Wie* und der Rekonstruktion von Wirkungszusammenhängen im Vordergrund. Wesentliche Orientierungspunkte hierfür bieten die Grundannahmen der Realist Evaluation (Pawson/Tilley 1997) sowie der qualitativen Evaluationsforschung (Flick 2009).³ Leitendes Erkenntnisinteresse unserer Wirkungsstudien war es demnach, danach zu fragen, „ob und warum, in welchen Kontexten und in Bezug auf welche Zielgruppen bestimmte Strategien der Demokratieförderung wirksam sind“ (Braun et al. 2023, S. 31). Die daraus abgeleiteten Leitfragen lauteten:

„Welche (un-)intendierten Veränderungen erzeugen die MP [Modellprojekte] in ihrer Arbeit? Welche förderlichen und hinderlichen Bedingungen gibt es für die vom MP intendierten Veränderungen? In welchem Zusammenhang stehen Projektmaßnahmen, Kontextfaktoren und Outcomes?“ (ebd.)

Um den für die Realist Evaluation im Fokus stehenden „kontextspezifisch wirkenden Mechanismen“ (Dollinger 2019, S. 250) und damit der Erhellung der „Black Box“ der Wirkweisen sozial-(politisch-)er und pädagogischer Interventionen (Hense/Taut 2021, S. 268; Dössegger et al. 2017, S. 101) auf die Spur zu kommen, verfolgten wir einen „konsequent multiperspektivischen Evaluationsansatz“ (Braun et al. 2023, S. 32), in dem die subjektiven Sichtweisen aller an einer Maßnahme beteiligten bzw. betroffenen Stakeholder (Modellprojektmitarbeitende, Durchführende, Adressat:innen, (semi-)externe Kontextakteur:innen) erhoben und ausgewertet wurden. Wirkungen sind in unserer Analyse konzeptionell folglich als „Ko-Produkte der Rekonstruktion multiperspektivischer, das heißt subjektiver und interindividueller Deutungsmuster und Handlungsstrategien

2 Eine weiterführende Auseinandersetzung mit den Konzepten von Plausibilität und Kausalität findet sich beispielsweise bei Kleinoscheg (2020).

3 Bis auf die Pilotierung einer quantitativen Vorher-Nachher-Befragung bei einem einzelnen Modellprojekt haben wir uns in den letzten beiden Jahren der Wirkungsfrage empirisch ausschließlich durch qualitative Ansätze angenähert (Braun et al. 2024).

(möglichst) aller Beteiligten an einer untersuchten Maßnahme/Intervention und deren analytische Einordnung“ (ebd.; Klawe 2006, S. 126).

Aufgrund der Vielfältigkeit von Ansätzen und Strategien von Modellprojekt(aktivität)en im Handlungsfeld (Ehnert et al. 2021, S. 23 ff.) haben wir uns für ein clusteranalytisches Vorgehen in Anlehnung an die programmtheoriebasierte Clusterevaluation nach Worthen und Schmitz (1997) bzw. Haubrich (2006, 2009) entschieden, um durch kontrastierendes Sampling und die (paar-)vergleichende Auswertung von Einzelprojektwirkungen zu projektübergreifenden und vorsichtig verallgemeinerbaren Aussagen zu wirksamen Maßnahmen der Demokratieförderung auf Handlungsfeldebene zu gelangen (Braun et al. 2023, S. 31). In den konkreten empirischen Blick kamen dadurch punktuell einzelne Wirkungsausschnitte ausgewählter Modellprojekte, welche in Projektpaaren zusammengefasst wurden (insgesamt zehn Modellprojekte in fünf thematischen Clustern). Die Cluster setzten sich aus Projekten zusammen, welche in vergleichbaren Feldern/Umsetzungskontexten (z. B. Schule, Kommune, Zivilgesellschaft) ähnliche Zielsetzungen bezogen auf ähnliche Zielgruppen (z. B. pädagogische Fachkräfte, Schüler:innen, politisch Engagierte) verfolgen (Haubrich 2006, S. 110). In den Fokus rückten – auch geschuldet dem begrenzten Bearbeitungszeitraum – jeweils spezifische Einzelprojektmaßnahmen als Wirkungsausschnitte und nicht die Modellprojektarbeit in Gänze.

Erhoben wurden die unterschiedlichen Perspektiven durch qualitative leitfadengestützte Interviews in Gruppen- und Einzelsettings sowie durch ethnografische Beobachtungen und deren Protokollierung. Bei Möglichkeit wurden außerdem projekteigene Dokumente und Produkte der Modellprojekte (Konzeptpapiere, Veranstaltungsdokumentationen etc.) in die Datensammlung eingebunden. Inhaltlich fokussierten die Erhebungen unter anderem die Durchführung und Ausgestaltung der Projektangebote, eingetretene Veränderungen bei Zielgruppen und in Zielinstitutionen, Gründe für das (Nicht-)Zustandekommen intendierter Veränderungen sowie institutionelle und sozialräumliche Kontextfaktoren. Die Auswertung und Ergebnisdarstellung erfolgte – vornehmlich inhaltsanalytisch, mit punktuellen Vertiefungen mithilfe der Dokumentarischen Methode (Nohl 2017) bei der Auswertung der Adressat:innenperspektive – orientiert an den für die Realist Evaluation wesentlichen „theoretischen Komponenten realistischer Erklärungen von Effekten: Kontext, Mechanismus und Outcome“ (KMO) (Braun et al. 2023, S. 33; Haunberger/Baumgartner 2017, S. 125; Pawson/Tilley 1997, S. 412). Um den Wirkungsausschnitten einzelner Projektaktivitäten und deren Spezifika theoretisch begründet sowie darstellungstechnisch gerecht zu werden, erweiterten wir das verwendete Kategorienschema (KMO) deduktiv um die Kategorie „Aktivität“ (A). So gelangten wir schließlich zu einer Auswertungssystematik der in den Blick genommenen Projektaktivitäten entlang eines Aktivität-Kontext-Mechanismus-Outcome-Schemas (kurz: AKMO-Schema bzw.

-Konfiguration).⁴ Die einzelnen Kategorien haben wir dabei wie folgt bestimmt (Braun et al. 2023, S. 33):

- **Aktivität:** die im Rahmen der zu betrachtenden Wirkausschnitte untersuchte(n) Maßnahme(n) des Modellprojekts und deren Ausgestaltung.
- **Kontext:** die Rahmenbedingungen mit Einfluss auf die Wirkweise der Intervention.
- **Mechanismus:** die vermittelnden (sozialen) Wirkweisen der Aktivität-Outcome-Ketten, „durch die Maßnahmen wirken (oder nicht)“ (Dollinger 2019, S. 251).
- **Outcome:** die plausibel auf die Maßnahme (Aktivität) zurückführbaren (un-)intendierten Effekte.

In den Ergebnispräsentationen der Wirkungsstudien (Braun et al. 2023; Braun et al. 2024) konnten wir schließlich einzelne Wirkungszusammenhänge empirisch plausibilisiert rekonstruieren, aus ihnen vorsichtige Schlüsse ziehen und Empfehlungen für die Praxis im Handlungsfeld abgeben (vgl. Braun/Hemmann/Rehse in diesem Band). Der Fokus der Auseinandersetzung mit Wirkungen lag dabei klar auf der Ebene der Adressat:innen. Erfahrungen des Lernens und Scheitern in der Modellprojektumsetzung traten zugunsten dieser Fokussierung in den Hintergrund.

6.3 Möglichkeiten der Auseinandersetzung mit Lernen und Scheitern in Wirkungsstudien

Im Folgenden gehen wir zunächst kurz darauf ein, welches Verständnis von Lernen und Scheitern und ihrem Zusammenhang in der Modellprojektarbeit unseren Überlegungen zugrunde liegt. Ausgehend von diesem Verständnis werfen wir einen Blick auf die Empirie und stellen an zwei Beispielen vor, inwiefern uns Lernen und Scheitern in den Erhebungen unserer Wirkungsstudien begegnet sind. Daran anschließend identifizieren wir Anschlusspunkte für die stärkere Berücksichtigung von Lernen und Scheitern im Rahmen unseres bestehenden Vorgehens in den Wirkungsstudien und gehen abschließend auf einige methodische Erfordernisse und die Grenzen der Auseinandersetzung mit Lernen und Scheitern ein.

4 In der internationalen Rezeption und praktischen Umsetzung realistischer Evaluationsansätze gibt es zahlreiche unterschiedliche Anwendungen des ursprünglich von Pawson und Tilley vorgeschlagenen CMO (Context-Mechanism-Outcome)-Schemas. So existieren auch abseits unserer Modifikationen Umsetzungen, welche die Kategorien *Intervention*, *Strategies* oder *Actors* gegenstandsbezogen in ihre Heuristiken einbeziehen. Wichtig sind dabei die Transparenz und Definition der genutzten Kategorien. Eine erste systematische Auseinandersetzung hierzu ist bei De Weger et al. (2020) zu finden.

6.3.1 Lernen und Scheitern in der Modellprojektarbeit

Scheitern kann, wie einleitend bereits angemerkt, als ein Charakteristikum von Modellprojekten und der Umsetzung ihrer Aktivitäten in einem Experimentier- raum verstanden werden. Die Frage, wie Scheitern und Lernen sowie der Zusammen- hang zwischen ihnen gefasst werden kann, stand bisher in unserer Ausein- andersetzung mit Modellprojekten jedoch weniger stark empirisch im Fokus. Dies betrifft insbesondere die Auseinandersetzung mit den Wirkungen der Modell- projektarbeit. Für unsere folgenden Ausführungen stellen jene Überlegungen einen Ausgangspunkt dar, mit denen Scheitern in der Modellprojektpraxis „als eine wichtige Ressource für erfahrungsbasiertes Lernen [angesehen wird], da es In- novationen, z. B. im Rahmen neuer Zielgruppenzugänge anregt und erzwingt“ (Figlesthler et al. 2019, S. 19; ähnlich Brand et al. 2021, S. 20). Im Anschluss an diese Überlegungen erscheint ein Verständnis von „Lernen als Erfahrung in Ver- hältnissen“ (Brinkmann 2022, S. 280) für die Auseinandersetzung mit Lernen in der Modellprojektarbeit einen Erkenntnisgewinn zu bringen. Lernen kann damit als „Erfahrungsprozess“ (ebd., S. 281) markiert werden, in welchem den vorhan- denen Vorerfahrungen auf der Seite der Lernenden Bedeutung beigemessen, Ler- nen aber nicht als Akkumulation von Wissen und Können in Form des Hinzufü- gens zu bereits Vorhandenem verstanden wird (ebd.). Vielmehr rückt in den Blick, dass „[n]eue Erfahrungen [...] sich im Lernprozess gegenüber alten profili- eren [können], die Vertrautheit mit ihnen kann unterbrochen, die routinisierte Praxis sowie Gewohnheiten und Habitus irritiert und möglicherweise verän- dert werden“ (ebd., S. 281). *Lernen* meint somit nicht nur Dazulernen, sondern auch Umlernen (ebd., dazu ausführlicher Mayer-Drawe 2008). Ausgehend von diesem Verständnis von Lernen können negative Erfahrungen (z. B. Erfahrungen des Scheiterns) zur Änderung vorhandener Erfahrungsbestände führen. Negative Erfahrungen respektive Scheitern als „Erfahrungsöffnung für Anderes und Fremdes“ (ebd., S. 283) sind als Folge „bedeutsame Anlässe für Lernen und Umlernen“ (ebd.). Grundlage für „eine lernende Auseinandersetzung mit den Themen des Scheiterns“ stellen „Reaktionsmodi der Reflexion“ dar (Hilliger 2024, S. 93).

Bezogen auf die Modellprojektarbeit beschreibt Lernen das Machen von Erfahrungen, die sich beispielsweise in der Veränderung bestehender Wissens- bestände aufseiten der Modellprojektmitarbeitenden niederschlagen können oder dazu führen, dass diese ihr routiniertes, auf Wissen basierendes Handeln im Modus der Reflexion hinterfragen und gegebenenfalls revidieren. Der Bezug auf Verhältnisse und das Verständnis von Lernen *in Verhältnissen* macht deutlich, dass Lernen im Kontext der Modellprojektkonzeption und -umsetzung erfolgt, die wiederum nicht im *luftleeren Raum* stattfinden, sondern von Kontextbedin- gungen beeinflusst werden. *Scheitern* als negative Erfahrungen kann in diesem Zusammenhang als produktives Moment der Modellprojektarbeit verstanden werden. Scheitern stellt somit eine Ressource für Lernen dar bzw. kann dieses

ermöglichen. Für die pädagogische Praxis der Modellprojektarbeit gilt dementsprechend, dass ein temporärer „Misserfolg, also ein graduelles Scheitern [...] Perspektiven für einen Wandel und erneut die Aussicht auf Erfolg“ eröffnen kann (Kessler/König 2024, S. 10; Hilliger 2024, S. 99).

6.3.2 Empirische Beispiele für Lernen und Scheitern

Der eingangs zitierte Interviewauszug „Das hätte ich, dann jetzt rückblickend betrachtet, noch anders gestaltet“, der aus dem Interview mit einer Modellprojektmitarbeiterin stammt und sich auf den Beziehungsaufbau mit Adressat:innen bezieht, verdeutlicht beispielhaft, welche Einblicke wir in unseren Wirkungsstudien in Erfahrungen des Lernens und Scheiterns in der Modellprojektarbeit bekommen haben. Ausgehend von den skizzierten Konzeptionen von (erfahrungsbasiertem) Lernen und Scheitern zeigt sich an diesem Interviewauszug und seiner Kontextualisierung, wie aufseiten von Modellprojekten gescheitert und gelernt wird. Die Aussage der Projektmitarbeiterin stammt aus dem Bericht zum Beziehungsaufbau mit Adressat:innen des Modellprojekts. Die Mitarbeiterin berichtet zuvor im Interview, dass viele Ressourcen in den Beziehungsaufbau mit vergleichsweise „kooperationsunwilligen“ Adressat:innen des Modellprojektangebots investiert wurden, dies aber nicht erfolgreich war – also gescheitert ist. Als Folge kam es zu zeitlichen Verschiebungen von Maßnahmen, die von der Modellprojektmitarbeiterin als temporäres Scheitern angesprochen werden.⁵ Sie kommt zum Schluss, dass sie den Beziehungsaufbau, „rückblickend betrachtet, noch anders gestaltet [hätte]“, und dies, wie sie im weiteren Interviewverlauf konkretisiert, zukünftig durch die stärkere Konzentration auf kooperationsbereite Adressat:innen tun wird. Die Erfahrung des Scheiterns – trotz Bemühungen kam die Kooperation nicht zustande und es kam zu zeitlichen Verschiebungen – führt zu einem Lerneffekt und zukünftig, so die Einschätzung zum Interviewzeitpunkt, zu einer veränderten Handlungspraxis.

Ähnliche Beispiele für Lernen und Scheitern finden sich auch in anderen Interviews unserer Wirkungsstudien. Begründet in unserer Auseinandersetzung mit Wirkungen, die ein multiperspektivisches Vorgehen und den Einbezug verschiedener Interviewpartner:innen erfordert, werden Lernen und Scheitern nicht nur von Modellprojektmitarbeitenden *zur Sprache gebracht*, wie ein zweites Beispiel verdeutlicht. Im Rahmen der Erhebungen, in denen ein Bildungsangebot für Jugendliche und junge Erwachsene eines Modellprojekts im Fokus stand, zeigten sich Erfahrungen des Lernens und Scheiterns zunächst im Interview

5 Dieses temporäre Scheitern wurde nicht nur von den Mitarbeitenden, sondern auch von weiteren Interviewpartner:innen aus dem Projektkontext thematisiert und mit einer zurückhaltenen Wirksamkeitseinschätzung in Verbindung gesetzt.

mit einer teilnehmenden Person. Diese berichtete von einer Änderung des thematischen Zuschnitts des Angebots in Form einer inhaltlichen Zuspitzung auf bestimmte Themen, die mit Veränderungen der adressierten Zielgruppe und der Zusammensetzungen von Teilnehmendengruppen einherging. Aus Sicht der teilnehmenden Person ermöglichte dies eine bessere Betreuung der Teilnehmenden durch das Modellprojekt. Im Interview wurde diese Änderung positiv bewertet und als wichtige Reflexionsleistung des Modellprojekts gerahmt. In unserer Weiterarbeit und folgenden Kontakten mit den Modellprojektmitarbeitenden zeigte sich auch seitens des Modellprojekts eine ähnliche Einschätzung. Es wurde deutlich: Die Modellprojektmitarbeitenden sind mit ihren Überlegungen zu den zu behandelnden Themenschwerpunkten und der mit ihnen angesprochenen Zielgruppe gescheitert, insofern die Themen und die Zielgruppenadressierung zu weit gefasst und mit den Ressourcen des Modellprojekts nicht ausreichend abzudecken waren. Dieses Scheitern hat zu einer Veränderung beider Aspekte (Themen und Zielgruppe) geführt. Es kann festgehalten werden, dass die Modellprojektmitarbeitenden aus den gemachten Erfahrungen gelernt haben und die Modellprojektaktivitäten im Verlauf der Durchführung weiterentwickelt wurden.

An den skizzierten Beispielen wird sichtbar, wie heterogen Erfahrungen des Lernens und Scheiterns in der Modellprojektarbeit sein können. Entsprechende Erfahrungen zeigen sich anhand der Zielgruppenauswahl, von zeitlichen Verschiebungen und inhaltlichen Zuspitzungen. Die Beispiele verdeutlichen auch, dass Lernen und Scheitern im Rahmen der Interviews unserer Wirkungsstudien nicht nur von Modellprojektmitarbeitenden thematisiert, sondern auch von anderen interviewten Personen wie den Teilnehmenden *ins Spiel gebracht werden*. Die Multiperspektivität unserer Auseinandersetzung mit Wirkungen ermöglicht relevante Einblicke in Lernen und Scheitern aus verschiedenen Perspektiven.

Gerahmt werden diese Beobachtungen jedoch von der Tatsache, dass unsere gegenwärtigen Einblicke in Lernen und Scheitern in der Modellprojektarbeit unterschiedlich weit in die Tiefe reichen. Dies hängt mit dem Fokus unserer Wirkungsstudien, ihrem Erkenntnisinteresse und den Fragestellungen zusammen.

6.3.3 Anschlusspunkte für die stärkere Berücksichtigung von Lernen und Scheitern

Ausgehend vom beschriebenen Vorgehen der Wirkungsstudien (Abschnitt 6.2.2) können zwei miteinander verbundene Anschlusspunkte skizziert werden, um Lernen und Scheitern systematischer in den Blick zu nehmen. Bei ihnen handelt es sich zum einen um das AKMO-Schema und zum anderen um die Konzeption von Lern- und Scheiternserfahrungen als Koproduktion.

6.3.3.1 Berücksichtigung von Lern- und Scheiternserfahrungen im AKMO-Schema

Einen ersten Anschlusspunkt für die systematischere Untersuchung von Lernen und Scheitern mit unserem Wirkungsstudiendesign stellen das AKMO-Schema und die ihm zugrundeliegenden Konzeptionen von Kontext, Aktivität und Wirkung dar. Mit dem Begriff des *Kontextes* beschreiben wir im AKMO-Schema die „Rahmenbedingungen mit Einfluss auf die Wirkweise der Intervention“ (Braun et al. 2023, S. 33). Der Kontext einer Intervention lässt sich dabei nicht als soziales Setting *an sich* verstehen (also beispielsweise als *die* Schule oder *die* Kommune). Dementsprechend liegt unserer Vorgehensweise nicht die Annahme zugrunde, dass wir auf einer solchen verallgemeinerbaren Ebene über Kontextbedingungen, die die Wirksamkeit einer Maßnahme ermöglichen oder verhindern, sprechen können. Unser Vorgehen zielt im Gegenzug darauf, jene Kontextbedingungen empirisch in den Blick zu nehmen, die sich in ihrer jeweiligen Spezifik ermöglichend oder verhindernd auswirken. Bedeutsam werden damit die in einem Kontext vorhandenen sozialen Beziehungen, Regeln, Normen, Erwartungen und Ressourcen sowie ihr spezifisches Zusammenspiel.⁶ Sie sind in der jeweiligen Konstellation untrennbar verflochten mit den Mechanismen, durch die eine Maßnahme wirkt. Dabei ist davon auszugehen, dass *der* Kontext im Rahmen seiner Interaktion mit Mechanismen potenziell zur Konstitution neuer Kontexte beiträgt (ausführlicher The RAMESES II Project 2017; auch Nielsen/Lemire/Tangsig 2022, S. 95). Angesichts dieser Konzeption von Kontext zeigt sich Lernen und Scheitern im Rahmen der von uns durchgeführten Wirkungsstudien dort, wo Kontextbedingungen derart hemmend vorliegen, dass sie das Eintreten von intendierten Veränderungen durch Mechanismen als deren generative Prozesse stören.⁷ Lern- und Scheiternserfahrungen in der Umsetzung einer Maßnahme als Bestandteil der Modellprojektarbeit verdeutlichen somit einerseits den großen Einfluss von Kontextbedingungen auf die Umsetzung. Sie können andererseits aber auch Leerstellen in den Konzepten der Modellprojekte hinsichtlich des Einbezugs und der Berücksichtigung von relevanten Rahmenbedingungen sichtbar machen, deren Reflexion zu einer modifizierten und verbesserten Ausgestaltung der Angebote führen kann.

6 Diese Fokussierung auf Kontextbedingungen in ihrer jeweiligen Spezifik bedeutet nicht, dass sich keine Abstrahierungen vornehmen ließen. So können mit der umgesetzten Vorgehensweise sehr wohl clusterübergreifende Kontextbedingungen identifiziert und anhand der vorhandenen Daten untermauert werden. Diese Abstrahierungen sind dann jedoch empirisch begründet und auf das jeweils erhobene Material zurückführbar. Kontextbedingungen werden dementsprechend nicht vorab gesetzt, wenn eine Maßnahme in einem bestimmten Setting stattfindet.

7 Laux formuliert bezogen auf Innovation eine ähnliche Einschätzung: „Neue Ideen scheitern häufig nicht an fehlendem Potenzial, sondern an der konkreten Umsetzung durch eine mangelhafte Synchronisation mit Ereignissen in der gesellschaftlichen Umwelt“ (2021, S. 172).

Das AKMO-Schema bietet mit der Kategorie der *Aktivität* einen zweiten Anschlusspunkt für das systematischere In-den-Blick-Nehmen von Lernen und Scheitern. Als Aktivitäten fassen wir „die im Rahmen der zu untersuchenden Wirkausschnitte untersuchte[n] Maßnahme[n]“ der Modellprojekte (Braun et al. 2023, S. 33). Qua unseres Verständnisses vom Gegenstand *Modellprojekt* (Abschnitt 6.2.1) ist zu berücksichtigen, dass es selbst im Durchführungszeitraum unserer Wirkungsstudien zu Modifikationen der Maßnahmen kommen kann. Denn Modellprojektmitarbeitende sind zur Reflexion ihrer Umsetzung und zur Modifikation aufgerufen, wenn sich ein entsprechender Bedarf zeigt.

Ein dritter Anschlusspunkt im AKMO-Schema für die stärkere Berücksichtigung von Lernen und Scheitern lässt sich anhand unseres Verständnisses von *Outcomes* identifizieren. Mit der Kategorie der Outcomes beschreiben wir (un-)intendierte Effekte, die plausibel auf die jeweils untersuchte Maßnahme (Aktivität) der Modellprojekte zurückführbar sind (Braun et al. 2023, S. 33). Kennzeichnend für unser Vorgehen war, dass wir die hervorgerufenen Effekte in Form von Wirkungen auf der Ebene von Zielgruppen in den Blick genommen haben. Dabei konnten wir je nach Maßnahme unterschiedliche Wirkungen und Wirkzusammenhänge plausibilisieren und zwischen intendierten und unintendierten Wirkungen unterscheiden. Im Rückblick auf die Wirkungsstudien und ihre Ergebnisse zeigt sich beispielhaft: Das Eintreten angestrebter Lerneffekte bei Teilnehmenden eines Workshops als untersuchter Maßnahme eines Modellprojekts kann mithilfe des AKMO-Schemas und des zugrundeliegenden Verständnisses von Outcomes als intendierte Wirkung beschrieben werden. Übertragen auf die Lerneffekte, die durch Erfahrungen des Scheiterns bei Modellprojektmitarbeitenden entstehen, kann geschlussfolgert werden, dass es sich auch hier um Wirkungen handelt. Es handelt sich in unserer Unterscheidung zwischen intendierten und unintendierten Wirkungen jedoch nicht um intendierte Wirkungen, denn sie waren in dieser Form nicht angestrebt und treten zudem bei den Modellprojektmitarbeitenden und nicht bei den Teilnehmenden einer Maßnahme auf. Lernen der Modellprojektmitarbeitenden in Folge von Scheitern kann im Rahmen unseres Vorgehens mit dem Fokus auf Wirkungen auf Zielgruppenebene als unintendierte Wirkung bei einer nicht-anvisierten Zielgruppe der Modellprojektarbeit verstanden werden.⁸ Mit dem Blick auf diese unintendierten Wirkungen wird aber auch ein Grundcharakteristikum der Modellprojektarbeit – nämlich das des potenziellen Scheiterns – sichtbar und in Wirkungsstudien einbezogen.

8 Zumindest partiell finden sich ähnliche Überlegungen in den von uns erstellten Wirkmodellen zu den Wirkungsstudien wieder. Hier sind als unintendierte Wirkungen der Modellprojektaktivitäten beispielsweise Lerneffekte von Kontextakteur:innen oder Empowerment-Erfahrungen von Modellprojektmitarbeitenden angeführt. Sie werden jedoch nicht tiefergehend einbezogen (Braun et al. 2024).

6.3.3.2 Verständnis von Lern- und Scheiternserfahrungen als Koproduktion

Eine zweite Anschlussstelle für den systematischeren Einbezug von Lernen und Scheitern zeigt sich anhand unseres Verständnisses von Wirkungen als Ergebnisse multiperspektivischer Koproduktion (Braun et al. 2023, S. 32). Lerneffekte als unintendierte, bei Modellprojektmitarbeitenden hervorgerufene Wirkungen und ihnen zugrundeliegende Erfahrungen des Scheiterns stellen im Anschluss an dieses Verständnis ebenfalls Koprodukte dar. Die Auseinandersetzung mit ihnen erfordert dementsprechend einen multiperspektivischen Evaluationsansatz und den Einbezug der Sichtweisen aller an einer Maßnahme beteiligten Stakeholder – und nicht nur der Sichtweisen der Modellprojektmitarbeitenden, deren Lernen und Scheitern im Fokus steht. Die Relevanz, die Sichtweisen der unterschiedlichen Stakeholder auf Lernen und Scheitern einzubeziehen, zeigt sich anhand des zweiten ausführlicher vorgestellten empirischen Beispiels. Hier ist es das Interview mit einer teilnehmenden Person an einem Bildungsangebot eines Modellprojekts, das Aufschlüsse über Lernen und Scheitern gibt und eine weitere Auseinandersetzung ermöglicht. Hinzu kommt die Frage, wann bzw. ab wann etwas (beispielsweise eine Maßnahme in den Wirkungsstudien oder einer ihrer Bestandteile) als gescheitert gilt und ab wann von Lerneffekten gesprochen werden kann. Unterschiedliche Stakeholder, genauso wie die wB, werden verschiedene Einblicke und Verständnisse von Lernen und Scheitern aufweisen, auf deren Basis die Bedeutung von Lern- und Scheiternserfahrungen für die Modellprojektarbeit rekonstruiert werden kann.

6.3.4 Methodische Erfordernisse und Grenzen der Auseinandersetzung mit Lernen und Scheitern

Wie unsere empirischen Beispiele zeigen, sind Modellprojektmitarbeitende und auch weitere Stakeholder bereit, in den Interviews unserer Wirkungsstudien Auskunft über Lernen und Scheitern zu geben. Ob dies für alle Modellprojekte gilt und dementsprechend verallgemeinert werden kann, ist jedoch zu hinterfragen, geht es doch im Kontext des Erhalts finanzieller Förderung um die „Offenheit des Felds [der Modellprojekte] und die Bereitschaft seiner Akteure, ‚sich in die Karten gucken zu lassen“ (Klawe 2006, S. 139). Dieser Aspekt wird virulenter angesichts unserer Position als wB, die auch die Bewertung der Modellprojektarbeit umfasst (Ehnert et al. 2021, S. 10), und erfordert die Reflexion unserer eigenen Eingebundenheit als wB in das Evaluationsgeschehen (Flick 2009, S. 13). In der Kombination verdeutlicht dies, dass das Sprechen insbesondere über Scheitern im Rahmen der Evaluation voraussetzungsvoll ist, was seitens der wB anerkannt und berücksichtigt werden muss.

Daran anschließend eröffnen sich weiterführende Erfordernisse hinsichtlich des Kontakts, des Vertrauensaufbaus mit Modellprojektmitarbeitenden und weiteren Stakeholdern und der Datenerhebungen: Notwendig erscheint nicht nur die Reflexion von unumgänglichen Hierarchien in der angestrebten kritisch-partnerschaftlichen Arbeitsbeziehung zwischen Modellprojekten und wB, sondern auch der Blick auf Möglichkeiten ihres (zumindest partiellen) Abbaus im Rahmen des Aufbaus einer vertrauensvollen Beziehung. In diesem Zusammenhang ist auch die weiterführende Auseinandersetzung mit Fragen der Transparenz von Bewertungskriterien nötig (ausführlicher Schau/Greuel 2021). Dies betrifft auch Transparenz im Hinblick auf die Frage, wann und auf welcher empirischen Grundlage etwas als Scheitern interpretiert wird und ob diesem Scheitern produktive Potenziale oder eine negative Bedeutung beigemessen werden (Kessler / König 2024, S. 11 f.).

Die Auseinandersetzung mit Lernen und insbesondere mit Scheitern erfordert darüber hinaus eine Begründung gegenüber den Modellprojekten und den weiteren einbezogenen Stakeholdern, warum diese Themenkomplexe für die wB von Bedeutung sind. Zudem stellt sich die Frage, wie Lernen und Scheitern beispielsweise in Interviewsituationen kontextsensibel *zur Sprache gebracht* werden können, was wiederum die Entwicklung von Erhebungsinstrumenten in den Fokus rückt. Zugleich legt der Hinweis auf Bewertungen durch die wB und deren Relevanz für die Modellprojektarbeit die Vermutung nahe, dass insbesondere Scheitern in der wissenschaftlichen Begleitung von Modellprojekten trotz jeglicher Fokussierungsversuche immer nur ausschnitthaft in den Blick rückt und es – aus Modellprojektperspektive durchaus nachvollziehbar – Akteur:innen gibt, die sich nicht *in die Karten gucken lassen*.

6.4 Potenziale des Einbezugs von Lernen und Scheitern – weiterführende Perspektiven für die wB

In den vorangegangenen Ausführungen haben wir zum einen theoretisch-konzeptionell aufgezeigt, inwiefern Modellvorhaben eng mit Lern- und Scheiternerfahrungen verknüpft sind und dass zur Erprobung wirksamer Konzepte das potenzielle und tatsächliche (graduelle) Scheitern von Maßnahmen zur Fachpraxis dazugehört. Zum anderen haben wir empirisch dargestellt, dass Lern- und Scheiternerfahrungen in unseren Wirkungsstudien zwar *zur Sprache kamen*, aber nicht systematisch und tiefergehend ergründet wurden. Dem folgten Überlegungen zu Anschlusspunkten für die stärkere Berücksichtigung von Lernen und Scheitern in unserem, an der Realist Evaluation orientierten Vorgehen in der Wirkungsuntersuchung von Modellprojektarbeit und der Blick auf daraus resultierende me-

thodische Erfordernisse und Grenzen der Auseinandersetzung mit Lernen und Scheitern.

In diesem Teil möchten wir Potenziale eines systematischeren Einbezugs von Lern- und Scheiternserfahrungen für die Aufgabe der wB als Handlungsfeldevaluation reflektieren. Damit bewegen wir uns weg von dem bisher vorgenommenen engen Bezug auf unser Vorgehen in den Wirkungsstudien und blicken stärker auf Transfermöglichkeiten und die Handlungsfeldebene. Für eine wirkungsorientierte Evaluation, welche sowohl summativ als auch formativ ausgerichtet ist, ist es unseres Erachtens nach nicht nur von Bedeutung, Wirkungen von Modellprojektmaßnahmen auf Adressat:innenebene und deren Wirkungsweisen nachzuvollziehen, sondern auch Veränderungs- und Anpassungsschleifen in der Projektumsetzung in den Blick zu bekommen.⁹ Schließlich ist es auch unsere Aufgabe als wB, an modellprojektübergreifende Erkenntnisse zur (Weiter-)Entwicklung guter (also wirksamer) Fachpraxis im Bereich der Demokratieförderung zu gelangen und diese systematisch für Transfer – also die interne und externe Weitergabe von Erkenntnissen, Erfahrungen und Konzepten (Brand et al. 2021, S. 74 ff.) – aufzubereiten. Der Transfer von Lern- und Scheiternserfahrungen und damit einhergehenden Weiterentwicklungen von Maßnahmen und Konzepten kann in diesem Sinne als Wirkung auf Handlungsfeldebene und Ergebnis von Innovationsproduktion verstanden werden. Gerade die gemachten Erfahrungen und das gesammelte Wissen der an den Modellprojekten beteiligten Personen – zentral: Modellprojektmitarbeitende, aber auch Adressat:innen und Kontextakteur:innen – stellen wertvolle Ressourcen für den Transfer in die Fachpraxis dar. Um diese Ressourcen zu nutzen, muss die wB für die erprobende Praxis von Modellprojekten entsprechende Reflexionsräume und -gelegenheiten schaffen. Dadurch wird ihr einerseits die Rolle als Ermöglicher von Reflexionsprozessen auf Einzelprojektebene zuteil, welche sie durch Workshopangebote für alle Projekte und die dortige Bereitstellung systematischer Befunde zu den Modellprojekten auf die Handlungsfeldebene erweitern kann. Andererseits kann die wB als zusammentragende Instanz selbst als Transferakteur und somit als Koproduzent von Wirkung auf Handlungsfeldebene bewusst auf den Plan treten. Denn im Sinne der Realist Evaluation bleibt die Herausforderung der Verallgemeinerbarkeit von Aussagen über die Wirksamkeit von Maßnahmen in sozialen Settings bestehen, „[i]mmerhin aber kann durch Abstraktion von einzelnen Maßnahme-Durchführungen und durch die Berücksichtigung von Erfahrungen aus verschiedenen Evaluationen langfristig angestrebt werden, ‚sets of ideas‘ (Pawson/Tilley 1997, S. 120) zu gewinnen, die über divergente Kontexte hinweg Vergleichbarkeiten gewährleisten“ (Dollinger

9 Und damit dem Umstand Rechnung zu tragen, dass die (pädagogische) Umsetzung von Modellprojekten nicht als „lineare Vorwärtsbewegung“ (Sattler 2016, S. 171 in Kessler/König 2024, S. 14) misszuverstehen ist.

2019, S. 252). Die handlungsfeldbezogene Aufgabe der wB betrifft schließlich also nicht die Produktion grundsätzlich und für jeden Kontext verallgemeinerbarer Aussagen über Wirkungen, sondern die bestmögliche Aufbereitung und Bereitstellung von Umsetzungserfahrungen und Praxiswissen, „das als Bezugspunkt fachlicher Reflexion“ genutzt werden kann (Albus/Micheel/Polutta 2010, S. 241 in Dollinger 2018, S. 253). Der Einbezug von Lernen und Scheitern kann diesem *Bezugspunkt* ein valideres und empirisch gesicherteres Fundament verleihen.

6.5 Abschluss – Zur Relevanz der Auseinandersetzung mit Lernen und Scheitern

Ausgehend von den empirischen Beobachtungen aus unseren Datenerhebungen und dem Ziel, Lernen und Scheitern in der Modellprojektumsetzung in Wirkungsstudien systematischer einzubeziehen, verdeutlicht unsere Auseinandersetzung mehrere Punkte: Wir haben anhand unseres, an der Realist Evaluation orientierten Vorgehens verschiedene Anschlusspunkte für den Einbezug von Lern- und Scheiternserfahrungen identifiziert. Unter anderem haben wir gezeigt, wie Lern- und Scheiternserfahrungen von Modellprojektmitarbeitenden als unintendierte Wirkungen bei einer nicht-anvisierten Zielgruppe in Wirkungsstudien einbezogen werden können. Darüber hinaus haben wir deutlich gemacht, dass sich das Potenzial der Auseinandersetzung mit Lernen und Scheitern in der Modellprojektarbeit nicht ausschließlich auf die Identifikation von unintendierten Wirkungen bezieht. Unser Blick auf die Möglichkeiten des Transfers von Lern- und Scheiternserfahrungen zeigt, inwiefern Lernen und Scheitern sowie die damit einhergehende Weiterentwicklung von Maßnahmen und Konzepten auch auf der Handlungsfeldebene eine Wirkung entfalten und Veränderungen anregen kann. Die wB von Modellprojekten rückt in diesem Zusammenhang potenziell als Ermöglicher von Reflexionsprozessen und als Transferakteur in den Fokus.

Zusammenfassend betrachtet zeigt unser Beitrag verschiedene Denkrichtungen für eine Blickverschiebung in Wirkungsanalysen auf, die zukünftig weiterverfolgt werden können. Dass es sich dabei um ein komplexes Unterfangen handelt, haben wir mit der Vorstellung der methodischen Erfordernisse für die Auseinandersetzung mit Lernen und Scheitern deutlich gemacht. Auch die Grenzen dieser Auseinandersetzung sind an dieser Stelle nicht außen vor zu lassen. Dennoch verstehen wir unseren Beitrag als Plädoyer für die Relevanz der Auseinandersetzung mit Lernen und Scheitern der Modellprojektumsetzung in Wirkungsstudien. Diese Auseinandersetzung eröffnet, wie wir gezeigt haben, äußerst relevante Einblicke in den „Experimentierraum“ (Ehnert et al. 2021, S. 7), in dem Modellprojekte ihre Konzepte erproben und dabei Lern- und Scheiternserfahrungen machen.

Als wertvolle Erfahrungen für die Fachpraxis sollten diese stärker in den Blick genommen werden. In den Überlegungen unseres Beitrags klingt zudem an, dass die Modellprojektarbeit zwar prädestiniert für die Auseinandersetzung mit Lernen und Scheitern ist, entsprechende Erfahrungen aber auch kennzeichnend für die (sozial-)pädagogische Praxis an sich (abseits von Modellhaftigkeit) sind (Kessler/König 2024). Dies verweist auf die grundsätzliche Relevanz des Einbezugs von Lernen und Scheitern in Wirkungsstudien – auch wenn diese sich nicht mit Wirkungen von Modellprojekten, sondern mit anderer (sozial-)pädagogischer Praxis auseinandersetzen.

Literatur

- Albus, S./Micheel, H.-G./Polutta, A. (2010): Empirie der Wirkungsorientierung. In: Otto, H.-G./Polutta, A./Ziegler, H. (Hrsg.): *What Works*. Opladen: Barbara Budrich, S. 231–244.
- Bischoff, U./Greuel, F./Heinze, F./König, F./Zimmermann, E. (2019): Ein Modellprojekt ist ein Modellprojekt? Gegenstandskonstruktionen im Bundesmodellprogramm „Vielfalt tut gut“. In: Frank, C./Jooß-Weinbach, M./Loick Molina, S./Schoyerer, G. (Hrsg.): *Der Weg zum Gegenstand in der Kinder- und Jugendhilfeforschung. Methodologische Herausforderungen für qualitative Zugänge*. Weinheim und Basel: Beltz Juventa, S. 245–265.
- BMFSFJ (2019): *Förderaufruf für Modellprojekte im Handlungsfeld „Demokratieförderung“ im Rahmen des Bundesprogramms „Demokratie leben!“*. Berlin.
- BMFSFJ (2023): „Bundesprogramm ‚Demokratie leben!‘“. <https://www.bmfsfj.de/bmfsfj/themen/engagement-und-gesellschaft/bundesprogramm-demokratie-leben-73948> (Abruf 18.04.2024).
- Brand, A./Johann, T./Leistner, A./Rehse, A./Roscher, T./Walter, E./Zimmermann, E. (2021): *Vierter Bericht: Modellprojekte D. Wissenschaftliche Begleitung der Modellprojekte zu GMF und Demokratiestärkung. Programmevaluation „Demokratie leben!“ Zwischenbericht 2018*. Halle (Saale): DJI.
- Braun, M./Ehert, K./Euteneuer, J./Hemmann, M./Johann, T./Rehse, A./Rüger, S./Walter, E. (2023): *Demokratieförderung im Kindesalter: Kontext- und Wirkungsanalysen. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024*. Unter Mitarbeit von H. Bilgili. Halle (Saale): DJI.
- Braun, M./Ehert, K./Euteneuer, J./Hemmann, M./Johann, T./Rehse, A./Rüger, S./Walter, E. (2024): *Demokratieförderung im Jugend- und jungen Erwachsenenalter – Wirkungsanalysen. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld „Demokratieförderung“ im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024*. Unter Mitarbeit von H. Bilgili. Halle (Saale): DJI.
- Brinkmann, M. (2022): *Lernen*. In: Feldmann, M./Rieger-Ladich, M./Voß, C./Wortmann, K. (Hrsg.): *Schlüsselbegriffe der Allgemeinen Erziehungswissenschaft. Pädagogisches Vokabular in Bewegung*. Weinheim und Basel: Beltz Juventa, S. 277–286.
- De Weger, E./Van Vooren, N. J. E./Wong, G./Dalkin, S./Marchal, B./Drewes, H. W./Baan, C. A. (2020): *What's in a Realist Configuration? Deciding Which Causal Configurations to Use, How, and Why*. In: *International Journal of Qualitative Methods*, 19. Jg. <https://doi.org/10.1177/1609406920938577> (18.04.2024).
- Dollinger, B. (2018): *Paradigmen sozial- und erziehungswissenschaftlicher Wirkungsforschung: Eine Analyse kausaltheoretischer Annahmen und ihrer Folgen für die Soziale Arbeit*. In: *Soziale Passagen*, 10. Jg., H. 2, S. 245–262.

- Dössegger, A./Weibel, D./Frei, K. M./Wissmath, B./Hense, J. (2017): Entwicklung eines Wirkmodells für die Evaluation des Programms Jugend und Sport. In: Zeitschrift für Evaluation, 16. Jg., H. 1, S. 97–120.
- Ehnert, K./Johann, T./Mielke, A./Rehse A./Walter, E./Zimmermann, E. (2021): Handlungsfeld und Handlungslogiken. Wissenschaftliche Begleitung der Modellprojekte im Handlungsfeld Demokratieförderung im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Unter Mitarbeit von C. Wagner. Halle (Saale): DJI.
- Figlesthler, C./Greuel, F./Grunow, D./Langner, J./Schau, K./Schott, M./Zierold, D./Zschach, M. (2019): Vierter Bericht: Modellprojekte E. Wissenschaftliche Begleitung der Modellprojekte zur Radikalisierungsprävention. Programmevaluation „Demokratie leben!“. Zwischenbericht 2018. Halle (Saale): DJI.
- Flick, U. (2009): Qualitative Methoden in der Evaluationsforschung. In: Zeitschrift für Qualitative Forschung, 10. Jg., H. 1, S. 9–18.
- Haubrich, K. (2006): Wirkungsannahmen sichtbar machen: Cluster-Evaluation innovativer multizentrischer Programme. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 101–122.
- Haubrich, K. (2009): Sozialpolitische Innovation ermöglichen. Bd. Die Entwicklung der rekonstruktiven Programmtheorie-Evaluation am Beispiel der Modellförderung in der Kinder- und Jugendhilfe. Internationale Hochschulschriften 536. Münster, New York, München und Berlin: Waxmann.
- Haubrich, K./Baumgartner, E. (2017): Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview. In: Zeitschrift für Evaluation, 16. Jg., H. 1, S. 121–145.
- Hense, J./Taut, S. (2021): Wie Wirkungsmodelle zur Wirkung kommen. Nutzungsvarianten, Mehrwert und Kosten ihrer Verwendung in der Evaluationspraxis. In: Zeitschrift für Evaluation, 20. Jg., H. 2, S. 267–292.
- Hilliger, B. (2024): Scheitern als Lernanlass für strukturelles Lernen. Eine systemische Perspektive auf den Gegenstand der Sozialen Arbeit. In: Kessler, S./König, K. (Hrsg.): Scheitern in Praxis und Wissenschaft der Sozialen Arbeit. Reflexions- und Bewältigungspraktiken von Fehlern und Krisen. Weinheim und Basel: Beltz Juventa, S. 93–106.
- Kessler, S./König, K. (2024): Einführung. Scheitern in Praxis und Wissenschaft der Sozialen Arbeit. In: Kessler, S./König, K. (Hrsg.): Scheitern in Praxis und Wissenschaft der Sozialen Arbeit. Reflexions- und Bewältigungspraktiken von Fehlern und Krisen. Weinheim und Basel: Beltz Juventa, S. 9–26.
- Klawe, W. (2006): Multiperspektivische Evaluationsforschung als Prozess-Wirkungsrekonstruktion. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 125–142.
- Kleinoscheg, A. K. (2020): Kausalität und Plausibilität – Evaluation zwischen Wissenschaft und Praxis. Frühjahrstagung 2019 des Arbeitskreises „Methoden in der Evaluation“ der DeGEval – Gesellschaft für Evaluation e. V. In: Zeitschrift für Evaluation, 19. Jg., H. 2, S. 372–378.
- Laux, H. (2021): Innovation als Prozess der Netzwerkbildung. Thomas P. Hughes und die Akteur-Netzwerk-Theorie. In: Blätzel-Mink, B./Schulz-Schaeffer, I./Windeler, A. (Hrsg.): Handbuch Innovationsforschung. Sozialwissenschaftliche Perspektiven. Wiesbaden: Springer VS, S. 163–179.
- Lüders, C./Haubrich, K. (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe. Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: DJI, S. 5–23.
- Mayer-Drawe, K. (2008): Diskurse des Lernens. München: Fink.
- Nielsen, S. B./Lemire, S./Tangsig, S. (2022): Unpacking context in realist evaluations: Findings from a comprehensive review. In: Evaluation, 28. Jg., H. 1, S. 91–112.

- Nohl, A.-M. (2017): Interview und Dokumentarische Methode. Anleitungen für die Forschungspraxis. 5., aktualisierte und erweiterte Auflage. Wiesbaden: Springer VS.
- Pawson, R./Tilley, N. (1997): Realistic Evaluation. Thousand Oaks, CA: Sage.
- Sattler, D. (2016): Rechne mit dem Unerwarteten! Überlegungen zum Scheitern in der Sozialarbeit. In: Evangelische Jugendhilfe, H. 3, S. 169–173.
- Schau, K./Greuel, F. (2021): Gegenstandsangemessenes und ausgewogenes Bewerten als Herausforderung – Umgangsweisen mit Komplexität, Mehrdeutigkeit und Kontextgebundenheit in der wissenschaftlichen Begleitung. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluation von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 108–127.
- Schulz-Schaeffer, I. (2021): Innovation im Verhältnis zu Neuheit und Wandel. In: Blättel-Mink, B./Schulz-Schaeffer, I./Windeler, A. (Hrsg.): Handbuch Innovationsforschung. Sozialwissenschaftliche Perspektiven. Wiesbaden: Springer VS, S. 25–42.
- Strübing, J./Hirschauer, S./Ayaß, R./Krähnke, U./Scheffer, T. (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. In: Zeitschrift für Soziologie, 47. Jg., H. 2, S. 83–100.
- The RAMESES II Project (2017): What realists mean by context; or, Why nothing works everywhere or for everyone. https://www.ramesesproject.org/media/RAMESES_II_Context.pdf (Abruf 11.01.2024).
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./Lobermeier, O./Heitmeyer, W. (Hrsg.): Evaluation von Programmen und Projekten für eine demokratische Kultur. Wiesbaden: Springer VS, S. 41–68.
- Worthen, B. R./Schmitz, C. C. (1997): Conceptual Challenges Confronting Cluster Evaluation. In: Evaluation, H. 3, S. 300–319.

7 Resonanz ist mehr als die Sehnsucht nach Wiederhall

Zur empirischen Erfassung von Resonanzen in der Evaluation komplexer Programme

Franziska Heinze, Sarah Langer, Steffen Loick Molina, Stefanie Reiter, Kornelia Sammet und Ellen Schroeter

In Programmevaluationen sind Evaluierende oftmals mit Fragen der Untersuchung bzw. des Nachweises von Wirkungen konfrontiert. Dabei stellt sich die nicht triviale Frage, welche Wirkungen eines Vorhabens konkret erwartet und wie eingetretene Wirkungen empirisch festgestellt und bewertet werden können. Im Vorfeld ist dafür zu klären: Wie lassen sich Wirkungen als Folge konkreter Maßnahmen beschreiben und empirisch prüfen bzw. beobachten? Wo muss man empirisch hinschauen, um Wirkungen welcher Art zu finden?

Ausgangspunkt unserer Überlegungen zu Wirkungsuntersuchungen sind zwei gegenstandsbezogene Herausforderungen, die wir kurz umreißen wollen: (1) Im Kontext komplexer, emergenter und sich dynamisch verändernder Bundesprogramme finden wir einen Evaluationsgegenstand vor, der sich über mehrere Ebenen (z. B. Gesamtprogramm, Teilbereiche, Maßnahmenbündel, Einzelmaßnahmen/-projekte) erstreckt. Für weitere empirische Untersuchungen muss dieser Gegenstand daher zunächst durch ein programmtheoriebasiertes Vorgehen konkretisiert werden (Heinze et al. 2021, S. 198 f.). In dieser Konstellation soll Evaluation programmbezogene Wirkungen feststellen und bewerten. Dabei soll sie insbesondere Wirkungen auf einer übergeordneten Ebene betrachten, die sich aus dem Zusammenspiel der Interventionen und (adressat:innenbezogenen) Wirkungen ergeben. (2) Vor diesem Hintergrund kommt den Adressat:innen von Maßnahmen eine in andere Ebenen hinein vermittelnde Funktion (Blatter / Schelle 2022) zu: Sie sollen dazu beitragen, die aus Interventionssicht angestrebten Wirkungen bei weiteren Zielgruppen zu erzielen. Dies birgt nicht nur Herausforderungen für die Zurechnung von Wirkungen zu konkreten Maßnahmen, sondern erfordert eine empirische Herangehensweise, die hinreichend offen für solche im Prozess erst entstehenden, kontingenten Wirkungen ist.

Bei der Suche nach einem erkenntnisleitenden Konzept, das diese Herausforderungen angemessen zu berücksichtigen erlaubt, sind wir auf verschiedene Konzepte von Resonanz gestoßen. Daran interessant erschien uns der erkenntnisoffene, prozessorientierte Fokus darauf, welchen Wiederhall konkrete Impulse

finden und warum manche Impulse ein vielfaches Echo erzeugen, andere hingegen verhallen. Das in diesem Zusammenhang prominente Begriffsverständnis von Rosa (2016), auf das der Titel des Beitrags anspielt, profiliert Resonanz soziologisch als Gegenkonzept zu gesellschaftlicher Beschleunigung und Entfremdung. Wir hingegen wollen nachvollziehen, wie Impulse sich über Individuen und Artefakte über mehrere Ebenen in Organisationen (und letztlich: Systemen) entwickeln, also Resonanz erzeugen. Ein solches Verständnis von Resonanz nutzen wir hier als erkenntnisleitende Perspektive und diskutieren, ob bzw. inwieweit das damit verbundene Vorgehen einen Erkenntnismehrwert für die von uns formulierten Herausforderungen einer wirkungsorientierten Evaluation bietet.

Im Folgenden stellen wir zunächst unseren Evaluationsgegenstand kurz vor und legen dar, welche gegenstandstheoretischen Vorbedingungen in einem wirkungsorientierten Evaluationsdesign berücksichtigt werden müssen. Wir erläutern anschließend das von uns genutzte Konzept der Resonanz und arbeiten heraus, welche erkenntnisleitenden Prämissen sowie methodologischen und methodischen Anforderungen dieses Konzept für die empirische Untersuchung des Evaluationsgegenstands bereithält. An zwei Fallstudien aus unserer Evaluationspraxis reflektieren wir dann unser methodisches Vorgehen. Dabei fokussieren wir die Frage, inwieweit wir die Ansprüche der Gegenstandsangemessenheit und des Resonanzkonzepts einlösen. Entlang der beiden Fallstudien zeigen wir auf, wie das empirische Vorgehen Erkenntnisse zu Resonanzen ermöglicht bzw. limitiert. Schließlich diskutieren wir im Fazit die Potenziale und Herausforderungen einer Resonanzperspektive unter Bezug auf andere Ansätze sowie Lernerfahrungen für wirkungsorientierte Evaluationen.

7.1 Wirkungsorientierte Evaluation von Transferaktivitäten

In der Evaluationsliteratur wird eine ganze Bandbreite an Ansätzen zur Untersuchung von Wirkungen mit jeweils unterschiedlichen Begrifflichkeiten und Erkenntnisinteressen diskutiert. Eine einheitliche Definition bzw. Verwendung des Wirkungsbegriffs findet sich nicht (Maikämper 2003, S. 18 ff.). Herangehensweisen, die Effekte von standardisierten Interventionsprogrammen messen (z. B. Campbell/Stanley 1963) werden häufig als Goldstandard proklamiert. Der kausale Wirkungsnachweis setzt unter anderem einen standardisierten Impuls und eine der Messung vorgängige und theoriebasierte Vorstellung darüber voraus, welche Wirkung(en) einer Intervention bei wem bzw. wo auftreten soll(en).

Daneben existiert eine breite Palette an weiteren Evaluationsansätzen, in denen Theorien im Rahmen von Wirkungsuntersuchungen eine bedeutende Rolle zugeschrieben wird. Solche programmtheoriebasierten Ansätze formulieren Annahmen darüber, wie Veränderungen bei den Adressat:innen und in Zielsystemen (z. B. Kinder- und Jugendhilfe, öffentliche Verwaltung) auf Interventionen („Pro-

gramme“) zurückgeführt werden können (Werthern 2020; Funnell/Rogers 2011).¹ Über eine solche theoriebasierte Attribuierung (Plausibilisierung) von Veränderungen bei den Adressat:innen bzw. in Zielsystemen und deren Einordnung als Wirkung hinaus können theoriebasierte Evaluationen wesentlich zum Verständnis davon beitragen, wie eine Intervention bei wem und unter welchen Bedingungen zu welcher Wirkung führt. Als gegenstandsbezogene Theorien sind sie dabei von begrenzter Reichweite: Gegenstand und (Programm-)Theorie stehen dabei in einem Verhältnis „phänomenologische[r] Nachgiebigkeit und [...] beständig[er] rekursiv[er] Selbstkorrektur“ (Strübing et al. 2018, S. 87). Programmtheoriebasierte Ansätze identifizieren beispielsweise den konkreten Beitrag einer spezifischen Maßnahme zu einer Veränderung (Contribution Analysis, z. B. Mayne 2019) oder nehmen die kausalen Zusammenhänge innerhalb der Wirkungskette von Intervention zu Outcomes in den Blick (Process Tracing, z. B. Schmitt/Beach 2015). Schließlich existieren auch Ansätze wie die Realistic Evaluation (z. B. Haunberger/Baumgartner 2017; Pawson/Tilley 1997), die die spezifischen Wirkungszusammenhänge von Kontexten, Mechanismen und Outcomes untersuchen und herausarbeiten, wie und warum Maßnahmen bei unterschiedlichen Adressat:innen wirken.

Auch wir nutzen ein programmtheoriebasiertes Vorgehen, um die zugrundeliegenden Veränderungsannahmen und die angestrebten Wirkungen unseres Evaluationsgegenstands herauszuarbeiten: Organisationsverbünde („Kompetenznetzwerk“) und Fachorganisationen („Kompetenzzentrum“) sollen im Bundesprogramm „Demokratie leben!“ unter anderem den „Transfer von erfolgreichen Arbeitsansätzen in Bundes-, Landes- und kommunale Strukturen“ übernehmen und zugleich andere Akteure in ihren Transferaktivitäten unterstützen (BMFSFJ 2019, S. 4). Als Wirklogik² arbeiteten wir heraus, dass Verbünde und Organisationen sich dieser Aufgabe mit vielfältigen Transferaktivitäten widmen, die sich häufig an (pädagogisches) Fachpersonal, Verwaltungsmitarbeiter:innen oder Multiplikator:innen richten, die die Transferimpulse wiederum in ihren

-
- 1 Gegenstandsbezogene Theorien bilden die Basis für die empirische Untersuchung der Intervention. Sie rücken in den Blick, welche Aspekte für eine empirische Untersuchung von Interesse sind, und beeinflussen die Wahl des Evaluationsdesigns. Dabei werden die konkreten methodischen Entscheidungen jedoch nicht nur durch die ontologischen Annahmen über die Beschaffenheit des Untersuchungsgegenstands geprägt. Sie enthalten auch unterschiedliche epistemologische (z. B. positivistische vs. post-positivistische) Positionen zu den Voraussetzungen und Grenzen von Erkenntnis sowie damit einhergehende methodologische Erwägungen, wie und mit welchen Methoden wir Erkenntnisse überhaupt nachvollziehbar machen können (Lemire/Whynot/Montague 2019; Kelle 2018).
 - 2 Ausgangspunkt programmtheoriebasierter Evaluationen ist häufig eine theoretische Modellierung von Ursachen-Kontext-Maßnahmen-Wirkungszusammenhängen (beispielsweise Heinze et al. 2021; Dössegger et al. 2017; Coryn et al. 2011), die empirisch konkretisiert wird.

jeweiligen Arbeitskontexten und Bezugssystemen weitenutzen sollen (Heinze et al. 2022, S. 54 ff.).

In Anschluss an aktuelle Erkenntnisse der Transferforschung betrachten wir Transfer „als komplexen, sozialen Prozess, der innerhalb von vielschichtigen und dynamischen Systemen stattfindet“ (Blatter/Schelle 2023, S. 9), das heißt als komplexen, mehrstufigen und wechselseitigen Kokonstruktionsprozess (Heinze et al. 2023, S. 48 ff.). Die Transferwege können sehr unterschiedlich sein und über verschiedene Ebenen verlaufen. Hieran werden die eingangs beschriebenen Herausforderungen für wirkungsorientierte Evaluation sichtbar: Transferimpulse sind in soziale Interaktionen eingebettet, die verschiedene Anschlüsse (z. B. Aufgreifen, Adaptieren, Weitertragen, Verklingen) ermöglichen und damit potenziell vielzählige und vielgestaltige Formen des Widerhalls hervorrufen.

Ausgehend von diesen gegenstandstheoretischen Annahmen untersuchen wir in einer wirkungsorientierten Evaluation die potenziell vielzähligen und vielgestaltigen Formen von Widerhall. Unser Erkenntnisinteresse richtet sich dabei nicht primär auf die unmittelbaren und vorab erwarteten Effekte einer Intervention oder auf kleinteilige Wirkmechanismen bzw. kausal aufeinander folgende Prozesselemente sowie deren Zusammenhänge. Mit dem Konzept der Resonanzen stellen wir im Folgenden vielmehr eine erkenntnistheoretische Heuristik zur Diskussion, die es zwar einerseits ermöglicht, die programmtheoretisch formulierten Wirkungen in den Blick zu bekommen. Andererseits ist eine solche Resonanzperspektive jedoch hinreichend offen für weitere im Transferprozess aufscheinende Formen des Widerhalls.

7.2 Resonanz als erkenntnisleitende Perspektive auf Transferaktivitäten

Für unsere Untersuchung von Transferprozessen greifen wir erkenntnistheoretische Überlegungen von Breyer und Gerner im Anschluss an das ursprünglich aus der Akustik stammende Konzept der Resonanz auf, wonach Resonanz „eine Bewegung voraus[setzt], die auf das Subjekt zukommt, seiner Reaktion also vorausgeht und es passiv affiziert. Um dann aber tatsächlich mitschwingen zu können, ist von Seiten des Angesprochenen eine Resonanzfähigkeit und ein wenn auch noch so minimales Gegenschwingen erforderlich, welches aktiv zu leisten ist“ (Breyer/Gerner 2017, S. 45).

Dieses Konzept der Resonanz betont das aktive Ein- oder Gegenschwingen aufseiten der Adressat:innen. In seiner Übertragung auf die Untersuchung von Transfer als Kokonstruktionsprozess kommen somit komplexe *Prozesse des Auf-*

*greifens- und Übersetzens bzw. Adaptierens*³ der von den Transferaktivitäten ausgehenden Impulse aufseiten der Adressat:innen in den Blick. Damit einhergehend verweist der Resonanzbegriff auf den „polythetischen“ Charakter von Resonanz (ebd., S. 41), also ihre *Vielstrahligkeit*, die „durch mehrere aufeinanderfolgende intentionale Bezugnahmen charakterisiert“ ist (ebd.). Die damit einhergehende multidirektionale Gerichtetheit von Resonanz gilt es für die empirische Umsetzung zu berücksichtigen und auch Rückwirkungsprozesse zwischen Impulsgebenden und -aufnehmenden in den Blick zu nehmen. Dabei stellt sich die Frage, inwieweit die Beteiligten sich in ihren Interaktionen oder der Rezeption eines Impulses mehr oder weniger intensiv affizieren und transformieren, indem sie sowohl sich selbst als auch ihre Beziehung zueinander verändern. So kommt vor dem Hintergrund ihrer Vielstrahligkeit in den Blick, dass Resonanzen sich gegenseitig verstärken, abschwächen oder gar aufheben können (Heinze et al. 2021, S. 219). Der Resonanzbegriff umfasst damit *unterschiedliche Ausprägungen von Veränderungen* (Affizieren, Verklingen, Repulsion). Anschlussfähig zur Prozesshaftigkeit von Transfer bezeichnet Resonanz schließlich ein „Zwischenereignis, das sich zwischen mehreren Entitäten abspielt“ (Breyer/Gerner 2017, S. 36), und fordert dazu auf, die *zeitliche Situiertheit* von Transferprozessen zu berücksichtigen. Für die Evaluation von Transferaktivitäten folgt daraus idealerweise eine multiperspektivische, begleitende und mehrstufige Erhebung zu verschiedenen Zeitpunkten, die sowohl den prozessualen Charakter von Resonanz als auch ihre Eingebundenheit in je spezifische Rahmenbedingungen in den Blick nimmt.

7.3 Erkenntnispotenziale der Resonanzperspektive: zwei Fallstudien

Basierend auf einem mehrstufigen, multiperspektivischen und multimethodischen Design haben wir im Jahr 2023 verschiedene Transferaktivitäten (Fortbildungen, Fachveranstaltungen, Beratungsprozesse und Produkte, ausführlich dazu Heinze et al. 2024) in einem zeitlich begrenzten Ausschnitt in den Blick genommen. Für die Reflexion der Erkenntnispotenziale der Resonanzperspektive in diesem Beitrag wählten wir zwei Fallstudien aus: einen Beratungsprozess sowie ein Modul der politischen Bildung. Diese Fallstudien unterzogen wir einer Metaanalyse, indem wir Erhebungsdesign (Konzeption, Interviewleitfäden etc.), Daten (Interviews, Beobachtungsprotokolle, Dokumente) und Forschungsno-

3 Beispielsweise verweisen Blatter/Schelle (2023, S. 9) unter Rückgriff auf Dewe (2005) und Beck/Bonß (1989) darauf, dass innerhalb von Transferprozessen Wissen nicht einfach weitergegeben werden kann, sondern in Interaktionen diskutiert, reflektiert und für das eigene Praxisfeld relationiert wird. Das Wissen wird währenddessen transformiert und entfaltet damit seine Relevanz für die Akteur:innen.

tizen auf den Umgang mit den zentralen erkenntnisleitenden Prämissen der Resonanzperspektive (siehe Abschnitt 7.2) hin auswerteten und verglichen.

Fallstudie 1: Resonanzen eines Beratungsprozesses

Als ersten Fall ziehen wir die Beratung eines Vereins heran, deren Ziel es war, Unterstützung beim Aufbau von Angeboten der Jugendarbeit zu leisten, zum Beispiel bei der Gewinnung von Jugendlichen sowie für die Akquise von Räumlichkeiten, Kooperations- und Fördermöglichkeiten. Der Beratungsprozess umfasste ein erstes Beratungsgespräch und zwei Folgegespräche zwischen dem Vereinsvorstand und zwei Berater:innen. Dabei wurden in das letzte Gespräch auch zwei jugendliche Engagierte aus dem Verein einbezogen.

Für die Untersuchung der Resonanzen dieses Beratungsprozesses führten wir zunächst ein Vorgespräch mit den beiden Berater:innen durch, bei dem unter anderem erste Kontextinformationen und Erwartungen an den Beratungsprozess erhoben, Möglichkeiten der empirischen Begleitung abgeklärt und die Einwilligung der beteiligten Personen eingeholt wurden. Anschließende Recherchen und Dokumentenanalysen zum Verein erlaubten einen ersten Einblick in dessen Kontext und Aktivitäten. Zur Wahrung der Vertraulichkeit und auf Wunsch der Beteiligten verzichteten wir auf eine teilnehmende Beobachtung des Beratungsgesprächs. Stattdessen entwickelten wir eine Vorlage zur Dokumentation des Verlaufs und der Ergebnisse des ersten Beratungsgesprächs. Mit einem zeitlichen Abstand von drei bzw. vier Wochen führten wir anschließend ein leitfadengestütztes Interview mit dem Vereinsvorstand sowie ein leitfadengestütztes Gruppeninterview mit den drei beratenden Personen durch. Damit sollte retrospektiv die Motivation zur Inanspruchnahme der Beratung, die mit der Beratung jeweils verbundenen Erwartungen und Ziele sowie eine bilanzierende Einschätzung der Beratungssituation und des tatsächlichen Beratungsverlaufs erhoben werden. Mit weiterem zeitlichen Abstand von zwei bzw. drei Monaten wurden sowohl die beratene als auch die primär verantwortliche beratende Person erneut interviewt. Dabei ging es darum, welche Impulse aus der Beratung aufgegriffen wurden, welche Veränderungen davon ausgehend angestoßen wurden und welche Bedingungen ein Aufgreifen der Beratungsimpulse ermöglichen oder verhindern. In allen Erhebungsphasen wurden die bereits erhobenen Daten im Sinne eines iterativen Vorgehens dafür genutzt, konkrete Nachfragen sowie zusätzliche Erzählimpulse in die weiteren Befragungen einzubeziehen.

Fallstudie 2: Resonanzen eines Moduls der politischen Bildung

Der zweite Fall betrachtet den Prozess, wie ein Modul für den Einsatz in Settings der (außer-)schulischen politischen Bildung erprobt und entwickelt wird. In einer Vorabversion wurde jenes mehrfach an Bildungseinrichtungen mit Gruppen junger Menschen erprobt sowie bei einer Fortbildung für Lehrkräfte vorgestellt und getestet. Daran waren eine oder zwei Personen, die das Modul entwickelten,

beteiligt. Die Ergebnisse der Erprobung der Vorabversion sollten fortlaufend in die Weiterentwicklung des Moduls einfließen.

Für die Untersuchung von Resonanzen der Modulerprobung fand zunächst ein Vorgespräch mit den Durchführenden statt. Parallel dazu wurden Dokumente zum Modul analysiert. Das Kernstück der Resonanzuntersuchungen bildeten teilnehmende Beobachtungen bei drei Erprobungen in Bildungssettings. Dabei zielte der methodische Zugang auf Interaktionsverläufe, Aneignungsprozesse und situative Reaktionen der Teilnehmenden. Im Anschluss daran wurden leitfadengestützte Einzelinterviews mit den beteiligten pädagogischen Fachkräften sowie ein bilanzierendes Nachgespräch mit den Durchführenden realisiert.

Im Folgenden reflektieren wir beide Fallstudien hinsichtlich der erkenntnistheoretischen Prämissen des Resonanzkonzeptes (siehe Abschnitt 7.2), unserer methodischen Umsetzung sowie der daraus resultierenden Erkenntnispotenziale und -grenzen.

7.3.1 Aktives Gegenschwingen: Prozesse des Aufgreifens, Adaptierens und Übersetzens

Aus den gegenstandstheoretischen Überlegungen zu Transfer (Abschnitt 7.1) leitet sich methodologisch der Anspruch ab, dass Prozesse der Kokonstruktion bei der Untersuchung von Transferaktivitäten zu berücksichtigen sind. Dementsprechend geht auch das Resonanzkonzept methodologisch von einem Subjektverständnis aus, das aktiv an der Herstellung von Resonanz mitwirkt. Daran anknüpfend konzentrieren sich unsere Erhebungen im untersuchten Beratungsfall auf den primär beteiligten Personenkreis. Die Anlage der Erhebung (Interviewbefragungen der beratenen und der beratenden Personen zu zwei Zeitpunkten) nimmt dabei eine – zunächst auf wenige Perspektiven enggeführte – Fokussierung vor, bei wem bzw. mit welchen subjektiven und situativen Perspektivierungen Resonanzen beobachtet werden (können). Beispielsweise reflektiert die beratene Person ca. drei Wochen nach dem ersten Beratungsgespräch, welche Veränderungen sich ausgehend von der Beratung entwickelt haben:

„[D]ie erste Sitzung war Beratung, aber die zweite, es ist wie eine Unterstützung, nicht nur Beratung. [...] und da haben wir das, wie können wir mein Vorhaben, mein Anliegen entwickeln, haben wir ausgetauscht. [...] Und wir suchen zusammen, wie können wir das eigentlich verwirklichen? [Der Berater], er ist sehr bereit, er kann mich unterstützen bei den [Förder-]Mitteln und bei dem Netzwerk und so weiter. Aber ich habe [dem Berater] mitgeteilt, dass ich eigentlich, ich habe jetzt momentan so gedacht, meine erste Priorität ist, eine richtige geeignete Zielgruppe zu finden. Und kann ich vielleicht im Sommer/bis Sommer diese Zielgruppe finden“ (Beratung-I_I_Beratungsnehmer:in_t1, Z. 250–261).

Im Sinne des Resonanzkonzeptes kommt hier ein aktives Aufgreifen der Impulse aus dem Beratungsgespräch in Form der reflexiven Konkretisierung des eigenen Vorhabens zum Ausdruck: der Aufbau der Jugendarbeit und die Identifikation konkreter Unterstützungsressourcen. Zeitlich nachgelagerte Resonanzen werden als geplante Schritte (Zielgruppengewinnung) vorausgedacht. Sie verbleiben nah an dem in die Beratung eingebrachten Anliegen und sind kongruent mit den angestrebten Zielen der Beratung.

Hinweise auf Resonanzen auf weiteren Ebenen finden sich auch im Interview mit der beratenden Person zum zweiten Zeitpunkt, in dem darüber berichtet wird, welche Impulse im Nachgang der Beratung aufgegriffen wurden. Deutlich wird daran, dass sich im zeitlichen Verlauf Resonanzen entwickeln, die über die zunächst angelegte Zielstellung des Beratungsprozesses hinausweisen.

„[A]usgehend von dem Beratungsprozess haben sich jetzt primär zwei Sachen entwickelt. Einmal haben wir eine feste Zusammenarbeit mit den schon mal angesprochenen Jugendlichen [des Vereins] entwickelt. [...] Genau, die nehmen mittlerweile regelmäßig an unseren Veranstaltungen teil [...] und haben eigene Beratungsgespräche jetzt in Anspruch genommen, um ihr eigenes Engagement [...] zu entwickeln [...]. Und auch aufgrund dieses Beratungsgesprächs, aber auch anderen, haben wir in Kooperation mit [einer weiteren Organisation] einen Workshop, einen Empowerment-Workshop gemacht oder werden wir machen [...], also ansonsten ist die direkte Zusammenarbeit mit [dem Vereinsvorstand] jetzt nur auf regelmäßige bilaterale Gespräche eingeschränkt“ (Beratung-I_I_Berater:in_t2, Z. 8–29).

Insbesondere der Beginn des Ausschnitts verdeutlicht, dass die Beratung auch nicht geplante Resonanzen bei weiteren Personen (Jugendliche) hervorbringt. Jene Jugendlichen waren zwar teilweise in den Beratungsprozess involviert, kommen aber im Rahmen der Resonanzerhebungen nicht selbst zu Wort. Ihre Bedeutung als wichtige Resonanzsubjekte wird vielmehr über die fachliche Perspektive der Berater:innen auf die Bedarfe und Anliegen der Jugendlichen (u. a. „eigene Beratungsgespräche [...] in Anspruch genommen“, „eigenes Engagement [...] zu entwickeln“) vermittelt in den Daten sichtbar.

Im Vergleich zum untersuchten Beratungsprozess sind für das Bildungsmodul Prozesse des Aufgreifens, Adaptierens und Übersetzens zentral im Gegenstand angelegt, da es zunächst erprobt wird. Auf Basis von Erfahrungswerten mit den zumeist jugendlichen Teilnehmenden als auch über die reflexiven Perspektiven von pädagogischen Fachkräften soll jenes im Prozess weiterentwickelt werden. Eine Lehrkraft, die an einer Erprobung teilgenommen hat, reflektiert im Interview, welche schulischen Bedingungen und Übersetzungsleistungen die Resonanzen der Arbeit mit dem Modul beeinflussen und in dessen Weiterentwicklung zu reflektieren wären:

„Darüber hinaus, würde ich sagen, die – es gibt bestimmte Schritte, die begünstigen können, dass eine Diskussion in Gang kommt, dass die Leute Problembewusstsein kriegen dafür, was sie da eigentlich debattieren sollen. [...] wenig Textarbeit, möglichst viel die Leute in Aktion halten“ (Produkt-II_I_Lehrkraft, Z. 54–62).

Die im Interview thematisierten Resonanzbedingungen bleiben situativ dem Erprobungskontext verhaftet: Welche thematischen oder didaktischen Adaptionen die Lehrkraft beim zukünftigen Einsatz des Moduls tatsächlich vornimmt und welche situativen Störfaktoren dabei zu bewältigen sind, können nur über erste Hinweise aus dem Material geschlussfolgert werden.

Mit dem methodischen Zugang über die teilnehmende Beobachtung wurden – anders als in der Fallstudie zur Beratung, in der erzählerische Perspektiven auf bereits eingetretene oder erwartete Resonanzen erhoben wurden – auch Interaktionsverläufe und situative Reaktionen fokussiert, zum Beispiel auf den Erprobungsablauf oder die Rahmenbedingungen. Damit wollten wir Prozesse der interaktiven Aneignung und der Kokonstruktion zwischen den an der Erprobung beteiligten Durchführenden sowie den Lehrkräften empirisch in den Blick bekommen. In den Beobachtungsprotokollen finden sich Sequenzen, in denen Lehrkräfte vor dem Hintergrund ihres spezifischen schulischen Kontextes die Umsetzungsbedingungen des Moduls hinterfragen und die Durchführenden dazu Anpassungen vorschlagen.

„Lehrkraft: Muss das im Block von fünf Stunden durchgeführt werden? Oder geht es auch wochenweise mal eineinhalb Stunden?“

Durchführende:r 1: Wir haben es immer im Block gemacht, hängt wohl von der Gruppe ab.

Durchführende:r 2: Hauptsache die Spielphasen werden nicht unterbrochen, so wird es auch beschrieben“ (Produkt-II_tB_Fortbildung).

Deutlich werden Übersetzungsprozesse in das Arbeitsfeld Schule: Aus einer programmtheoretischen Perspektive sind es die pädagogischen Fachkräfte, die aus ihrer Perspektive über die Einsatzmöglichkeiten des Moduls im Unterricht sowie über relevante Adaptionen entscheiden und denen damit eine zentrale Bedeutung für das Weiterschwingen des Impulses zukommt.

Neben den fachlichen Perspektiven der Lehrkräfte auf die Adaption des Moduls sind von den Durchführenden Resonanzen aufseiten der Jugendlichen im iterativen Erprobungsprozess angelegt: Die beobachteten Reaktionen der teilnehmenden Schüler:innen sollen Aufschluss über mögliche Anpassungsbedarfe geben. Jugendliche und junge Erwachsene fungieren in dieser Hinsicht

auch im Rahmen von Befragungen als Wissensressource zur bedarfsgerechten Weiterentwicklung des Moduls:

„Durchführende:r 1 [...] kommt auf das [Modul] zu sprechen: ‚Wir möchten das [Modul] erproben und weiterentwickeln. Dafür sind wir auf Eure Hilfe angewiesen. Wir möchten eine Befragung durchführen‘“ (Produkt-II_tB_Erprobung_II).

Aus der programmtheoriebasierten Entscheidung, die teilnehmende Beobachtung der Modulerprobung auf die Resonanzen der Fachkräfte zu fokussieren, ergibt sich eine erkenntnisbezogene Leerstelle bezüglich der Resonanzen bei den Jugendlichen und jungen Erwachsenen: Welchen situativen oder längerfristigen Mehrwert ziehen sie aus der Erprobung? Wie greifen sie die Impulse der Teilnahme am Modul in weiteren lebensweltlichen Kontexten auf? Die Perspektiven der Jugendlichen werden in der empirischen Anlage der Resonanzerhebung lediglich vermittelt durch die Wahrnehmung und Deutung der Durchführenden und der Fachkräfte sichtbar:

„Lehrkraft: Es fehlt im [Modul] teilweise der Lebensweltbezug, weil nicht alle Schüler:innen aus dem städtischen Raum kommen.

Durchführende:r 1: Auch wenn der direkte Bezug fehlt, sind die Schüler:innen bereit, sich damit auseinanderzusetzen, zumal es immer auf sie zukommen kann. [...] Außerdem wird vorgeschlagen, dass ‚stillere‘ Schüler:innen mit den Rollen so bedacht werden, dass sie sich gut einbringen können“ (Produkt-II_tB_Fortbildung).

Die in diesem Ausschnitt aus einem Beobachtungsprotokoll sichtbaren Übersetzungen durch die pädagogischen Fachkräfte und die Durchführenden der Modulerprobung sind Ausgangspunkt für unterschiedliche Resonanzen, die wir im Folgenden in ihrer Vielstrahligkeit in den Blick nehmen.

7.3.2 „Klangvarianzen“ und „Verklängen“: Vielstrahligkeit und unterschiedliche Ausprägungen von Resonanz

Im Zusammenhang mit der methodologischen Annahme, dass sich Resonanzen innerhalb kokonstruktiver Beziehungsgeflechte entfalten, gehen wir davon aus, dass Resonanzen unterschiedliche Richtungen und Intensitätsgrade aufweisen können. Resonanzen auf einen initialen Impuls können sich im Transferprozess wechselseitig verstärken, abschwächen oder aufheben. So wird im betrachteten Beratungsfall trotz der empirischen Fokussierung der Erhebungen auf den Kreis der primär am Beratungsprozess beteiligten Personen der vielstrahlige Charakter von Resonanz deutlich: Resonanzen entfalten sich aufseiten des Vereins,

aufseiten der Jugendlichen, aber auch aufseiten des Dachverbandes und des Kompetenznetzwerks bzw. -zentrums. Allerdings deutet sich auch ein potenzielles Verklingen an, wenn die beratende Person darauf hinweist, dass – neben den Resonanzen bei den Jugendlichen und im Kompetenznetzwerk bzw. -zentrum – die weitere konkrete, „direkte Zusammenarbeit mit [dem Verein] [...] bisher jetzt nur auf weitere Gespräche, Kaffeetreffen und sowas beschränkt war“ (Beratung-I_I_Berater:in:t2, Z. 32–34).

Der vielstrahlige Charakter des Resonanzkonzeptes umfasst dabei auch Rückwirkungsprozesse. In der betrachteten Beratung werden Rückwirkungen einerseits über das Erhebungsdesign (z. B. als Fragestellung in Interviewleitfäden) in den Erkenntnisraum gehoben, andererseits lassen sie sich auch im Datenmaterial rekonstruieren. Wie oben bereits dargestellt, berichtet die beratende Person, wie der Dachverband spezifische Anliegen der Jugendlichen (z. B. Unterstützung beim Empowerment) aufgreift und in die Umsetzung eigener Angebote für jugendliche Zielgruppen, z. B. Empowerment-Workshops, einfließen lässt.

Der Beratungsprozess und die daraus entstehenden Resonanzen gehen aus der Interaktion von Individuen, Organisationen und Artefakten (z. B. Gegenstände, Dokumente, Handreichungen) in einer konkreten sozialen Situation hervor. Die Beratung selbst ist damit immer als eine kokonstruktive Leistung anzusehen, die Einfluss darauf nimmt, welche Reaktionen und Anklänge sie bei wem zu erreichen vermag. Da das Erhebungsdesign im Fall dieses Beratungsprozesses die Interaktionssituationen und die gemeinsame Erzeugung von Resonanzen nicht in situ in den Blick nehmen konnte, macht sich die Untersuchung die retrospektive und perspektivierte Erzählung über konkrete Interaktionsmomente zunutze. Darüber hinaus wurde in der Erhebungskonzeption eine Vermittlung zwischen unterschiedlichen Perspektiven angelegt, indem das Datenmaterial iterativ für die Generierung von Nachfragen, Erzählimpulsen und Perspektivierungen herangezogen wurde. So zeigt sich, wie die interviewende Person im Gespräch mit der beratenen Person zum ersten Befragungszeitpunkt eine vermittelnde und zeitlich entkoppelte Vermittlung der verschiedenen Perspektiven bzw. Daten vornimmt:

„Interviewer:in: Du hattest es gerade angeschnitten mit der Jugendarbeit, vielleicht ist das die Überleitung ja zu dem Gespräch [erstes Beratungsgespräch, Anmerkung d. A.]. Ich habe halt das Fallprotokoll gesehen, also wurde ja dokumentiert, darüber gesprochen, Jugendarbeit war da ja ein Thema. Und da war auch dokumentiert, dass schon davor Beratungsgespräche stattgefunden haben. Jetzt für mich zur Einordnung in diesem Prozess, diesem Beratungsprozess, wo steht ihr da gerade, am Anfang oder in der Mitte? Gab es vorher schon Gespräche? Das war jetzt ja bestimmt nicht das erste Gespräch, [...] [Beratene:r] Das war erste“ (Beratung-I_I_Beratene:r_t1, Z. 39–47).

Die interviewende Person rekurriert hier nicht nur auf andere Daten(-Erhebungsmomente) und die darüber verfügbaren Informationen, sondern nutzt sie auch für eine Perspektiventriangulation bzw. die Validierung ausgewählter Informationen. Dadurch werden die unterschiedlichen Ausprägungen von Resonanz erst sichtbar. Gleichzeitig ist methodisch zu beachten, dass die Perspektiven der Befragten durch den Bezug auf vorhandene Informationen nicht überformt werden. Beispielsweise führt die Art und Weise der Interviewführung im hier dargestellten Fall Beratung I dazu, dass unterschiedliche Rahmungen (z. B. „Beratungsgespräch“, „Kaffee treffen“) durch gezielte Nachfragen unter Verwendung von Kontextinformationen aus den anderen Interviews nivelliert werden. In der Folge werden verschiedene Gespräche im Verlauf des Interviews als Beratung gerahmt. So wird überhaupt erst möglich, die entstehenden Resonanzen als Resonanzen der Beratung in den Blick zu nehmen. Dies gilt es jedoch methodisch mit Blick auf die Grenzen der Zurechenbarkeit der Resonanzen zum Ausgangsimpuls zu reflektieren.

Auch im Falle der Modularprobung zeigt sich der Einfluss des methodischen Vorgehens auf die Art und Weise, wie die Vielstrahligkeit der Resonanzen im Datenmaterial sichtbar und weiterverfolgt werden. Beispielsweise deuten sich im Protokoll eines Dialoges zwischen den Durchführenden und Schüler:innen situativ Repulsionen hinsichtlich des angestrebten Bildungserfolges bei Letzteren an.

„Durchführende:r 1 fragt die Schüler:innen: Wie viel habt Ihr gelernt? Zusammengefasste Antwort: starke Tendenz: nein. [...]

Durchführende:r 1 fragt die Schüler:innen: Hat das [Modul] motiviert, Euch mehr zu informieren?

Zusammengefasste Antwort: Tendenz: nein“ (Produkt-II_tB_Erprobung_II).

Die Beobachtungssequenz zeigt in ihrem weiteren Verlauf, dass dieses Zurückweisen des Impulses von den Durchführenden nicht weiterverfolgt wird. Entlang von Verständnisfragen von Schüler:innen überlegen sie stattdessen, welche Anpassungen erfolgen könnten, damit eine zukünftige Zurückweisung durch Schüler:innen weniger wahrscheinlich wird:

„Nachfrage Schüler:in: ‚Was ist ein Flash Mob?‘

Erläuterung Durchführende:r 1: ‚Eine spontane Versammlung. Das gab es früher, das kennt ihr vielleicht nicht mehr. Das müsste vielleicht durch etwas Moderneres ausgetauscht werden“ (Produkt-II_tB_Erprobung_II).

Die Resonanzen und damit auch die Repulsionen der Jugendlichen und jungen Erwachsenen kommen in der Erprobungssituation primär im Rahmen der pädagogischen Austauschbeziehung zwischen den beteiligten Fachkräften und den Moduldurchführenden in den Blick. Dabei ist die teilweise Zurückweisung des Impulses als konstitutives Moment für die Optimierung des Moduls und nicht als situativer Bildungsprozess der Jugendlichen und jungen Erwachsenen gerahmt. Vor diesem Hintergrund werden vor allem die zurückwirkenden Resonanzen, die sich auf die Weiterentwicklung des Moduls beziehen, in den Blick genommen.

7.3.3 Intermezzi: Resonanzen als Zwischenereignisse

Mit der zugrunde gelegten erkenntnistheoretischen Perspektive werden Resonanzen als kleinschrittige Zwischenereignisse zwischen verschiedenen Elementen verstanden, die zeitlich situiert sind. Das impliziert, dass Resonanzen als Zwischenereignisse bedeutsame Verknüpfungen zwischen beteiligten Akteur:innen und Artefakten innerhalb eines größeren (zeitlichen) Kontextes darstellen. Da in beiden Fällen zu unterschiedlichen Erhebungszeitpunkten Befragungen stattfanden, ermöglichen sie Erkenntnisse über bereits zurückliegende, sich gerade entfaltende und zukünftige Resonanzen (z. B. geplante nächste Schritte).

Im Fall des Moduls zur politischen Bildung wurden im Interview mit den Durchführenden zunächst mehrere anvisierte Resonanzpfade deutlich: Neben der Weiterentwicklung des Moduls soll auch dessen Dissemination in Bildungseinrichtungen im Rahmen der Erprobungen angestoßen werden. Indem auch pädagogische Fachkräfte in die praktische Durchführung des Moduls eingebunden werden, sollen zum Beispiel etwaige Berührungsängste gegenüber seinem Einsatz im Unterricht abgebaut werden. In dem beobachteten Ausschnitt sind die erfassten Resonanzen zwar auf die Erprobungsmomente bezogen. Im Ergebnis wurde die Erprobung aber in ihrer zugleich vermittelnden Funktion sichtbar als ein Setting, in dem sich Resonanzen sowohl in Richtung der angestrebten Produktweiterentwicklung und -dissemination als auch in Richtung der Lehrkräfte, sogar zueinander, entfalten. Am nachfolgenden Beobachtungsauszug aus einer Lehrkräftefortbildung lassen sich affektive Reaktionen der teilnehmenden Lehrer:innen auf die gemeinsam simulierte Modulerprobung nachzeichnen.

„Nach und nach stellen sich die Gruppen mit viel schauspielerischem Talent vor. [...] Es kristallisieren sich klare Positionen heraus, die Lehrkräfte lachen viel. [...] Es entspinnt sich eine lebhafte und lautstarke Diskussion. Die Gruppen unterbrechen sich, rufen rein, scheinen sich zu streiten. Durchführende:r 1 muss die Diskussionsrunde unterbrechen und schlägt vor, dass alle symbolisch ihre Namenskarten abnehmen, um wieder aus der Rolle zu schlüpfen. Es gibt Applaus für [Name Spielrolle]“ (Produkt-II_tB_Fortbildung).

Die im Verlauf sichtbar gewordenen Resonanzpfade (Produktentwicklung bzw. dessen Dissemination) lassen sich anhand der Momentaufnahme der Affizierung der Lehrkräfte weiterverfolgen. Was jedoch daraus folgt und wie sich das zum Beispiel auf den Abbau möglicher Hemmschwellen zum Einsatz des Moduls im Unterricht auswirken mag, lässt sich auf Basis des zeitlichen Erhebungsausschnittes nicht schlussfolgern. Allerdings lässt sich die Funktion des Erprobungssettings in der Lehrkräftefortbildung als bedeutsames Zwischenereignis für die Verknüpfung unterschiedlicher Resonanzpfade – die Weiterentwicklung des Produkts einerseits, Vorbereitung einer weiteren Nutzung im Praxisfeld andererseits – genauer beschreiben.

7.4 ... mehr als Wiederhall: Diskussion und Fazit

In diesem Beitrag stellen wir unser Vorgehen zur Untersuchung von Transferwirkungen dar. Dabei betrachten wir Resonanzen von Transferaktivitäten, die sich ausgehend von Impulsen über verschiedene Ebenen (Individuen, Artefakte, Organisationen) und vielgestaltige Pfade entfalten, das heißt sich ausbreiten oder verhalten. Entlang unserer Empirie reflektieren wir, wie es gelingt, die gegenstands- und erkenntnistheoretischen Prämissen (Abschnitte 7.1 und 7.2) im konkreten Vorgehen einzulösen. Im Folgenden diskutieren wir, welchen Mehrwert und welche Fallstricke das Konzept der Resonanz als erkenntnisleitende Perspektive für programmtheoriebasierte Evaluationen bereithält.

Mit Blick auf die im Gegenstand angelegten Prozesse des Aufgreifens, Adaptierens und Übersetzens lässt sich zunächst festhalten, dass sich die im Rahmen der Programmtheorie als relevant definierten Perspektiven in der Auswahl der befragten Personen(-gruppen) sowie im Erhebungsfokus spiegeln (siehe Abschnitt 7.3). Die Perspektiven der Durchführenden und der Adressat:innen lassen sich anhand des Datenmaterials wechselseitig auf die jeweilige Transferaktivität beziehen und somit als Veränderungen durch ein aktives Gegen- und Weiter-schwingen des Impulses nachvollziehen. Die Resonanzperspektive ermöglicht, diese Veränderungen als unabgeschlossene „Zwischenereignis[se]“ (Breyer/Gerner 2017, S. 36) zu explorieren. Dabei bleiben die in den Daten sichtbaren Übersetzungs- und Anpassungsprozesse weitgehend auf Impulsgebende und Primäradressat:innen beschränkt. So kommen weitere Resonanzen (beispielsweise von Jugendlichen) primär vermittelt über die Perspektive der Befragten in den Blick (Abschnitt 7.3.1). Damit wird deutlich, dass die theoretisch informierten „Sehinstrumente“ (Lindemann 2008, S. 114) den empirischen Blick lenken. Dies gilt es, bei der Einordnung der Befunde zu reflektieren.

Anhand der methodischen Umsetzung der zwei Fallstudien lassen sich zudem Potenziale und Grenzen aufzeigen, wie Varianzen in den Blick kommen. Deutlich wurde, dass unterschiedliche Resonanzen zwar systematisch in der me-

thodischen Umsetzung berücksichtigt wurden. Beispielsweise explorieren beide Fallstudien auch Repulsionen, die sich aufseiten der beteiligten Jugendlichen zeigen (Abschnitt 7.3.2). Zugleich ließen sich Aspekte des Verklingens von Impulsen nur teilweise in den betrachteten Wirkausschnitten erfassen: So verklingt die direkte Zusammenarbeit von Berater:innen und Verein bzw. die Impulse aus dem Modul verklingen im Bildungsprozess der Jugendlichen (Abschnitt 7.3.2). Hier bleibt die Schwierigkeit bestehen, dass eine auf Interventionserfolge und intendierte Resultate abzielende Perspektive sich oftmals in evaluatives Handeln einschreibt (z. B. in eine implizite Schwerpunktsetzung in der Interviewführung, siehe Abschnitt 7.3.2). Umgekehrt stellt sich die prinzipielle Frage, wie etwas, das nicht stattfindet, methodisch erfasst werden kann. Dem polythetischen Charakter entsprechend gilt es, in der praktischen Umsetzung der Resonanzuntersuchungen immer wieder bewusst auch Repulsionen und das Verklingen von Impulsen zu verfolgen.

In der methodologischen Reflexion wurde zudem deutlich, dass die beobachteten Resonanzpfade (z. B. die Affizierung der Lehrkräfte in der Modulerprobung) nur in einem begrenzten zeitlichen Ausschnitt zugänglich sind. So können veränderte Handlungspraktiken bzw. verändertes Wissen immer nur als „Zwischenereignisse“ (Abschnitt 7.3.3) zu einem bestimmten Zeitpunkt, nicht jedoch bei einer längerfristigen Umsetzung oder in einer möglichen Zukunft erfasst werden. Dennoch bieten die zeitlich situierten Einblicke in die Verkettung verschiedener Zwischenereignisse Aufschluss darüber, wie sich Resonanzen in Entstehungsmomenten gegenseitig beeinflussen können (etwa hinsichtlich unterschiedlicher Zielorientierungen bei der Moduldurchführung). Um die Momentaufnahmen einordnen zu können, bedarf es daher weiterer systematischer Erhebungs- und Auswertungsschritte, z. B. prozessbezogene Vergleichsperspektiven mit einem theoretischen Sampling oder die Berücksichtigung von Erkenntnissen aus anderen Studien und Theorien. Ein Potenzial der im Resonanzkonzept angelegten Prozessperspektive liegt insofern darin, dass auch Erkenntnisse über das Zustandekommen und situative Rahmenbedingungen der Resonanzen generiert werden. Dies geht über die in summativen Wirkungsevaluationen übliche Feststellung von Transferergebnissen zu einem bestimmten Zeitpunkt hinaus. Somit wird die Resonanzperspektive dem kokonstruktiven Charakter des Gegenstands gerecht, wenngleich es in der Umsetzung bei einer zeitlichen und perspektivischen Beschränkung auf Wirkausschnitte bleibt.

Ausgehend von den an unseren Fallstudien aufgezeigten Erkenntnismöglichkeiten und Limitationen der Resonanzperspektive stellt sich die Frage nach dem Erkenntnismehrwert für eine programmtheoriebasierte, wirkungsorientierte Evaluation. Zunächst weitet das Resonanzkonzept den Blick für vielzählige und vielgestaltige Resultate einer Intervention. Dafür werden ausgehend von einer Programmtheorie Prozesse, die plausibel von einer Intervention ausgehen, daraufhin betrachtet, wie sie sich im Zusammenspiel von Akteur:innen, Organi-

sationen und Artefakten untereinander sowie gegebenenfalls über mehrere Ebenen in zeitlichen Etappen und in Varianzen entfalten. Dabei unterscheidet sich das Erkenntnisinteresse nicht nur von summativen Kontroll- bzw. Vergleichsgruppensdesigns, die vorab festgelegte Wirkungszusammenhänge mithilfe von (quasi-)experimentellen Ansätzen überprüfen. Die Resonanzperspektive geht auch über andere (theoriebasierte) Evaluationsansätze wie das Process Tracing (Camacho Garland/Beach 2023) oder die Realistic Evaluation (Haunberger/Baumgartner 2017; Pawson/Tilley 1997) hinaus, die versuchen, die Prozesse zwischen Intervention und Outcome in theoretisch abgeleiteten Bahnen zu erhellen. Das Resonanzkonzept öffnet dagegen den empirischen Blick für alle möglichen (auch in den Zielen nicht angelegten) Resultate einer Intervention und weitet damit den Erkenntnishorizont, wie beispielsweise Ansätze einer *goal free evaluation* (Scriven 1991) oder das *outcome harvesting* (Wilson-Grau 2018). Anders als die beiden letztgenannten Ansätze, die zunächst auf eine theoriebasierte Fokussierung verzichten und möglichst alle Resultate erfassen, verbindet das hier vorgestellte Vorgehen die Erkenntnisoffenheit des Resonanzkonzeptes mit der iterativen Theorie-Empirie-Integration einer programmtheoriebasierten Herangehensweise. Potenzial, das sich aus der Resonanzperspektive ergibt, liegt insofern darin, die theoretische Fokussierung des programmtheoriebasierten Vorgehens im iterativen Prozess auszuweiten und damit die Programmtheorie auszdifferenzieren: So können beispielsweise die Erkenntnisse aus dem Beratungsfall genutzt werden, die Programmtheorie um Aspekte von fluiden, organisationalen Beratungsprozessen zu erweitern, die auch Adressat:innen aus unterschiedlichen Kontexten einbinden. Sowohl die im Forschungsprozess emergierenden Perspektiven, die sich im Sinne des Gegen- und Weiterschwingens eines Impulses als bedeutsam erweisen, als auch unvorhergesehene Resultate können dazu genutzt werden, Wirkpfade auszdifferenzieren. Letzteres bleibt – ähnlich der Suche nach unintendierten oder negativen Wirkungen – eine empirisch herausforderungsvolle und ambivalente Angelegenheit. Ein programmtheoriebasiertes Vorgehen ermöglicht, die Aufmerksamkeit und Ressourcen der Evaluation auf zentrale Aspekte des Programms zu fokussieren (Pawson 2003, S. 486 f.; Weiss 1998, S. 72). Dies birgt zugleich die Gefahr der Engführung der Perspektive auf Erfolg versprechende Wirkpfade und blendet in der Evaluationspraxis oftmals nicht-intendierte und negative Outcomes aus.⁴ Die Resonanzperspektive bietet hier die Möglichkeit, wichtige Aspekte für Interventionen an ihrem „Nicht-Erfolg“ auch jenseits der Überprüfung vorab festgelegter Wirkungszusammenhänge zu

4 Gleichwohl ist es durchaus möglich, auch nicht-intendierte und negative Outcomes a priori oder ex-post in die Entwicklung bzw. Fortschreibung der Programmtheorie einzubeziehen. Dies erfolgt beispielsweise mit dem Ansatz der von Carol Weiss 1998 entwickelten *negative program theory* als Ergänzungsmöglichkeit etablierter Ansätze von Programmtheorien (Rogers 2002, S. 231).

erkennen und dabei die Rahmenbedingungen komplexer Programme zu berücksichtigen. Jedoch ist damit ein erhöhter Ressourcenaufwand verbunden. Zudem besteht die Gefahr, dass einerseits der Begriff der Resonanz und andererseits die Programmtheorie überfrachtet werden mit Aspekten, die für die (Bewertung der) Programmumsetzung nicht relevant sind. Durch den iterativen Prozess muss gewährleistet sein, sich nicht in der Suche nach allen möglichen Resonanzen zu verirren, sondern eine empirisch fundierte Plausibilisierung von – für die weitere Programmgestaltung und -bewertung relevanten – Wirkungen einer Intervention zu ermöglichen und Erkenntnisse über deren Entstehungsprozess zu generieren.

Literatur

- Beck, U./Bonß, W. (1989): Verwissenschaftlichung ohne Aufklärung? Zum Strukturwandel von Sozialwissenschaft und Praxis. In: Beck, U./Bonß, W. (Hrsg.): *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens*. Frankfurt am Main: Suhrkamp, S. 7–45.
- Blatter, K./Schelle, R. (2022): *Wissenstransfer in der frühen Bildung. Modelle, Erkenntnisse und Bedingungen*. München: DJI.
- Blatter, K./Schelle, R. (2023): *Wissenstransfer in der Frühen Bildung. Empirische Erkenntnisse einer qualitativen Studie*. München: DJI.
- BMFSFJ (2019): *Förderauftrag für den Handlungsbereich Bund im Rahmen des Bundesprogramms „Demokratie leben!“ des Bundesministeriums für Familie, Senioren, Frauen und Jugend (02.05.2019)*. Berlin.
- Breyer, T./Gerner, A. (2017): Resonanz und Interaktion. Eine philosophische Annäherung anhand zweier Proben. In: Breyer, T./Buchholz, M. B./Hamburger, A./Pfähder, S./Schumann, E. (Hrsg.): *Resonanz – Rhythmus – Synchronisierung*. Bielefeld: transcript, S. 33–46.
- Camacho Garland, G./Beach, D. (2023): Theorizing how interventions work in evaluation: Process-tracing methods and theorizing process theories of change. In: *Evaluation*, 29. Jg., H. 4, S. 390–409.
- Campbell, D. T./Stanley J. C. (1963): *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.
- Coryn, C. L. S./Noakes, L. A./Westine, C. D./Schröter, D. C. (2011): A Systematic Review of Theory-Driven Evaluation Practice from 1990 to 2009. In: *American Journal of Evaluation*, 32. Jg., H. 2, S. 199–226.
- Dewe, B. (2005): Von der Wissenstransferforschung zur Wissenstransformation. Vermittlungsprozesse – Bedeutungsveränderungen. In: Antos, G./Wichter, S. (Hrsg.): *Wissenstransfer durch Sprache als gesellschaftliches Problem. Transferwissenschaften, Bd. 3*. Frankfurt am Main: Peter Lang, S. 365–379.
- Dössegger, A./Weibel, D./Frei, K. M./Wissmath, B./Hense, J. (2017): Entwicklung eines Wirkmodells für die Evaluation des Programms Jugend und Sport. In: *Zeitschrift für Evaluation*, 16. Jg., H. 1, S. 97–120.
- Funnell, S. C./Rogers, P. J. (2011): *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco: John Wiley & Sons.
- Haunberger, S./Baumgartner, E. (2017): *Wirkungsevaluationen in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview*. In: *Zeitschrift für Evaluation*, 16. Jg., H. 1, S. 121–145.

- Heinze, F./Hemmann, M./Langer, S./Reiter, S./Sammet, K. (2022): Untersuchungen zur Nachhaltigkeit der Förderungen auf Bundesebene. Wissenschaftliche Begleitung Handlungsbereich Bund im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. (Schwerpunktbericht 2021). Halle (Saale): DJI.
- Heinze, F./König, F./Reiter, S./Schroeter, E. (2021): Wirkmodelle praxisnah erarbeiten und für Wirkungsuntersuchungen in komplexen Mehrebenenprogrammen nutzen. In: Milbradt, B./Greuel, F./Reiter, S./Zimmermann, E. (Hrsg.): Evaluation von Programmen und Projekten der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Gegenstand, Entwicklungen und Herausforderungen. Weinheim und Basel: Beltz Juventa, S. 198–221.
- Heinze, F./Langer, S./Loick Molina, S./Reiter, S./Sammet, K./Schroeter, E. (2024): Resonanzen der Angebote von Kompetenznetzwerken und -zentrum. Wissenschaftliche Begleitung Handlungsbereich Bund im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Halle (Saale): DJI.
- Heinze, F./Loick Molina, S./Meijer, L./Reiter, S./Sammet, K./Schroeter, E. (2023): Themenfeldentwicklung. Wissenschaftliche Begleitung Handlungsbereich Bund im Bundesprogramm „Demokratie leben!“ in der Förderphase 2020 bis 2024. Halle (Saale): DJI.
- Kelle, U. (2018): „Mixed Methods“ in der Evaluationsforschung – mit den Möglichkeiten und Beschränkungen quantitativer und qualitativer Methoden arbeiten. In: Zeitschrift für Evaluation, 17. Jg., H. 1, S. 25–52.
- Lemire, S./Whynot, J./Montague, S. (2019): How We Model Matters: A Manifesto for the Next Generation of Program Theorizing. In: Canadian Journal of Program Evaluation, 33. Jg., H. 3, S. 414–433.
- Lindemann, G. (2008): Theoriekonstruktion und empirische Forschung. In: Kalthoff, H./Hirschauer, S./Lindemann, G. (Hrsg.): Theoretische Empirie. Zur Relevanz qualitativer Forschung. Frankfurt am Main: Suhrkamp, S. 107–128.
- Maikämper, M. (2023): Wirkungsanalysen zu Beteiligungsprozessen in der Stadtentwicklung. Hemmnisse, Stellschrauben und Perspektiven. opus4.kobv.de/opus4-btu/files/6140/Maikämper_Moritz.pdf (Abruf 18.04.2024).
- Mayne, J. (2019): Revisiting contribution analysis. In: Canadian Journal of Program Evaluation, 34. Jg., H. 2, S. 171–191.
- Pawson, R. (2003): Nothing as Practical as Good Theory. In: Evaluation, 9. Jg., H. 4, S. 379–510.
- Pawson, R./Tilley, N. (1997): Realistic evaluation. Thousand Oaks, CA: Sage.
- Rogers, P. J. (2002): Program Theory. No Whether Programs Work but How They Work. In: Stufflebeam, D. L./Madaus, G. F./Kellaghan, T. (Hrsg.): Evaluation Models. Viewpoints in Educational and Human Services Evaluation. 2. Auflage. New York: Kluwer Academic Publishers, S. 209–232.
- Rosa, H. (2016): Resonanz. Eine Soziologie der Weltbeziehung. Berlin: Suhrkamp.
- Schmitt, J./Beach, D. (2015): The contribution of process tracing to theory-based evaluations of complex aid instruments. In: Evaluation, 21. Jg., H. 4, S. 429–447.
- Scriven, M. (1991): Pros and Cons about Goal-Free Evaluation. In: Evaluation Practice, 12. Jg., H. 1, S. 55–62.
- Strübing, J./Hirschauer, S./Ayaß, R./Krähnke, U./Scheffer, T. (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. In: Zeitschrift für Soziologie, 47. Jg., H. 2, S. 83–100.
- Weiss, C. H. (1998): Evaluation. Methods for Studying Programs and Policies. 2. Auflage. Englewood Cliffs, NJ: Prentice Hall.
- Werthern, A. v. (2020): Theoriebasierte Evaluation: Entwicklung und Anwendung eines Verfahrensmodells zur Programmtheoriekonstruktion. Wiesbaden: Springer VS.
- Wilson-Grau, R. (2018): Outcome Harvesting. Principles, Steps, and Evaluation Applications. Charlotte, NC: Information Age Publishing Inc.

8 Wie bringt ein Bundesprogramm die Zivilgesellschaft „in Aktion“?

Empirische Befunde und Überlegungen zur Identifikation von Wirkung des Bundesprogramms „Demokratie leben!“ anhand einer regionalen Vergleichsstudie

Frank König, Alexander Stärck und Diana Zierold

8.1 Einleitung

Das Bundesprogramm „Demokratie leben!“ fördert eine Vielzahl vor allem zivilgesellschaftlicher Projekte, die dazu beitragen wollen, Demokratie zu stärken, gesellschaftliche und kulturelle Vielfalt anzuerkennen sowie Rassismus, Diskriminierung, Antisemitismus und politischem Extremismus vorzubeugen. Die geförderten Akteur:innen bearbeiten diese Themen mit Handlungsansätzen primär der Bildung, Beratung und Prävention. Das geschieht sowohl auf der Bundes- als auch auf der Landes- und kommunalen Ebene. Dazu sind im Programm mehrere Programmbereiche eingerichtet, in denen Gruppen von Projekten nach Themen und den staatlichen Ebenen, auf denen sie tätig sind, zusammengefasst wurden.¹

Die Wirkungen eines Programms wie „Demokratie leben!“ – im Sinne von Veränderungen, die durch die Aktivitäten der dort geförderten Projekte angestoßen werden – kann man auf verschiedenen Ebenen untersuchen. Infrage kommen zum Beispiel die Ebene der geförderten Projekte, die Ebene der einzelnen Programmbereiche oder die Ebene des Gesamtprogramms, das heißt der Gesamtheit aller Programmbereiche (Widmer 2012). Weitere Ebenen sind die der Individuen, Organisationen, Zusammenschlüsse einzelner Akteur:innen (z. B. Netzwerke) oder der der örtlichen Gemeinschaften.

Als Mitarbeiter:innen des Evaluationsteams im Bundesprogramm (nachfolgend kurz „Gesamtevaluation“ genannt), welches – anders als die wissenschaftlichen Begleitungen der einzelnen Programmbereiche – eine programmbereichsübergreifende Perspektive einnimmt, haben wir uns ausdrücklich mit der Ebe-

1 Das sind die Programm- bzw. Handlungsbereiche Modellprojekte, Kommune (Partnerschaften für Demokratie), Bund (Kompetenzzentren und -netzwerke) sowie Land (Landes-Demokratiezentren und Beratungsarbeit im Kontext von Rechtsextremismus sowie islamistischem Extremismus).

ne des Gesamtprogramms beschäftigt und Fragen behandelt, die alle Programmbereiche und ihr Zusammenwirken betreffen. Für diesen Beitrag haben wir uns zwei Fragen herausgegriffen: Zum einen: Wie trägt das Programm dazu bei, zivilgesellschaftliches Engagement anzuregen und zu stärken? Zum anderen: Welche Wirkungen erzeugt das Zusammenspiel verschiedener programmgeförderter Akteur:innen vor Ort?

Aus vorangegangenen Evaluationen wussten wir, dass wir die Frage nach der Zusammenarbeit der in „Demokratie leben!“ geförderten Akteur:innen nicht allein auf Basis von programmweiten Projektbefragungen zufriedenstellend klären konnten. Dazu bleiben deren Ergebnisse in der Regel zu abstrakt. Daher entschieden wir uns, zusätzlich Untersuchungen dort zu machen, wo sich das Handeln dieser Akteur:innen beobachten lässt: auf der Ebene von Kreisen und Gemeinden (Regionen bzw. Kommunen). Dort haben wir zusammen mit Kolleginnen und Kollegen aus dem Team der wissenschaftlichen Begleitung des Programmbereichs² Kommune mithilfe eines Regionenvergleichs Wirkungen der praktischen Umsetzung des Bundesprogramms untersucht. Uns interessierten in diesem Zusammenhang nicht nur die Erträge von Kooperation, sondern als zweite Wirkdimension auch die Anregung und Stärkung von zivilgesellschaftlichem Engagement für Demokratie und gegen politischen Extremismus sowie Gruppenbezogene Menschenfeindlichkeit. Dabei wollten wir nicht nur Wirkungen feststellen, sondern auch ermitteln, welche zentralen Faktoren deren Zustandekommen beeinflussen.³

In Bezug auf die Untersuchung von Zusammenarbeit legten wir ein besonderes Augenmerk auf das Zusammenspiel zivilgesellschaftlicher und staatlicher Akteur:innen vor Ort. Denn die Arbeit zu den Kernthemen von „Demokratie leben!“ ist zum einen eine Gemeinschaftsaufgabe von Staat und (Zivil-)Gesellschaft. Zum anderen wissen wir aus der eigenen Evaluationspraxis und der wissenschaftlichen Literatur (u. a. Zimmer 2010; Grande/Grande/Hahn 2021), dass dieses Verhältnis nicht spannungsfrei ist. Das wiederum beeinflusst sowohl das Zustandekommen und die Qualität der Zusammenarbeit der Akteur:innen als auch deren Erträge.

Mit den Regionalanalysen gingen wir in der Gesamtevaluation davon aus, dass vergleichende Fallstudien dieser Art nicht nur spezifische Wirkungen auf regionaler Ebene zutage fördern, sondern in Verbindung mit weiteren, vor allem quantitativen Daten, auch Generalisierungen mit Blick auf das Bundesprogramm ins-

2 In der zweiten Förderphase des Bundesprogramms „Demokratie leben!“ (2020 bis 2024) wurde für die Bezeichnung eines Programmbereichs der Begriff „Handlungsbereich“ verwendet. Die drei Programmzielbereiche „Demokratieförderung“, „Vielfaltgestaltung“ und „Extremismusprävention“ hießen „Handlungsfelder“.

3 Die Ausführungen in diesem Artikel zum Untersuchungsdesign und zu ausgewählten Ergebnissen unserer Regionalstudien basieren im Wesentlichen auf einem ausführlichen Bericht dazu (Autor:innengruppe DJI/ISS i. E.).

gesamt erlauben. Dadurch – so unsere Annahme – sollte es möglich sein, Wirkungen, die aus dem Zusammenspiel der im Programm geförderten und weiterer Akteur:innen auf regionaler Ebene erwachsen, als „Programmwirkungen“ zu fassen und zu plausibilisieren (Balzer/Beywl 2018, S. 97f.).

Vor diesem Hintergrund wollen wir in den nun folgenden Ausführungen der Frage nachgehen, welche Möglichkeiten und Grenzen regionale Fallstudien bieten, um in einem Bundesprogramm wie „Demokratie leben!“ empirisch belastbare Aussagen zu seiner Wirksamkeit auf der kommunalen Ebene im Bundesmaßstab treffen zu können. Wirksamkeit wird hierbei als die Fähigkeit verstanden, (intendierte) Wirkungen hervorzubringen (Ottmann/König 2023, S. 18 f.). Dazu ziehen wir beispielhaft und kontrastierend zwei Regionen aus der Studie heran.

Unser Beitrag gliedert sich wie folgt: Zuerst stellen wir die Regionalanalysen vor (Abschnitt 8.2). Dies umfasst eine Beschreibung des methodischen Vorgehens (Abschnitt 8.2.1) sowie ausgewählter wirkungsbezogener Ergebnisse aus zwei der untersuchten Regionen und deren Vergleich (Abschnitt 8.2.2). Danach wird das Vorgehen mit Blick darauf reflektiert, wie gut es geeignet war, Wirkungen des Gesamtprogramms zu plausibilisieren (Abschnitt 8.3). Der Beitrag schließt mit einem kurzen Fazit (Abschnitt 8.4).

8.2 Regionalanalysen als Ansatz zur Untersuchung von Programm-Wirkungen

8.2.1 Methodisches Vorgehen

Die Auswahl geeigneter Untersuchungsfälle (Regionen), die Entwicklung angemessener Erhebungsinstrumente und die Abstimmung mit anderen Evaluationsteams beansprucht in einem vielgestaltigen Programm wie „Demokratie leben!“ typischerweise einige Zeit. Bereits im Jahr 2020 begannen wir die Konzeption und Vorbereitung der Regionalstudien. In unserem Fall hatten wir nicht nur der Vielfalt lokaler Problemlagen, Themen und handelnden Akteur:innen im Bundesprogramm Rechnung zu tragen, sondern erfuhren auch Einschränkungen unserer Forschungs- und Evaluationsarbeit im Zuge der Corona-Pandemie (2020–2023). Vor diesem Hintergrund waren einerseits geeignete Regionen zu identifizieren. Im Ergebnis haben wir anhand der Merkmale „bearbeitete Programmt Themen vor Ort“, „sozioökonomische Rahmenbedingungen“, „geografische Lage“ sowie „Anwesenheit von programmgeförderten Akteur:innen“ neun Regionen für unsere Studie ausgewählt. Andererseits waren Zugänge zu den Regionen und den regionalen Akteur:innen zu erarbeiten. Um Doppelerhebungen zu vermeiden, geschah das in Absprache mit anderen Teams in der wissenschaftlichen Begleitung innerhalb des Evaluationsverbundes von „Demokratie leben!“.

Flankiert wurde das Ganze von Sondierungsgesprächen mit verschiedenen situationskundigen Akteur:innen aus Verwaltung und Zivilgesellschaft in den ausgewählten Regionen. Das hat uns ebenfalls dabei geholfen einzuschätzen, inwieweit die Regionen für die geplanten Untersuchungen geeignet sind. In diesen Vorgesprächen überprüften wir unsere ersten Einschätzungen hinsichtlich der für unsere Studie relevanten Themen bzw. Problemlagen und der Möglichkeit, dort Erhebungen durchzuführen. Wir konnten diese dann entweder bestätigen oder mussten – wie in zwei Fällen geschehen – vorab ausgewählte Regionen wieder verwerfen. Diese zeitaufwendigen Vorbereitungen dauerten ungefähr ein Jahr.

Die Erhebungen führten wir von Ende 2021 bis Anfang des Jahres 2023 durch. An der Untersuchung haben sowohl Mitarbeitende unterschiedlicher staatlicher und nichtstaatlicher Organisationen⁴ teilgenommen, die in den Programmbereichen des Bundesprogramms gefördert wurden, als auch Akteur:innen der Zivilgesellschaft, des Bildungsbereiches und weiterer relevanter Institutionen, die keine Programmmittel aus „Demokratie leben!“ erhielten, aber in dessen Themenbereichen arbeiten. Insgesamt haben wir in jeder der ausgewählten Regionen im Durchschnitt sieben problemzentrierte Einzel- und Gruppeninterviews geführt. Die Interviews erfolgten entweder face-to-face vor Ort oder per Videokonferenz. Grundlage bildeten kurze Interviewleitfäden, die sich um verschiedene Aspekte der Zusammenarbeit von Akteur:innen in den Regionen, die konkrete Arbeit an den relevanten Themen vor Ort, die Rolle von „Demokratie leben!“ sowie die Effekte der Programmförderung drehten. Der Leitfaden und die Interviewführung wurden dabei zum einen an die konkreten Gegebenheiten vor Ort angepasst und zum anderen daran, ob die Befragten unmittelbare Begünstigte des Bundesprogramms waren oder nicht. Das so zustande gekommene Material haben wir inhaltsanalytisch (nach Mayring 2015) sowie orientiert an der Dokumentarischen Methode (nach Bohnsack/Nentwig-Gesemann/Nohl 2013; Bohnsack 2014 sowie Nohl 2017) ausgewertet, um anhand von Schlüsselstellen geteilte Orientierungen der Befragten zu deren Verständnis bezüglich Themenbearbeitung, Kooperation Staat – Zivilgesellschaft und Einschätzung des Bundesprogramms herauszuarbeiten. In mehreren der untersuchten Regionen realisierten wir außerdem im Erhebungszeitraum (Stadt-)Raumbegehungen und Soziale Netzwerkanalysen.⁵ Die

4 Träger der Koordinierungs- und Fachstellen sowie sogenannte Federführende Ämter der Partnerschaften für Demokratie, Träger von Modellprojekten, Mitarbeitende in Landes-Demokratiezentren, Träger der Mobilien Beratung und Träger in sogenannten Kompetenznetzwerken (<https://www.demokratie-leben.de/das-programm/ueber-demokratie-leben>, Abruf 03.05.2024).

5 Ziel der Sozialen Netzwerkanalysen war es unter anderem, die Vernetzungs- und Kooperationsbeziehungen verschiedener Akteur:innen vor Ort abzubilden. Den Befragten wurde dabei die Möglichkeit geboten, auf zwei Netzwerkkarten jeweils die aus ihrer Sicht relevanten Akteur:innen zu einem Thema einzuordnen: in einer ersten, egozentrierten Netzwerkkarte entsprechend

hierbei entstandenen Netzwerkkarten wurden, wenn möglich, angelehnt an die qualitative strukturelle Analyse in die Interpretationen einbezogen (Herz/Peters/Truschkat 2015). Zusätzlich verwendeten wir Ergebnisse aus quantitativen Erhebungen der wissenschaftlichen Begleitung des Programmbereichs Kommune in allen Partnerschaften für Demokratie (Pfd)⁶ des Bundesprogramms, um die qualitativ gewonnenen Befunde einzuordnen und zu kontrastieren. Außerdem erfassen wir Daten zur politischen Situation, zur Sozial- und Engagementstruktur sowie zur wirtschaftlichen Situation in den Regionen als Kontextfaktoren der Arbeit der regionalen Akteur:innen.⁷

Um die bei der Auswertung des Interviewmaterials identifizierten Entwicklungen und Erträge des Handelns der befragten Akteur:innen als Effekte der Programmförderung und der Zusammenarbeit von Akteur:innen vor Ort auszuweisen, haben wir – evaluationstheoretisch betrachtet – in einem ersten Schritt den Ansatz des „Outcome-Harvesting“ (outcomeharvesting.net, Abruf 03.05.2024) verwendet. Es handelt sich hierbei um eine von Ricardo Wilson-Grau (2018) entwickelte Kombination von nutzungsorientierter Evaluation und Outcome-Mapping, die es erlaubt, empirisch erhobene Veränderungen bzw. Entwicklungen „ex post“ als Wirkungen zu beschreiben. Zentrale Mechanismen des Zustandekommens der so dem Programm „Demokratie leben!“ und der regionalen Kooperation der Akteur:innen zugeschriebenen Entwicklungen deckten wir auf, indem wir aus dem erhobenen Material mithilfe „sensibilisierender Konzepte“ die Art und Weise rekonstruiert haben, mit der sowohl die Zusammenarbeit von Akteur:innen als auch die Programmförderung zu den beobachteten Veränderungen beigetragen haben. Mechanismen begreifen wir hierbei als empirisch nur begrenzt direkt zu bestimmende „Kräfte“, die zwischen dem Handeln von Akteur:innen und den durch sie erzeugten Wirkungen „vermitteln“. Die hierbei verwendeten „sensibilisierenden Konzepte“ aus der Stadt-, Netzwerk- und Governance-Forschung können als implizite, kleinteilige „Theories of Change“ begriffen werden, mit deren Hilfe empirische Befunde als Mechanismen beschreibbar und Handeln sowie berichtete Effekte einander zurechenbar gemacht werden.

der Intensität der Zusammenarbeit (je näher die Akteur:innen in der Mitte liegen, desto intensiver sind die Beziehungen) und in einer zweiten Netzwerkkarte nach der Bedeutung für die Themenbearbeitung in der Region. Im Abschnitt 8.2.2 sind zwei dieser Netzwerkkarten beispielhaft abgebildet.

6 Pfd sind als ein Programm- bzw. Handlungsbereich von „Demokratie leben!“ auf kommunaler Ebene angesiedelt und sollen im Zusammenwirken von Zivilgesellschaft, Verwaltung und Politik lokale Strategien zur Demokratieförderung und Vielfaltgestaltung entwickeln und deren Umsetzung unterstützen (Näheres dazu siehe: <https://www.demokratie-leben.de/das-programm/ueber-demokratie-leben/partnerschaften-fuer-demokratie>, Abruf 03.05.2024).

7 Hierbei orientierten wir uns in den Regionen am Zeitpunkt der Auswahl bzw. dem Start in die Untersuchungen, das heißt dem Jahr 2021.

Die Kombination von Outcome-Harvesting und der Rekonstruktion von Mechanismen war nötig, da wir mithilfe der Regionalstudien sowohl intendierte als auch nicht intendierte Erträge feststellen wollten und darauf verzichtet haben, ex ante dafür eine gegenstands- oder ebenenbezogene Programmtheorie zu explizieren. In den folgenden Kapiteln werden ausgewählte Ergebnisse der Untersuchungen für die Fälle Annegretstadt und Bettinastadt dargestellt.

8.2.2 Beispiele aus den Regionalanalysen⁸

8.2.2.1 Porträt Annegretstadt⁹

Annegretstadt ist eine Großstadt mit einer heterogenen sozialen Lage. Sie ist davon geprägt, dass sowohl die Arbeitslosenquote (Stand 2023) als auch der Anteil von Menschen, die Leistungen zur Sicherung des Lebensunterhaltes bezogen (Stand 2021), doppelt so hoch lagen wie der Bundesdurchschnitt und auch höher als im Landesdurchschnitt. Zudem haben knapp 40 Prozent der dort lebenden Menschen einen Migrationshintergrund (Stand 2022).¹⁰ Die dominierenden Parteien sind SPD, CDU und Bündnis 90/Die Grünen. Bei den Kommunalwahlen konnten extrem rechte Parteien über mehrere Wahlperioden einzelne Sitze im Parlament erringen.

In den Interviews wird hinsichtlich der Kernthemen von „Demokratie leben!“ Rechtsextremismus als das in der Region vordringlich zu bearbeitende Thema benannt. Viele Jahre war es zunächst vor allem die ehrenamtliche Zivilgesellschaft, die sich den zunehmenden Aktivitäten der rechtsextremen Szene entgegenstellte und auf das Problem hinwies, bevor städtische Akteur:innen aus Verwaltung und Politik sich des Themas annahmen und Maßnahmen dagegen ergriffen.

Durch einen seit dem Jahr 2007 bestehenden und kontinuierlich fortgeschriebenen kommunalen Aktionsplan sowie einen Beschluss des Kommunalparlaments stehen inzwischen auch kommunale Gelder für Projekte zur Stärkung von Demokratie und zur Prävention von vor allem Rechtsextremismus zur Verfügung. Seit mehr als zehn Jahren erhält die Stadt zudem eine Förderung durch die einschlägigen Bundesprogramme¹¹ für ihre „Partnerschaften für Demokratie“

8 Für diesen Artikel wurden exemplarisch zwei Regionen ausgewählt. Die Ergebnisse über alle Regionen werden demnächst auf der Projekthomepage der Programmevaluation „Demokratie leben!“ des DJI veröffentlicht.

9 Die Städtenamen wurden anonymisiert.

10 Die Daten sind den Statistikportalen von Annegretstadt und der Agentur für Arbeit entnommen (Zeitpunkt: 01/2024).

11 Im Bundesprogramm „Toleranz fördern – Kompetenz stärken“ (2011–2014) wurden entsprechende Aktivitäten als „Lokale Aktionspläne“ gefördert, in der ersten (2015–2019) und zweiten Förderperiode (2020–2024) von „Demokratie leben!“ erfolgt deren Förderung im Rahmen von „Partnerschaften für Demokratie“.

(kurz: PfD). Daraus finanziert die Kommune eine Koordinierungs- und Fachstelle (KuF¹²) für die PfD sowie verschiedene Einzelprojekte. Dabei rückten zuletzt auch die Themen Vielfaltgestaltung und Demokratieförderung stärker in den Blick.

Daneben existiert eine städtische Koordinierungsstelle mit drei Mitarbeiter:innen zur Umsetzung des bereits genannten Aktionsplans, die direkt beim Oberbürgermeister angesiedelt ist. Außerdem ist ein ehrenamtlicher Sonderbeauftragter des Oberbürgermeisters für die Themen bestellt. Personen des öffentlichen Lebens, unter anderem aus der Polizei, Justiz, Wirtschaft, positionieren sich für alle sichtbar zu den genannten Themen. Sie unterstützen Politik und Verwaltung in der strategischen Arbeit dazu und bei entsprechenden Maßnahmen.

Eine Vielzahl von Arbeitskreisen, Initiativen, Bündnissen, Vereinen sowie Fach- und Netzwerkstellen prägen die zivilgesellschaftliche Landschaft in Annegretstadt in den genannten Themenfeldern. Das Spektrum bewegt sich dabei von rein ehrenamtlichen Zusammenschlüssen, Initiativen verschiedener Organisationen, hauptamtlicher Fachkräfte bis hin zu Gremien von Entscheidungsträger:innen. Es gibt mehrere Bündnisse und Runde Tische, die oftmals stadtteilbezogen agieren, aber auch stadtweit aktivierbar sind. In Annegretstadt gibt es zahlreiche Kooperationen zwischen Akteur:innen des Bundesprogramms und Projekten sowie Initiativen, die nicht über „Demokratie leben!“ gefördert werden. Sie finden in der Regel themenbezogen statt, verlaufen aber nicht immer reibungslos. So berichteten verschiedene Gesprächspartner:innen vor Ort von Parallelstrukturen sowie Unstimmigkeiten zwischen den einzelnen Akteur:innen oder von Konkurrenzen. Kontroversen kristallisieren sich beispielweise entlang der Frage heraus, welche Akteur:innen der Zivilgesellschaft sich in welcher Form an der Bearbeitung von Rechtsextremismus beteiligen können oder sollen.

Ehrenamtlich Aktive in Annegretstadt berichteten in den Interviews auch über eine von ihnen wahrgenommene Dominanz der städtischen Akteur:innen, insbesondere der Verwaltung. Diese äußert sich zum einen im kommunalen Agenda-Setting im Rahmen des städtischen Aktionsplans. Themensetzungen finden hier oftmals ohne nennenswerte Beteiligung der Zivilgesellschaft statt. Zum anderen bemängeln die befragten Interviewpartner:innen, dass die von der KuF und der städtischen Koordinierungsstelle angestoßenen Aktivitäten oft mit strikten Vorgaben zur Art und Weise der Themenbearbeitung auch für die adressierten zivilgesellschaftlichen Akteur:innen einhergehen. Dadurch entsteht bei den Ehrenamtlichen der Eindruck, dass es den staatlichen Akteur:innen nicht um partner-

12 Die KuF ist zwar de jure bei einem freien Träger angesiedelt, hat ihr Büro aber in der Verwaltung beim Federführenden Amt bzw. bei der städtischen Koordinierungsstelle. Sie wird in der Bevölkerung daher der Verwaltung und nicht der Zivilgesellschaft zugehörig betrachtet.

schaftliche Zusammenarbeit geht, sondern darum, das zivilgesellschaftliche Engagement verwaltungskonform zu managen.

Diesen Eindruck bestätigen die Interviews mit den Mitarbeiter:innen der städtischen Koordinierungsstelle und der KuF der PfD teilweise. Diese gaben zum Beispiel an, dass sie eine ihrer Aufgaben darin sehen, Einfluss auf die Zusammensetzung neu gegründeter Gremien und Akteursarrangements unter zivilgesellschaftlicher Beteiligung zu nehmen. Entsprechende Aussagen lassen sich dabei als Anzeiger dafür interpretieren, dass die KuF und die städtische Koordinierungsstelle versuchen, sich in ihrer Koordinierungsarbeit im Verhältnis zur Zivilgesellschaft im Wesentlichen einer Top-down-Steuerung zu bedienen. Dabei legen sie Themen und Zielrichtung der geplanten Aktivitäten eigenmächtig fest, bevor sie die breite Zivilgesellschaft einbeziehen. Im Ergebnis zeigt sich hier ein etatistisches Verwaltungs- bzw. Rollenverständnis, das in der Tendenz mit einem distanzierten Verhalten gegenüber der Zivilgesellschaft einhergeht und dabei Rückhalt sowohl im Stadtrat als auch in der Kommunalverwaltung zu besitzen scheint.

Letztlich schränkt dieser Koordinierungsansatz aus Sicht einiger Befragter aus dem zivilgesellschaftlichen Bereich ihre Handlungsfähigkeit nicht nur ein, sondern führt aufseiten der ehrenamtlichen Akteur:innen zu Demotivation. Vereinzelt trägt das dazu bei, dass sie ihr Engagement in den Themenfeldern des Bundesprogramms einschränken.

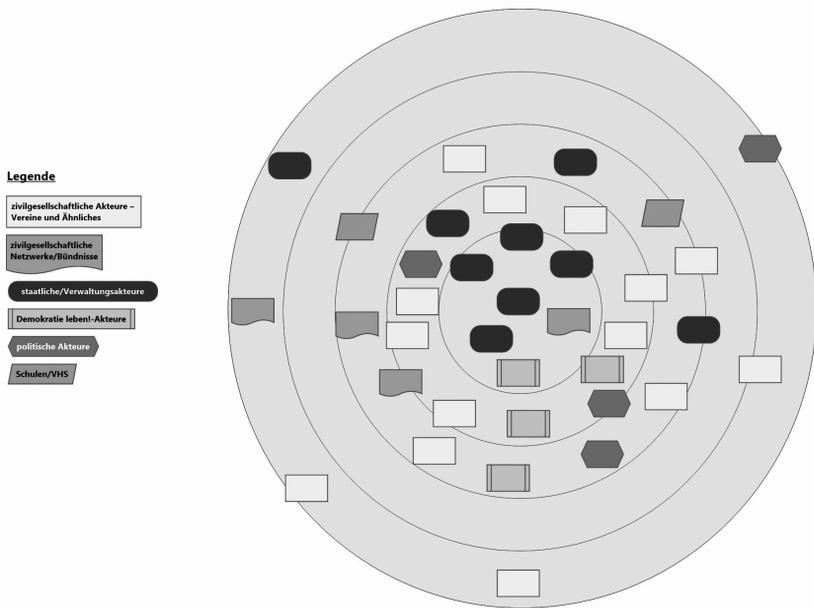
Die zentrale Position der Verwaltungsakteur:innen in der Themenbearbeitung zeigt sich auch in den für die Region Annegretstadt erstellten Netzwerkkarten deutlich (siehe Abb. 1). Ebenso gut lässt sich dort die Vielzahl an zivilgesellschaftlichen Akteur:innen und Netzwerken erkennen, die sich in Annegretstadt engagieren.¹³ Während manche schon seit vielen Jahren bestehen, haben sich andere mit der und durch die Förderung über das Bundesprogramm entwickelt.

In Annegretstadt zeigt sich, dass das Bundesprogramm regionales zivilgesellschaftliches Engagement aus den Netzwerken heraus motiviert und unterstützt, indem es den PfD Fördermittel in begrenztem Umfang für verschiedene Projekte in der Zivilgesellschaft bereitstellt, die unter anderem über das regionale Begleitemium der PfD, den Begleitausschuss, vergeben werden. Die inhaltliche Dominanz der städtischen Akteur:innen, gepaart mit einem starken Steuerungsan-

13 Interviewaussagen zur anfänglich stark verzögerten Reaktion der städtischen Akteur:innen (sowohl der Verwaltung/Verwaltungsspitze als auch der Kommunalpolitik) auf die Problemlagen in der Stadt legen nahe, dass sich die hier abgebildete Zentralität dieser Akteur:innen erst in den letzten Jahren entwickelt hat. In einer ersten Vorstellung von Ergebnissen im Begleitausschuss der PfD bestätigten die dabei Anwesenden diesen Eindruck.

14 Für die Darstellung wurde eine Karte ausgewählt, die zusammenfassend für alle vor Ort angefertigten Netzwerkbilder steht. Dies trifft auch auf die Abbildung 2 für Bettinastadt zu. Beide Karten wurden von zivilgesellschaftlichen, nicht über das Bundesprogramm geförderten Akteur:innen erstellt.

Abbildung 1: Karte der wichtigsten Akteur:innen in Annegretstadt im Bereich der benannten zentralen Themen/Problemlagen¹⁴



Quelle: Eigene Darstellung der Gesamtevaluation

spruch, führt in diesem Zusammenhang unserer Einschätzung nach zwar dazu, dass Zivilgesellschaft für ein Engagement in den Themenbereichen des Bundesprogramms gewonnen wird. Im Vergleich zu anderen Regionen unserer Studie, wie der nachfolgend vorgestellten Region Bettinastadt, gelingt dies jedoch in einem weitaus geringeren Maße und weniger zielgerichtet.

Dessen ungeachtet lässt sich am Beispiel Annegretstadt gut zeigen, dass es gerade für (Rechts-)Extremismusprävention in Verbindung mit Demokratieförderung von großem Vorteil ist, wenn sich das Stadtoberhaupt, die Mehrheit des Stadtrates und weitere wichtige Repräsentant:innen, beispielsweise Leitungspersonen aus den Bereichen Sport und Sicherheit, als Unterstützer:innen dieser Arbeit positionieren. Eine eigene, programmunabhängige finanzielle Förderung für Projekte sowie für Personal, das sich aus der Verwaltung heraus diesen Themen widmet, kann die öffentliche Sensibilisierung und Debatte fördern und weitere Akteur:innen animieren, sich für diese Themen einzusetzen. Sichtbar wird, dass in Annegretstadt im Zusammenspiel von städtischen Aktivitäten und solchen aus dem Bundesprogramm heraus – bei aller Kritik an der teilweise hierarchisch geprägten Zusammenarbeit – eine tragfähige Basis für die Bearbeitung von Rechtsextremismus und die Stärkung demokratischer Widerstandsfähigkeit

existiert. Dieses Zusammenspiel von administrativer Steuerung, Commitment der Kommunalpolitik sowie die breite Vernetzung in Verwaltung und Zivilgesellschaft lassen sich hier als Elemente eines Wirkmechanismus identifizieren, über den demokratische Resilienz in der Stadtgesellschaft gestärkt wird. Das Bundesprogramm leistet hierbei einen Beitrag.¹⁵ Eine stärkere partizipative Einbindung in die Planungen und Umsetzungen von Aktivitäten zu diesen Themen und die Mitnahme insbesondere der ehrenamtlich (bereits) engagierten Zivilgesellschaft, kann diese Basis noch weiter stärken.

Durch die Vielzahl an Netzwerken, Gremien und Ähnlichem erhöht sich auch für interessierte Bewohner:innen die Chance, ein passendes „Angebot“ für ein mögliches Engagement oder Ansprechpartner:innen bei Fragen, Ideen oder Problemlagen zu finden. Darüber hinaus schafft ein breites Angebot die Möglichkeit, flexibel, bedarfsgerecht und vor allem über ehrenamtliche und andere zivilgesellschaftliche Strukturen niedrigschwellig auf (neue) Situationen und Herausforderungen reagieren zu können.

Wir konnten am Beispiel Annegretstadt zum einen sehen, dass eine breite Unterstützung durch die kommunalen Entscheidungsgremien, gepaart mit eigenen städtischen finanziellen und personellen Ressourcen im Bereich der Demokratieförderung und Rechtsextremismusprävention, sich sowohl positiv auf das Verständnis und die Problemwahrnehmung in der Bevölkerung auswirkt als auch den Fokus der Akteur:innen in den Programmbereichen von „Demokratie leben!“ lenken kann. Zugleich lässt sich in der Region belegen, dass sich administrative Steuerung von Engagement, wenn sie hierarchisch angelegt ist, auf das Anliegen der „Aktivierung von Zivilgesellschaft“ nachteilig auswirkt.

8.2.2.2 Porträt Bettinastadt

Auch Bettinastadt ist eine Großstadt, in der sich eine heterogene Sozialstruktur mit einer über die Stadtteile ungleichen Verteilung verschiedener Bevölkerungsschichten (arm oder reich, mit oder ohne Migrationshintergrund) zeigt. Der Anteil an Einwohner:innen mit Migrationshintergrund betrug Ende 2022 beinahe 50 Prozent. Die Arbeitslosenquote lag mit rund sieben Prozent über dem Durchschnitt des Bundeslandes (Stand 2021). Die stärksten Fraktionen des Gemeinderates stellen Bündnis 90/Die Grünen, SPD und CDU. Die AfD erreichte in der letzten Kommunalwahl knapp unter zehn Prozent.¹⁶

Analog zum hohen Anteil von Einwohner:innen mit Migrationshintergrund ist aus dem Kanon der Programmenthemen von „Demokratie leben!“ die Vielfaltgestaltung in der Stadt besonders zentral. Mit Blick auf Demokratieförderung

15 Dieser kann im Verhältnis zum Eintrag der regionalen Inputs (Ressourcen) und Maßnahmen gegenwärtig aufgrund fehlender Daten jedoch nicht weiter qualifiziert werden.

16 Alle statistischen Angaben sind der Homepage der Stadt entnommen (Zeitpunkt: 01/2024).

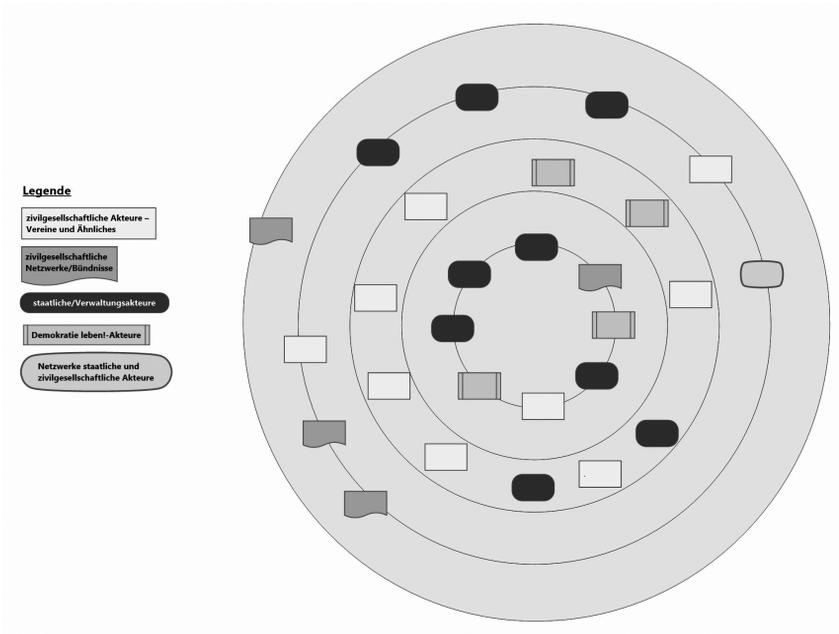
und die Prävention von Rechts- und islamistischem Extremismus haben die in der Region interviewten Personen außerdem Gruppenbezogene Menschenfeindlichkeit, Islamophobie, Rassismus und Antisemitismus als Arbeitsgegenstände benannt.

In Bettinastadt nehmen staatliche und zivilgesellschaftliche Akteur:innen schon seit vielen Jahren Förderungen über einschlägige Bundesprogramme in Anspruch. So gab es bereits im Jahr 2009 eine Modellprojektförderung aus dem Xenos-Programm „Integration und Vielfalt“. Seit dem Programm „Toleranz fördern – Kompetenz stärken“ (2011–2014) wird die lokale Demokratiarbeit der Stadt im Rahmen eines „Lokalen Aktionsplans“ bzw. einer PfD durch den Bund unterstützt. Die Stadt verfügt über eigene Organisationsstrukturen, die sich mit den Themen Demokratie und Vielfalt beschäftigen. Zu nennen ist vor allem ein entsprechender Fachbereich in der Stadtverwaltung. Dort ist neben der Koordinierungs- und Fachstelle der PfD auch die Koordinierungsstelle eines Bündnisses für Vielfalt angesiedelt, teilweise in Personalunion. Die dem Bündnis zugrundeliegende „Bettinastädter Erklärung für Vielfalt“ haben bereits über 350 Institutionen, Vereine und Organisationen unterschrieben. Sie geht auf eine Initiative der Stadt zurück. Für das zivilgesellschaftliche Engagement gegen Rechtsextremismus, Muslimfeindlichkeit, Antisemitismus und Antiziganismus stehen neben den erwähnten Bundesmitteln auch Gelder in sechsstelliger Höhe aus dem städtischen Haushalt zur Verfügung.

Bettinastadt verfügt ähnlich wie Annegretstadt über eine Vielzahl an zivilgesellschaftlichen Akteur:innen, Initiativen und Vereinen sowie Netzwerken, die neben den Verwaltungs- und anderen staatlichen Akteur:innen für die Förderung der Akzeptanz und Bewältigung von gesellschaftlicher Vielfalt als wichtig angesehen werden. Im Unterschied zu Annegretstadt finden sich in Bettinastadt jedoch keine Akteur:innen, die als dominierend oder tonangebend betrachtet werden. Das zeigt sich auch in den von den Interviewten erstellten Netzwerkkarten, wovon eine, die die Informationen aller Karten gut zusammenfasst, als Abbildung 2 eingefügt ist. Dort ist kein:e Akteur:in ins unmittelbare Zentrum der erstellten Karten gerückt worden. Dies begründeten die Befragten unter anderem damit, dass sie eine Vielzahl an Akteur:innen als wichtig erachten und diese auf Augenhöhe agierten. Dementsprechend sind die einzelnen Akteur:innen auch in konzentrischen Kreisen angeordnet.

Die Zusammenarbeit der Akteur:innen vor Ort in den genannten Themen gründet sich auf historisch gewachsene Strukturen, personelle Überschneidungen in zivilgesellschaftlichen Organisationen (vor allem durch ehrenamtliche Mehrfachmitgliedschaften) sowie Beziehungen, die durch die Förderung aus dem Bundesprogramm angeregt wurden. Hervorzuheben sind dabei sogenannte „Themenpartnerschaften“. Es handelt sich hierbei um eine Form von „außen“ veranlasster Zusammenarbeit zwischen bisher nicht kooperierenden Akteur:innen. Die Anregung passiert, da die Bewilligung von Fördermitteln aus „Demokratie

Abbildung 2: Karte der wichtigsten Akteur:innen in Bettinastadt im Bereich der benannten zentralen Themen/Problemlagen (Gesamtnetzwerk)



Quelle: Eigene Darstellung der Gesamtevaluation

leben!“ daran gebunden ist, dass die Geförderten neue Kooperationsbeziehungen miteinander eingehen. Daraus erwachsen zum einen neue, gegebenenfalls nachhaltige (dauerhafte) Arbeitsbündnisse. Zum anderen entstehen neue Ideen, um die Arbeit im Bereich der Vielfaltgestaltung und Demokratieförderung in der Stadt voranzubringen, die nach Einschätzungen der Interviewpartner:innen ohne diesen Anstoß nur sehr unwahrscheinlich entstanden wären. Zahlreiche Ergebnisse und Handlungsempfehlungen aus der Arbeit der „Themenpartnerschaften“ finden – so die Aussagen einzelner Befragter – in verschiedenen Gremien des Stadtrates im Kontext kommunalpolitischer Entscheidungen einen Widerhall.

Wie in den Interviews zu erfahren war, nutzten einige Akteur:innen die Förderung der Zusammenarbeit auch, um Konzepte für eigene Folgeprojekte zu erarbeiten. Hier leistet das Bundesprogramm eine Art „Anschubfinanzierung“. Die dabei entstehenden Konzepte werden dann anderen Geldgeber:innen zur Förderung vorgeschlagen, wobei die vormalige Förderung der Projektträger aus „Demokratie leben!“ mitunter als eine Art „Gütesiegel“ fungiert.

Weitere Impulse erhält die lokale Arbeit an den Programmtiteln aus der engen Zusammenarbeit zwischen einem in der Stadt tätigen Modellprojekt

von „Demokratie leben!“ und den PfD-Mitarbeiter:innen.¹⁷ Das Handeln dieser beiden Programmakteure wirkt aus Sicht der von uns Befragten stark in die Zivilgesellschaft hinein. Dabei betrachten Akteur:innen der Zivilgesellschaft, Verwaltung, Politik und Wirtschaft die (städtischen) Koordinierenden der PfD und des Bündnisses als wichtige, aber – anders als in Annegretstadt – nicht als bestimmende Akteur:innen. Indem die Koordinierenden eine vermittelnde Position zwischen Engagierten einnehmen, die Akteur:innen der Stadtgesellschaft beraten, öffentlichkeitswirksame Maßnahmen organisieren und die Förderung der „Themenpartnerschaften“ vorbereiten, stärken sie die Zusammenarbeit in der Stadt.

Als begünstigender Faktor wirkt sich hierbei aus, dass die KuF und die Bündnis-Koordinierungsstelle in der Stadt in der Zuständigkeit eines Bereiches liegen. Als Kontextfaktor für ihr Handeln ist ebenfalls bedeutsam, dass sie enge Kontakte in die sogenannten regelfinanzierten Einrichtungen und Angebote unter anderem der Jugend- und Schulsozialarbeit, in Schulen und in Verbände wie den Stadtjugendring hinein besitzen. Damit erreichen sie Akteure, die eine Multiplikationsfunktion in der Demokratie- und Vielfaltarbeit übernehmen können. Das aus „Demokratie leben!“ im Bereich „religiös begründeter Extremismus“ geförderte Modellprojekt verfügt darüber hinaus über spezielle Zugänge zu muslimisch geprägten Communities in Bettinastadt, die hier für die Arbeit in der Gestaltung von Vielfalt unabdingbar sind.

Für viele Aktive aus der Zivilgesellschaft, aber auch für Institutionen wie Schulen, sind die KuF und die städtische Bündnis-Koordinierungsstelle unserer Erkenntnisse nach wichtige Anlaufstellen, um einerseits Kontakt zu anderen Aktiven herzustellen und andererseits auch, um handlungsfähig zu bleiben oder zu werden. Die Aktionsfähigkeit wird dadurch hergestellt, dass die Koordinierenden (Fach-)Informationen weitergeben oder auf weitergehende Angebote hinweisen. Damit leisten sie im Sinne von Wirkungen einen nachvollziehbaren Beitrag, um zivilgesellschaftliches Engagement anzuregen und zu unterstützen.

Aus den Erzählungen der Interviewpartner:innen geht hervor, dass neben der Handlungskoordination durch KuF und Bündnis-Koordination auch die Mitgliedschaft in den verschiedenen Themennetzwerken selbst für die Mobilisierung von zivilgesellschaftlichem Engagement bedeutsam ist. Der Begleitausschuss der PfD fungiert als ein solches Netzwerk. Dessen Mitglieder tauschen untereinander engagementrelevante Informationen aus ihren Institutionen, Vereinen, Bündnissen und Ähnlichen aus und besprechen Bedarfe, die die in den Themen des Bundesprogramms engagierten Akteur:innen haben, um darauf durch gezielte materielle Unterstützungsangebote oder inhaltliche Anregungen reagieren zu können.

17 Koordinierungs- und Fachstelle sowie Federführendes Amt.

In Bettinastadt lässt sich zeigen, dass die Zivilgesellschaft angemessen in die Bearbeitung gesellschaftlicher Probleme eingebunden werden kann, wenn die Kommune über das Bundesprogramm hinaus Ressourcen dafür bereitstellt und die Förderung des zivilgesellschaftlichen Engagements partizipativ angelegt ist.

8.2.2.3 Vergleich der Regionen

Bettinastadt und Annegretstadt sind beide Großstädte mit einer überdurchschnittlichen hohen Arbeitslosenquote, einer heterogenen Sozialstruktur und einer Vielzahl an zivilgesellschaftlichen Akteur:innen inner- und außerhalb des Bundesprogramms. Entsprechend der jeweiligen Problemlagen vor Ort legt Bettinastadt einen Schwerpunkt auf die Förderung von Diversität und Annegretstadt auf die Prävention von Rechtsextremismus.

In Bettinastadt wird im Vergleich zu Annegretstadt, bezogen auf die Frage nach den treibenden Kräften bei der Entstehung von Wirkungen in der Dimension „Aktivierung von Zivilgesellschaft“, aber auch bei der Neubildung von Kooperationsbeziehungen, deutlich, dass die städtischen Koordinierenden eine „weiche Steuerung“ (Göhler 2010) bzw. Governance durch Konditionierung, Koordinierung und „Überzeugung“ über den Weg der Vermittlung von bestimmtem (vorausgewähltem) Wissen und Know-how anwenden. Das kann als ein wichtiger Mechanismus betrachtet werden, über den aus dem Handeln der Koordinierenden Wirkungen entstehen. Dabei treten im Fall Bettinastadt als Akteursmerkmale eine der Zivilgesellschaft zugewandte Haltung und ein partizipationsorientiertes, zeitgemäßes Verwaltungs- bzw. Rollenverständnis hervor. Für das Handeln der Akteur:innen in den Themennetzwerken wie dem Begleitausschuss ist Handlungskoordination auf gleicher Augenhöhe und Ressourcentausch die treibende Kraft zur Engagementförderung. Im Gegensatz dazu versuchen in Annegretstadt die kommunalen Akteur:innen in der Verwaltung das zivilgesellschaftliche Engagement auf eher hierarchischem Wege an den von der Verwaltung formulierten Zielen auszurichten. Das führt teilweise zu Frustration und Demotivation bei den ehrenamtlichen Akteur:innen und zu Spannungen zwischen den Akteur:innen.

Ergänzend dazu konnten wir im Vergleich beider Regionen sehen, dass Wirkungen in der Dimension „Aktivierung von Zivilgesellschaft“ nicht nur von den Akteurskonstellationen vor Ort, sondern auch von den jeweils dort bearbeiteten Themen, beeinflusst werden. Während in Bettinastadt das Thema „Gestaltung gesellschaftlicher Vielfalt“ als originäre Aufgabe kommunaler Integrationsarbeit (Bogumil/Holtkamp 2023, S. 263 ff.) bereits seinem Wesen nach dazu einlädt, dass sich viele zivilgesellschaftliche Akteur:innen einbringen (können), beinhaltet das Thema Rechtsextremismusprävention – wie in Annegretstadt – immer auch sicherheitsrelevante Aspekte. Das ruft auch eine Reihe staatlicher Akteur:innen auf den Plan, die – im Gegensatz zu zivilgesellschaftlichen Handlungslogiken – primär sicherheitsbehördliche Handlungslogiken verfolgen.

Insoweit zeigt sich in Bettinastadt, wie eine unterstützende und vermittelnde Verwaltung die Beteiligung der Zivilgesellschaft im Bereich der Vielfaltgestaltung fördern kann, während in Annegretstadt ein eher hierarchischer, weniger partizipativ ausgerichteter Verwaltungsstil in der Zusammenarbeit mit der Zivilgesellschaft bei der Bekämpfung und Prävention von Rechtsextremismus dazu beiträgt, die Motivation und Kapazität zivilgesellschaftlicher Akteur:innen zu beeinträchtigen.

8.3 Diskussion

8.3.1 Methodologische Reflexion

Unser methodisches Vorgehen in der Untersuchung der einzelnen Regionen war vielgestaltig, sowohl was die Erhebungs- als auch die Auswertungsverfahren betrifft. Dass wir eine methodisch vielfältige Studie durchführen konnten, lag zum einen an der interdisziplinären Zusammensetzung des Teams der Gesamtevaluation und zum anderen an der Kooperation mit der für die kommunale Ebene zuständigen wissenschaftlichen Begleitung.

Spezifische Evaluationsfragestellungen, die sich aus der Betrachtung von „Demokratie leben!“ als Gesamtprogramm ergaben, und unser Vorwissen zum Bundesprogramm prägten die Erhebung und lenkten unseren Blick. Das ermöglichte uns einerseits unsere Untersuchungen effizient umzusetzen, andererseits hatte es aber auch zur Folge, dass wir Phänomene, die erst in extensiver Interpretation oder Reflexion als solche erkennbar werden, nicht in die Analyse einbeziehen konnten, wie das etwa bei einem ethnografischen Vorgehen mit langem Feldaufenthalt möglich gewesen wäre, der eine ganzheitlichere Erfassung erlaubt.¹⁸ Vor allem die Fokussierung auf ein vorab definiertes Set an spezifischen Evaluationsfragen schränkte unseren Blick ein.

Unser Vorgehen erlaubte uns dennoch, im Rahmen der uns zur Verfügung stehenden zeitlichen, finanziellen und personellen Ressourcen Einsichten in die Art und Weise von regional verdichtetem Handeln und Zusammenwirken im Kontext des Bundesprogramms sowie in dessen Erträge zu gewinnen, wie es im Rahmen der Evaluation von „Demokratie leben!“ bisher selten möglich gewesen ist.

18 Dies betrifft generell genaue Beschreibungen dessen, was die Akteur:innen in einer Region unter Containerbegriffen wie Demokratieförderung fassen und inwieweit sprachlich explizierte Handlungsziele dem tatsächlichen Handeln vor Ort entsprechen, also ob zum Beispiel als Zielvorstellung genannte Formate der Partizipation auch umgesetzt werden.

8.3.2 Erkenntnispotenziale und Grenzen der Regionalanalysen

Die von uns durchgeführten Regionalstudien haben unseres Erachtens zunächst gezeigt, dass die Plausibilisierung von Wirkungen (Balzer/Beywl 2018, S. 97 f.) auf Basis solcher Fallstudien überhaupt möglich ist. So konnten wir am Fall Bettinastadt nachzeichnen, dass und wie das Bundesprogramm dazu beiträgt, Kooperationsbeziehungen zu etablieren. Dort zeigte sich, dass die gezielte Lenkung über Fördermittel („Themenpartnerschaften“) wirkungsvoll ist und damit als Mechanismus betrachtet werden kann, um Akteur:innen zu veranlassen, Kooperationen mit neuen Partner:innen zu beginnen. Im Rahmen der Evaluation eines Programms, welches unter anderem Kooperationen anregen möchte, ist das ein relevanter Befund.

Belastbare Aussagen dazu, welche Wirkungen solche neuen, aber auch andere Kooperationsarrangements erzeugen und wie solche Wirkungen zustande kommen, ließen sich dagegen auf Basis unseres Untersuchungsdesigns nicht ohne Weiteres treffen. In den geführten Interviews erhielten wir zwar Hinweise, dass es die Akteur:innen beispielsweise beflügelt, wenn deren Arbeitsergebnisse von Schlüsselakteur:innen der Stadt wertgeschätzt und aufgenommen werden. Es hätte hier jedoch weiterer Untersuchungen bei den jeweiligen Kooperationspartner:innen und deren Adressat:innen sowie anderen Stakeholdern bedurft, um aus solchen ersten Hinweisen weitere Wirkungspfade abzuleiten.

Für die Wirkdimension „Aktivierung von Zivilgesellschaft“, einem der Kernanliegen von „Demokratie leben!“, ist es uns wiederum gelungen, einen von mehreren Wirkungspfaden und den dahinterliegenden Mechanismus zu rekonstruieren. Dieser verbindet das teilweise programmgeförderte, koordinierende Handeln von Akteur:innen in der kommunalen Verwaltung mit dem Grad an Engagement zivilgesellschaftlicher Akteur:innen. Hierbei hat sich vor allem der Regionenvergleich als gewinnbringend erwiesen. Dadurch konnten wir den „Modus der Steuerung bzw. Handlungskoordination“ im Verhältnis von Exekutive und der Zivilgesellschaft als eine der treibenden Kräfte identifizieren, die beeinflussen, inwieweit sich Menschen ehrenamtlich für die Themen des Bundesprogramms engagieren.

Zugleich ist in der Betrachtung von Annegretstadt deutlich geworden, dass Wirkungen zwar hinsichtlich einzelner Dimensionen untersucht werden können, sie jedoch nicht eindimensional sind. So ruft die Programmförderung aus „Demokratie leben!“ auch ambivalente Effekte hervor: In unserem Fall unterstützten kommunale Entscheidungsgremien zwar in einem positiven Sinne die Bearbeitung relevanter Themen bzw. Probleme vor Ort, wie die Prävention von Rechtsextremismus. Der Modus, in dem sie das machten – teilweise durch hierarchische Steuerung – hat aber gleichzeitig dazu beigetragen, die Aktivierung (bzw. das Aktiv-Halten) von Zivilgesellschaft zu beeinträchtigen. Dieses Zusammenspiel wird nicht zuletzt durch die Haltung der Akteur:innen zueinander bzw. ihr Rollenver-

ständnis sowie durch die Verfügbarkeit eigener materieller Ressourcen in der Region moderiert. Das heißt, der identifizierte Mechanismus kann zwar die Wirkung ex post als solche plausibilisieren bzw. erklären, allerdings gibt er per se noch keine Auskunft über die Ausprägung der Wirkung.

8.4 Fazit

Zusammenfassend lässt sich festhalten, dass es möglich ist, auf der Basis von Fallstudien auf der kommunalen Ebene Wirkungen in ausgewählten Wirkdimensionen eines Bundesprogramms – hier „Aktivierung von Zivilgesellschaft“ – nicht nur zu identifizieren, sondern auch hinreichend zu plausibilisieren. Darauf aufbauend lassen sich außerdem Aussagen zur Wirksamkeit des betreffenden Bundesprogramms insgesamt treffen. Unter methodischen Gesichtspunkten war dabei sowohl das Mapping relevanter Akteur:innenkonstellationen in den untersuchten Regionen ex-ante fruchtbringend als auch die Befragung nicht nur programmgeförderter sondern auch nicht geförderter Akteur:innen auf der kommunalen und auf der Landesebene. Dadurch konnten sowohl die Kooperationsbeziehungen auf der kommunalen Ebene als auch die für das Handeln der regionalen Akteur:innen bedeutsamen Arbeitsbereiche von „Demokratie leben!“ auf der Landesebene (v. a. Landeskoordinierungsstellen und Beratungsangebote) in der Analyse berücksichtigt werden. Darüber hinaus haben sich der Vergleich der Regionen miteinander ebenso bewährt wie die Netzwerkanalysen und das an der dokumentarischen Methode orientierte, sequenzanalytische Vorgehen in der Interviewsauswertung.

Im Rückblick zeigt sich, dass es umfangreiche Ressourcen braucht, um solche wirkungsorientierten Analysen in einem komplexen Programm wie „Demokratie leben!“ auf eine methodisch vor allem qualitative, „tiefergehende“ Art und Weise durchzuführen. Dennoch halten wir regionale Fallstudien für eine adäquate Möglichkeit, um Wirkungen in einem Programm wie „Demokratie leben!“ zu untersuchen und zu plausibilisieren. Aufgrund des damit verbundenen Aufwandes sollten sich Wirkungsuntersuchungen dieser Art auf zentrale Wirkungsdimensionen und deren Mechanismen fokussieren. Um mit elaborierten Ansätzen der Wirkungsplausibilisierung und -erklärung wie der Kontributionsanalyse (Starke 2022) zu arbeiten, ist es außerdem essenziell, vorab eine Programmtheorie zu formulieren, die den sich anschließenden Untersuchungen als Ausgangspunkt dient.

Zugleich kann es sich lohnen, die Erkenntnisse zu den Wirkungszusammenhängen, die wir im Rahmen der hier vorgestellten sowie weiterer Fallstudien gefunden haben, in umfassenden quantitativen Untersuchungen in einem Vergleichsgruppendesign zu verifizieren (Pollermann 2021, S. 115 ff.). Bezogen auf „Demokratie leben!“ könnten beispielsweise alle Regionen, die aus dem Bundes-

programm eine Förderung für eine Pfd erhalten, mit solchen verglichen werden, in denen das nicht der Fall ist. Außerdem ist es möglich, die Ergebnisse solcher Regionalstudien mit denen aus anderen Untersuchungen zu triangulieren, um ein gesteigertes Maß an Evidenz für die gefundenen Wirkungen zu erhalten. So haben zum Beispiel unsere Online-Erhebungen bei allen geförderten Akteur:innen in allen Programmbereichen des Bundesprogramms sowie die Befragung programm-externer Akteur:innen¹⁹ Befunde erbracht, die auf eine Bestätigung der Ergebnisse aus den Regionalanalysen hindeuten.

Literatur

- Autor:innengruppe (i. E.): Regionale Kooperation und Erträge der Programmförderung im Bundesprogramm „Demokratie leben!“ Ergebnisse aus acht Regionalanalysen. Halle (Saale): DJI/ Frankfurt am Main: ISS.
- Balzer, L./Beywl, W. (2018): evaluiert. Erweitertes Planungsbuch für Evaluationen im Bildungsbe- reich. Bern: hep.
- Bogumil, J./Holtkamp, L. (2023): Kommunalpolitik und Kommunalverwaltung. Eine praxisorien- tierte Einführung. Bonn: Bundeszentrale für politische Bildung.
- Bohnsack, R. (2014): Rekonstruktive Sozialforschung. Einführung in qualitative Methoden. 9., über- arbeitete und erweiterte Auflage. Opladen und Toronto: Barbara Budrich.
- Bohnsack, R./Nentwig-Gesemann, I./Nohl, A.-M. (2013): Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen qualitativer Sozialforschung. 3., aktualisierte Auflage. Wiesba- den: Springer VS.
- Göhler, G. (2010): Neue Perspektiven politischer Steuerung. In: Aus Politik und Zeitgeschichte, H. 2–3, S. 34–40.
- Herz, A./Peters, L./Truschkat, I. (2015): How to do qualitative strukturelle Analyse? Die qualitative Interpretation von Netzwerkkarten und erzählgenerierenden Interviews. In: Forum Qualitative Sozialforschung, 16. Jg., H. 1, Art. 9.
- Mayring, P. (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12., überarbeitete Aufla- ge. Weinheim und Basel: Beltz.
- Nohl, A.-M. (2017): Interview und Dokumentarische Methode. Anleitungen für die Forschungspra- xis. 5., aktualisierte und erweiterte Auflage. Wiesbaden: Springer VS.
- Ottmann, S./König, J. (2022): Wirkungsorientierung in der Sozialen Arbeit. Eine Einführung für Studium und Praxis. Stuttgart: Kohlhammer.
- Pollermann, K. (2021): Regional Governance: Begriffe, Wirkungszusammenhänge und Evaluations- ansätze. Thünen Report 87. Braunschweig: Johann Heinrich von Thünen-Institut. <https://10.3220/REPI626701303000>.
- Starke, P. (2022): Process Tracing in der Policy-Forschung. In: Wenzelburger, G./Zohlnhöfer, R. (Hrsg.): Handbuch Policy-Forschung, Wiesbaden: Springer Fachmedien, S. 453–482.
- Widmer, T. (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Strobl, R./ Lobermeier, O./Heitmeyer, W. (Hrsg.): Evaluation von Programmen und Projekten für eine de- mokratische Kultur. Wiesbaden: Springer VS, S. 41–68.

19 Bei der Befragung programm-externer bzw. semi-externer Akteur:innen wurden sachkundige Fachleute aus den Bereichen der Sozialen Arbeit, der Wohlfahrtspflege und Ähnliche sowie aus der Wissenschaft und Politik, die kein Projekt in „Demokratie leben!“ durchführen, gebeten, eine Einschätzung des Bundesprogramms vorzunehmen.

- Wilson-Grau, R. (2018): Outcome Harvesting: Principles, Steps, and Evaluation Applications. Charlotte, NC: Information Age Publishing Inc.
- Zimmer, A. (2010): Zivilgesellschaft und Demokratie in Zeiten des gesellschaftlichen Wandels. In: der moderne staat – dms, 3. Jg., H. 1, S. 147–163.

9 Die Analyse von Bedarfen in den Bereichen Demokratieförderung und Extremismusprävention

Grundsätzliche Überlegungen zur Konzeption und zum Zusammenspiel mit Wirkungsevaluationen

Laura Meijer und Pia Sauermann

9.1 Einleitung¹

In Deutschland existiert eine vielfältige Landschaft von Projekten, die sich im Rahmen des Bundesprogramms „Demokratie leben!“ der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention widmen. Diese Projekte agieren im Kontext dynamischer gesellschaftlicher Entwicklungen. Hierzu zählen unter anderem neue Partizipationsmöglichkeiten wie Bürgerräte sowie eine gestiegene Sensibilität für verschiedene Arten von Diskriminierung, aber auch neue Erscheinungsformen antidemokratischer Bewegungen wie die Querdenker:innen und antifeministische Influencer:innen oder auch die verstärkte Verbreitung von Hatespeech und Desinformation im digitalen Raum. Insbesondere für die Programmplanung und -steuerung ist dabei relevant zu wissen, auf welche aktuellen Entwicklungen Projekte und Programme reagieren und welche Phänomene und Zielgruppen sie verstärkt in den Blick nehmen sollten. Darüber hinaus sind auch die Bedürfnisse der Zielgruppen (u. a. Jugendliche und junge Erwachsene, Fachkräfte, die mit diesen arbeiten) hinsichtlich der Programmziele für die Programmplanung relevant. Hierfür können wissenschaftliche Erkenntnisse zu den Phänomenen und zu den entsprechenden Bedürfnissen bei den Zielgruppen und beteiligten Akteur:innen wichtige Hinweise liefern. Aus diesem Grund wurde im Rahmen des Bundesprogramms „Demokratie leben!“, im Übergang zur nächsten Förderperiode (ab 2025), die Umsetzung einer sogenannten Bedarfsanalyse vom BMFSFJ finanziert. Ziel dieses Projekts ist es, anhand von empirischen Untersuchungen wie beispielsweise qualitativen und quantitativen Befragungen von Zielgruppen und beteiligten Akteur:innen relevante Bedarfe für die Handlungsfelder Demokratieförderung und Extremismusprävention abzuleiten, um darauf basierend das Förderprogramm weiterzuentwickeln.

1 Wir danken Franziska Heinze und Frank Greuel für die wertvollen inhaltlichen Anmerkungen. Susanne Timling und Petra Göricke danken wir für das aufmerksame Korrektorat und Lektorat unseres Beitrags.

Die Frage nach einer auf wissenschaftlichen Daten basierenden Programm-entwicklung im Bereich der (sozialen) Arbeit mit jungen Menschen ist nicht neu. Auch zum Beispiel auf europäischer Ebene wird unter dem Begriff der evidenzbasierten Politik schon länger betont, dass die EU-Jugendpolitik auf wissenschaftlichen Erkenntnissen zur Lebensrealität und zu den Bedürfnissen junger Menschen basieren sollte (Hofmann-van de Poll/Pelzer 2023). In einer ähnlichen Logik sollte durch die Ermittlung von Bedarfslagen und Bedürfnissen von jungen Menschen und Fachkräften zu den Themen Demokratieförderung und Extremismusprävention eine Grundlage für die politische Planung der neuen Förderperiode von „Demokratie leben!“ geschaffen werden. Während jedoch Bedarfsanalysen, zum Beispiel in Unternehmen, in Gesundheitsprogrammen oder im Rahmen von Erwachsenenbildung, etablierte Verfahren sind (Choi/Park 2024; Jäger 2017; Schwikal/Steinmüller/Rohs 2017), liegen für derartige Bedarfsanalysen in den komplexen gesellschaftlichen Feldern Demokratieförderung und Extremismusprävention kaum Beispiele vor.

In diesem Beitrag setzen wir uns daher explorativ mit Bedarfsanalysen auseinander mit dem Ziel, sowohl zentrale Aspekte von Bedarfsanalysen in komplexen gesellschaftlichen Feldern zu diskutieren als auch zu ermitteln, ob und wie Bedarfs- und Wirkungsanalysen zusammenwirken. Zunächst stellen wir die Bedarfsanalyse kurz vor, die im Rahmen des Bundesprogramms „Demokratie leben!“ durchgeführt wurde (Abschnitt 9.2). In einem zweiten Schritt gehen wir der Frage nach, was überhaupt Bedarfe sind (Abschnitt 9.3). Denn obwohl der Bedarfsbegriff weitverbreitet ist, bleibt (nicht nur im alltäglichen Sprachgebrauch) oft unklar, was als Bedarf eingestuft wird bzw. wie und von wem Bedarfe festgelegt werden. Im dritten Teil (Abschnitt 9.4) diskutieren wir, wie Bedarfe ermittelt werden können. Anhand einer Einordnung von Bedarfsermittlungen im Rahmen von evidenzbasierter Politikgestaltung untersuchen wir, welche Herausforderungen eine Bedarfsanalyse in einem gesellschaftspolitischen Feld mit sich bringt und wie mit diesen umgegangen werden kann. Zuletzt (Abschnitt 9.5) gehen wir auf die Beziehung zwischen Bedarfen und Wirkungsevaluationen ein und analysieren, welchen Einfluss die Durchführung einer Bedarfsanalyse auf die spätere Wirkungsevaluation haben kann und welche Fallstricke dabei beachtet werden sollten.

9.2 Fallbeispiel: die Ermittlung von Bedarfen im Rahmen von „Demokratie leben!“

Am Deutschen Jugendinstitut haben wir in den Jahren 2023 und 2024 in Kooperation mit dem „SOCLES International Centre for Socio-Legal Studies“ das Projekt „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremis-

musprävention² durchgeführt. Dieses Forschungsprojekt sollte die Wissensbasis zu Bedarfslagen in den Bereichen Demokratieförderung und Extremismusprävention in Deutschland erweitern (für einen Überblick siehe Milbradt et al. 2024). Das Deutsche Zentrum für Integrations- und Migrationsforschung (DeZIM) widmete sich zeitgleich als externer Partner und in einem gesonderten Projekt Bedarfen im Bereich Vielfaltgestaltung.

Die Idee für die Bedarfsanalysen in den Bereichen Demokratieförderung, Vielfaltgestaltung und Extremismusprävention entstand im Rahmen des Bundesprogramms „Demokratie leben!“. Ziel war, die Weiterentwicklung des Programms stärker bedarfsorientiert zu gestalten und so eine Grundlage für die Entwicklung des Programms ab dem Jahr 2025 zu schaffen. Dabei sollte sich das Projekt nicht auf Bedarfe aus der derzeitigen Förderphase (2020 bis 2024) beschränken, sondern einen breiten Blick auf potenzielle Bedarfslagen in den genannten Bereichen werfen.

Um der Komplexität des Forschungsgegenstands Rechnung zu tragen, berücksichtigte das Projekt eine große Bandbreite an empirischer Evidenz. So erfolgte die anvisierte Erweiterung der Wissensbasis erstens über die systematische Aufbereitung bereits vorhandener Befunde sowie Daten und zweitens über die Durchführung eigener Erhebungen. Für beide Wege wendeten wir wiederum verschiedene methodische Vorgehensweisen an und berücksichtigten als Zielgruppen sowohl junge Menschen als auch die Fachpraxis, die mit Jugendlichen und jungen Erwachsenen arbeitet. In vier Teilprojekten wurden verschiedene Arbeitspakete umgesetzt: Um vorhandene Befunde und Daten möglichst effektiv zu nutzen, erstellten wir einen systematischen Literaturüberblick (Meijer et al. 2024) und führten Sekundäranalysen bereits verfügbarer Datensätze durch (Rottach/Wielath 2024). Darüber hinaus fanden im Rahmen des Projekts vier Primärerhebungen statt, in denen wir die Perspektiven junger Menschen sowie der Fachkräfte jeweils in einer qualitativen und in einer quantitativen Befragung eingefangen haben (Henschelmann/Meijer/Sauermann 2024; Kadera/Kindler/Witte 2024; Brandt et al. 2024).

Die Thematik der hier vorgestellten Bedarfsanalyse ist einerseits sehr breit und andererseits im Rahmen eines spezifischen Programmkontextes entstanden, sodass uns keine vergleichbaren Studien vorlagen, an denen wir uns orientieren konnten. Folglich begegneten uns in der Umsetzung immer wieder Fragen und Herausforderungen, insbesondere in der konzeptionellen Anfangsphase und bei der Interpretation der Ergebnisse. In den folgenden Abschnitten werden wir an verschiedenen Stellen Beispiele aus unserem Projekt einstreuen, zunächst aber auf die Frage eingehen, was eigentlich eine Bedarfsanalyse beinhaltet. Denn auch

2 <https://www.dji.de/ueber-uns/projekte/projekte/ermittlung-von-bedarfslagen-im-bereich-demokratieforderung-und-extremismuspraevention/projekt-publikationen.html> (Abruf 25.04.2024).

wenn dieser Beitrag letztlich den Fokus auf den Zusammenhang von Bedarfsanalysen und Wirkungsfragen legt, ist es sinnvoll, erst einmal einen Schritt zurückzutreten und zu klären, was eigentlich eine Bedarfsanalyse ist und was genau sie ermittelt. In einem ersten Schritt wird daher der Bedarfsbegriff näher beleuchtet.

9.3 Bedarfe und Bedürfnisse: eine Begriffsannäherung

Der Begriff „Bedarf“ findet im alltäglichen Sprachgebrauch, in politischen Entscheidungsprozessen sowie in wissenschaftlichen Erhebungen regelmäßig Verwendung. Zum Ausdruck gebracht wird dabei, dass etwas benötigt oder gewünscht wird. Bei genauer Betrachtung ist der Begriff jedoch weniger eindeutig als es auf den ersten Blick scheinen mag. Wie unterscheiden sich Bedarfe von Bedürfnissen oder von Wünschen? Ab wann gilt etwas als Bedarf? Und wer legt eigentlich fest, was ein Bedarf ist? Die Antworten auf diese und weitere Fragen beeinflussen, wie eine Bedarfsanalyse angelegt und durchgeführt wird.

Betrachtet man die deutsche und internationale Literatur zu Bedarfen, zeigt sich relativ schnell ein Unterschied: Während die deutsche Sprache zwischen „Bedarfen“ und „Bedürfnissen“ unterscheidet, spricht man im Englischen in beiden Fällen von „need“. Was ist also die Trennlinie zwischen Bedarfen und Bedürfnissen? Verschiedene Autor:innen weisen auf die unterschiedlichen Verwendungen des Bedürfnisbegriffs hin. So werden in der Ökonomie Bedürfnisse als etwas Subjektives, „ein Gefühl eines Mangels“ (Denninger 2022, S. 25), verstanden. In der Psychologie wird der Begriff zu den „inneren Beweggründe[n] des Verhaltens geknüpft, mit dem der Mensch zu seinem Wohlergehen strebt“ (Wendt 2023, S. 124), wobei zwischen unterschiedlichen Ebenen von Bedürfnissen, zum Beispiel Grundbedürfnissen oder Selbstverwirklichungsbedürfnissen, unterschieden wird (ebd.). Diese und weitere Verwendungen des Bedürfnisbegriffs teilen die Auffassung, dass Bedürfnisse individuell verankert sind und in der Regel von der (agierenden) Person selbst empfunden bzw. geäußert werden. Es handelt sich hier also um ein subjektives Verlangen nach etwas.

Aus manchen Bedürfnissen lassen sich Bedarfe ableiten, aber nicht alle Bedürfnisse sind unbedingt Bedarfe. Nullmeier beschreibt diese Trennung folgendermaßen:

„Bedarf zielt auf eine über das subjektive Empfinden hinausgehende Bestimmung von bestimmten Wünschen als notwendigen [sic] und zwingenden [sic] Anforderungen. Bedarf teilt die Gesamtheit aller Wünsche und Anliegen in solche, die ‚bloße Wünsche‘ sind, und solche, die als ‚notwendig‘ und deshalb als ‚vorrangig zu berücksichtigend‘ gelten“ (Nullmeier 2017, S. 1).

Bei Bedarfen kommt also eine normative Ebene ins Spiel, auf der festgelegt wird, welche Wünsche als Bedarf erachtet werden. Nach Nullmeier sind diese das Resultat gesellschaftlicher Aushandlungsprozesse, die bestimmten Wünschen „einen herausgehobenen Stellenwert“ (Nullmeier 2017, S. 3) zuordnen. In seiner „Taxonomy of Social Need“ differenziert auch Bradshaw (1972) zwischen sogenannten „felt needs“ und „normative needs“. Beim „felt need“ handelt es sich, ähnlich wie bei Nullmeier (2017), um ein individuelles Bedürfnis, einen Wunsch oder ein Anliegen. Ein „normative need“ hingegen ist ein Abgleich zwischen einem „desirable standard“ und dem existierenden Standard (Bradshaw 1972, S. 72). Auch für Bradshaw ist dieser wünschenswerte Standard nicht objektiv, sondern wird durch Expert:innen, Politik oder Wissenschaft ausgehandelt und definiert. Bedarfe sind also eine Lücke zwischen dem, was ist, und dem, was sein sollte – (normative) Festlegungen von Bedarfen sind somit „in no sense absolute“ (ebd.), sondern in starkem Maße von Wertorientierungen abhängig. Watkins und Kavale betonen deswegen, dass „[normative] needs typically shift over time“ (Watkins/Kavale 2014, S. 25), sich also im zeitlichen Verlauf verändern.

Bedarfe sind folglich einerseits eine Klassifizierung von Bedürfnissen, sie sind aber nicht auf die Individuen begrenzt. Auch auf organisationaler/institutioneller und auf gesellschaftlicher Ebene können Bedarfe bestehen. Beispiele für letztere wären ein gesamtgesellschaftlicher Bedarf nach dem Schutz von Minderheiten und der Demokratie (Denninger 2022, S. 29; Watkins/Kavale 2014, S. 25). Damit wird auch klar, dass nicht jedem Bedarf unbedingt ein subjektives Bedürfnis vorausgehen muss. Es können sogar – wie im Folgenden noch ausführlicher diskutiert wird (Abschnitt 9.4) – Widersprüche zwischen subjektiven Bedürfnissen und den gesellschaftlichen Bedarfen bestehen. Darüber hinaus muss zwischen latenten, manifesten und artikulierten Bedarfen unterschieden werden: Nicht immer sind Individuen sich ihrer Bedürfnisse bewusst. So kann manchmal ein latenter Bedarf erst durch geeignete Forschungsmethoden oder auch durch ein „Interesse generierendes Angebot“ (Käpplinger 2022, S. 63) entstehen (Schwikal/Steinmüller/Rohs 2017, S. 81; Jäger 2017, S. 12).

Die aufgeführten Betrachtungen zeigen, dass schon die Frage, was Bedarfe sind, nicht so einfach beantwortet werden kann. Bedarfe sind kontext- und wertabhängig. Welche Implikationen sich daraus für eine Bedarfsanalyse in einem gesellschaftspolitischen Themenfeld ergeben, wird im folgenden Abschnitt diskutiert. Dafür werden exemplarische Herausforderungen aus unserer Projektpraxis berichtet.

9.4 Bedarfsanalysen in gesellschaftspolitischen Feldern

Die begrifflichen Auseinandersetzungen (Abschnitt 9.3) implizieren, dass Bedarfe durch irgendeine Art von Prozess untersucht werden müssen. In der Regel

werden derartige Verfahren als Bedarfsanalyse oder Bedarfsermittlung bezeichnet. Einsatzbereiche für Bedarfsanalysen sind vielfältig und finden sich sowohl im privatwirtschaftlichen Bereich als auch mit Blick auf öffentliche Güter. Ein klassischer Anwendungsfall ist die Analyse von (Weiter-)Bildungsbedarfen, die sowohl firmenintern als auch von (Weiter-)Bildungsanbietern durchgeführt werden kann (siehe z. B. Schwikal/Steinmüller/Rohs 2017). Gerade Letztere sind breit aufgestellt und berücksichtigen verschiedene Interessengruppen (den Markt, die Gesellschaft, die Zielgruppe) sowie technologische und politische Entwicklungen (Jäger 2017, S. 11). Durch den deutlichen Bezug zu gesellschaftlichen Bedarfen stehen sie den hier diskutierten Bedarfsanalysen in den Bereichen Demokratieförderung und Extremismusprävention konzeptionell relativ nahe.

Werden die Ergebnisse von Bedarfsanalysen, wie in dem hier betrachteten Fall, zur Konzeption politischer Programme oder als Informationsgrundlage für politische Entscheidungen herangezogen, können sie als Komponente einer evidenzbasierten Politik erachtet werden. Diese bezeichnet ein Prinzip, nach dem politische Entscheidungen unter Einbezug wissenschaftlicher und systematischer Evidenz getroffen werden (Sutcliffe/Court 2005). Allerdings werden Bedarfsanalysen in der Literatur zu evidenzbasierter Politik nur selten thematisiert. Zwar unterscheiden sich die Vorstellungen darüber, was Evidenz in diesem Zusammenhang überhaupt umfasst, es zeigt sich jedoch häufig eine deutliche Begrenzung auf die Frage „what works, including costs and benefits“ (Evidence-Based Policymaking Collaborative 2016, S. 3). Diese Frage soll im Rahmen von Programmevaluationen beantwortet werden und zielt letztlich auf die Beurteilung von Effektivität oder Wirkungen von Programmen. Auf eine breitere Berücksichtigung von Evidenz, die nicht nur auf Evaluationen abstellt, verweisen zum Beispiel Hofmann-van de Poll und Pelzer (2023) im Kontext evidenzbasierter EU-Jugendpolitik. Evidenzen sind hier wissenschaftliche Erkenntnisse über die Lebensrealität junger Menschen, die dazu führen, dass „relevante Themen identifiziert werden“ (ebd., S. 170). Ein expliziter (wenn auch knapper) Hinweis darauf, dass auch Analysen von Problemen und Bedarfen, die einer politischen Intervention bedürfen, einen wenig beachteten, aber wichtigen Aspekt von Evidenz darstellen, findet sich bei Sanderson (2002). Er argumentiert, dass ein besseres Verständnis der spezifischen sozialen Probleme eine Voraussetzung für effektive politische Antworten darstellt (ebd., S. 4).

Die in diesem Beitrag fokussierte Bedarfsanalyse für die Bereiche Demokratieförderung und Extremismusprävention dient, im Sinne einer evidenzbasierten Politik, als empirische Grundlage für politische Entscheidungen. Darüber hinaus weisen deren Einsatzbereiche spezifische Eigenschaften auf. Im Folgenden wird diskutiert, welche Implikationen und Herausforderungen sich aus diesem Zusammenspiel ergeben. Hierfür wird Literatur zu evidenzbasierter Politik sowie zu Bedarfsanalysen herangezogen und in Bezug zu konkreten Erfahrungen

gesetzt, die wir bei der Umsetzung einer Bedarfsanalyse in den Bereichen Demokratieförderung und Extremismusprävention gemacht haben.

Wie bereits dargelegt wurde (Abschnitt 9.3), spielt Normativität bei Bedarfen – im Unterschied zu Bedürfnissen – immer eine Rolle, da sie stets ein Werturteil bezogen auf den erwünschten Zustand enthalten (Bradshaw 1972, S. 72). Bei der Durchführung einer Bedarfsanalyse muss also zunächst reflektiert werden, was überhaupt als Bedarf erachtet wird und warum. Die Bereiche Demokratieförderung und Extremismusprävention sind dabei in besonderem Maße von Normativität geprägt. Mit Blick auf politische Einstellungen, Haltungen, Kompetenzen, Verhaltens- und Sichtweisen sind konfligierende Vorstellungen über das Wünschenswerte in den unterschiedlichen politischen Lagern vorprogrammiert. Für Bedarfsanalysen ist es daher wichtig, dass der wünschenswerte Zustand, zu dem ein Bezug hergestellt wird, transparent gemacht, explizit als normativ markiert und gut begründet wird. Zum einen soll dadurch verhindert werden, dass Bedarfe als rein empirische Fakten missverstanden werden (Bradshaw 1972, S. 72), zum anderen sollen die wünschenswerten Standards für den demokratischen Diskurs zugänglich gemacht werden. Diese normativen Festlegungen sind als Arbeitsschritt zu betrachten und mit den entsprechenden finanziellen und personellen Ressourcen zu unterlegen.

Bei der Durchführung unserer Analysen zu Bedarfen der Demokratieförderung und Extremismusprävention, die von einem sehr eng getakteten Zeitplan geprägt waren, zeigte sich, dass dieser zu knapp bemessen war, um diesen vorangehenden Arbeitsschritt zu erfüllen. Infolgedessen mangelte es an einer systematischen Aushandlung mit den verschiedenen Stakeholdern, sowohl bezogen auf den erwünschten Zustand als auch auf die Auswahl zu berücksichtigender Zieldimensionen der Bedarfsanalysen. Letztere haben wir daher projektintern und mehr implizit als explizit durch die methodische Ausrichtung der Teilstudien festgelegt, in denen Bedarfe auf zwei Arten ermittelt wurden: Erstens wurden die Fachpraxis, aber auch die Jugendlichen selbst, nach Bedürfnissen gefragt. Geäußerte Bedürfnisse, beispielsweise nach Fortbildungen im Falle von Fachpraktiker:innen oder nach spezifischen politischen Bildungsangeboten im Falle der Jugendlichen und jungen Erwachsenen, wurden dann als Bedarfe interpretiert. Die normative Festlegung hinter diesem Vorgehen war, dass junge Menschen sowie Fachpraktiker:innen Expert:innen ihres eigenen Lebens bzw. ihres Arbeitsbereichs sind und ihre Angaben sowie Einschätzungen daher einen zentralen Stellenwert im Rahmen der Bedarfserhebung haben sollten. Zweitens wurden Problemlagen analysiert, aus denen sich (indirekt) Bedarfe der Demokratieförderung und Extremismusprävention ableiten lassen. Was als problematisch und bedarfsanzeigend bzw. wünschenswert erachtet wird, hängt hierbei sehr von etablierten methodischen Herangehensweisen ab, zum Beispiel von den verfügbaren Items zur Messung von politischen Einstellungen, die als demokratiegefährdend oder extremistisch gelten.

Bei der Planung der methodischen Ausrichtung besteht die Herausforderung, dass die übergreifenden Zielstellungen, insbesondere der Demokratieförderung, aber auch der Extremismusprävention, vergleichsweise unkonkret sind. Während es bei einer Weiterbildungsbedarfsanalyse im Bereich Erwachsenenbildung um einen „Soll-Ist-Abgleich zwischen erwarteten und tatsächlichen Qualifikationen, Kenntnissen und Fertigkeiten“ (Faulstich/Zeuner 2011, S. 44) geht, ist für eine Bedarfsanalyse in den Bereichen Demokratieförderung und Extremismusprävention ein breiteres Spektrum an individuellen Merkmalen und Aspekten potenziell relevant. Zu nennen sind hier zum Beispiel bestimmte demokratiegefährdende politische Haltungen und Einstellungen. Für eine Bedarfsanalyse ist eine in mehrfacher Hinsicht breite und möglichst ergebnisoffene Anlage daher von großer Bedeutung. Für den betrachteten Fall ist es in besonderem Maße plausibel, dass eine zu enge Ausrichtung auf bestimmte Themenfelder und antizipierte Problemlagen dazu führen kann, dass der Bedarf durch das gewählte Design quasi vorbestimmt ist oder dass zum Beispiel neuere gesellschaftliche Entwicklungen, denen ein demokratiegefährdendes Potenzial innewohnt, übersehen werden. Die gebotene Offenheit betrifft nicht nur die Themensetzung, sondern auch die Methoden. Nach Sutcliffe und Court solle „a wide breadth of sources of evidence, not just hard evidence“ (2005, S. 4) in den politischen Prozess einfließen. Sie sprechen sich damit gegen eine Position aus, nach der quantitative Daten als „hard evidence“ bezeichnet und den als „soft evidence“ betrachteten qualitativen Daten vorgezogen würden (Sutcliffe/Court 2005, S. 4). Altschuld und Watkins (2014b, S. 9) verweisen in diesem Zusammenhang auch darauf, dass (rein) quantitative Bedarfsmessungen als Top-Down-Prozess wahrgenommen werden können und subjektiv empfundene Bedürfnisse sich besser durch qualitative Erhebungen erfassen lassen. Darüber hinaus ist es wichtig, dass verschiedene beteiligte Akteur:innen und Zielgruppen befragt werden. Hierzu zählen im engeren Sinne junge Menschen, die eine wesentliche Zielgruppe von Maßnahmen der politischen Bildung und Demokratieförderung darstellen, sowie (pädagogische) Fachkräfte. Denkbar ist jedoch auch die ergänzende Befragung weiterer Gruppen, zum Beispiel Eltern oder Expert:innen.

In unserem Fall wurde der Breite des Themenbereichs durch den Einbezug verschiedener Zielgruppen bzw. Akteur:innen Rechnung getragen (Jugendliche, junge Erwachsene, Fachkräfte aus verschiedenen Bereichen). Darüber hinaus haben wir sowohl qualitative als auch quantitative Forschungsdesigns eingesetzt. Während die Festlegung von Bedarfen in den qualitativen Teilstudien dabei in starkem Maße von den Befragten ausgeht, ist sie bei den quantitativen Befragungen auch von den Forschenden beeinflusst, welche die geäußerten Bedarfe durch die geschlossene Fragestellung schon im Vorfeld begrenzen und auf bestimmte Themen fokussieren. Mit Blick auf die Bedarfsabfrage bei Jugendlichen, für die bisher wenig Erfahrungsberichte vor allem im Rahmen abstrakter gesellschaftspolitischer Themenkomplexe vorliegen, zeigen unsere Erfahrungen, dass

junge Menschen in quantitativen Befragungen durchaus unproblematisch auf geschlossene Fragen antworten konnten (z. B. zu ihrem Bedürfnis nach verschiedenen Arten der Unterstützung nach einer Begegnung mit Hass im Netz). Es fiel aber manchen jungen Menschen schwer, im Rahmen einer qualitativen Befragung auf offene Fragen nach ihren Bedürfnissen zu reagieren. Dafür kann es eine Vielzahl an möglichen Gründen geben, z. B., dass die Themenbereiche für junge Menschen zu abstrakt sind, um dazu klare Bedürfnisse zu äußern, oder auch, dass ihnen ihre Bedürfnisse nicht bewusst sind. Entsprechend könnte es sinnvoll sein, hierfür neue Forschungsdesigns zu entwickeln und zu erproben, die speziell für die Erhebung von Bedürfnissen junger Menschen eingesetzt werden können.

Es wurde bereits dargelegt, dass in den von uns betrachteten Bereichen Demokratieförderung und Extremismusprävention ein Spezifikum darin besteht, dass die individuellen Bedürfnisse teilweise deutlich von den normativ festgelegten Bedarfen abweichen. Dieser Aspekt muss sowohl in der konzeptionellen Phase als auch bei der Interpretation der Befunde Berücksichtigung finden. Zwar gilt generell, dass Bedürfnisse von Bedarfen abweichen können, die durch Politik, Wissenschaft oder Fachkräfte festgelegt werden (Käpplinger 2022, S. 63). Ein kritisches Moment ist zudem „[d]er Weg von einem Bedarf hin zu einer tatsächlichen Nachfrage“ (Denninger 2022, S. 36), das heißt die Frage, ob ein geäußelter Bedarf auch tatsächlich handlungsleitend wird. In den Bereichen Demokratieförderung und Extremismusprävention kann es darüber hinaus vorkommen, dass die normativen gesellschaftlichen Bedarfe den individuellen Bedürfnissen entgegengesetzt sind. Beispielsweise entspringen pädagogische Angebote für junge Menschen mit extremistischen Tendenzen weniger deren individuellen Bedürfnissen als einem gesamtgesellschaftlichen Bedarf nach Demokratie und Schutz von Minderheiten. Wie bereits erwähnt, ist es daher wichtig, dass Bedarfsanalysen in diesen Bereichen dringend über die Abfrage von Bildungs- oder Unterstützungsbedürfnissen hinausgehen. So sollten zusätzlich Problemlagen, zum Beispiel die Verbreitung bestimmter antidemokratischer Einstellungen, analysiert werden, die indirekt auf Bedarfe nach Angeboten der Demokratieförderung und Extremismusprävention verweisen. Wie in unserem Forschungsprojekt geschehen, können die fokussierten Gruppen und Akteur:innen auch nach ihrer Wahrnehmung von und Betroffenheit sowie Belastung durch demokratiegefährdende Phänomene befragt werden. Bei der Ergebnisinterpretation stellt sich dann die Aufgabe, diese Stränge gut zusammenzubringen. Es hat sich in unserem Forschungsprojekt gezeigt, dass dieser Schritt Zeit kostet und entsprechend in den Zeitplänen verankert werden muss.

Insgesamt zeigt sich, dass Bedarfsanalysen mit der thematischen Spezifik des betrachteten Anwendungsfalls Herausforderungen gegenüberstehen, die über diejenigen der klassischen Anwendungsgebiete hinausgehen. Zusammenfassend lässt sich festhalten: (1) Die notwendigen normativen Setzungen sollten zu Beginn einer Bedarfsanalyse transparent gemacht werden, (2) das Forschungs-

design ist thematisch sowie mit Blick auf die methodische Ausrichtung und die anvisierten Zielgruppen und Akteur:innen vielfältig aufzustellen und (3) Bedarfe können auch indirekt aus der Analyse von Problemlagen und der Beschreibung von Betroffenheit und Belastungen abgeleitet werden.

Im folgenden Abschnitt wird die Schnittstelle und das Zusammenspiel zwischen Bedarfsanalysen und Wirkungsevaluation, die ein zentraler Bestandteil evidenzbasierter Politik ist, betrachtet.

9.5 Bedarfe und die Evaluation von Wirkungen: ein wenig diskutierter Zusammenhang

Bedarfsanalysen stehen typischerweise am Anfang einer Programm- oder Projektentwicklung, denn sie sollen Informationen dazu liefern, welche Bedarfe bisher nicht gedeckt sind oder welche gesellschaftlichen Problemlagen eine Reaktion erfordern. Die Vorstellung, dass Bedarfsermittlungen auch (automatisch) feststellen, welche Art von Intervention oder Projekt erforderlich ist, um den ermittelten Bedarf zu decken, greift jedoch zu kurz. Bedarfe sind keine Lösungen: Sie zeigen in erster Linie die Lücke zwischen dem, was ist, und dem, was sein soll, auf. Daran anschließend gibt es eine Vielzahl an Möglichkeiten, auf diese Lücke zu reagieren (Watkins/Kavale 2014, S. 24). Beispielsweise war ein Befund unserer Jugendbefragung, dass viele junge Menschen mit digitalen Phänomenen wie Falschnachrichten und Hass im Netz konfrontiert sind und gerne mehr darüber lernen würden, wie man mit diesen Phänomenen umgeht (Henschelmann/Meijer/Sauermann 2024, S. 24 f.). Auch ließ sich aus der Befragung ableiten, dass es einen Bedarf gibt, die politische Selbstwirksamkeit von Mädchen und jungen Frauen zu stärken (ebd., S. 43 f.). Diese Befunde können die (Weiter-)Entwicklung von Bundesprogrammen wie „Demokratie leben!“ anregen. Schwieriger ist es jedoch, aus diesen Befunden konkrete Empfehlungen für geeignete Maßnahmen abzuleiten. Welche Aktivitäten am besten geeignet sind, um auf die ermittelten Bedarfe zu reagieren, ist alles andere als trivial und lässt sich nur unter Hinzunahme weiterer Expertise beantworten. Hier kommen unter anderem Evaluationen ins Spiel.

Scriven und Roth (1990) differenzieren zwischen „performance deficit needs“ (z. B. ein Bedarf einer höheren Selbstwirksamkeit bei Mädchen und jungen Frauen) und „treatment deficit needs“ (die Bearbeitung dieses Bedarfs). Ihrer Meinung nach werden diese zwei Bedarfe oft verwechselt oder sogar gleichgesetzt, was dazu führt, dass zum Beispiel aus dem Bedarf einer verbesserten Selbstwirksamkeit von Mädchen und jungen Frauen ein Bedarf nach mehr Selbstwirksamkeitstrainings für junge Mädchen und junge Frauen wird („verbally similar treatment-need claims“, ebd., S. 138). Dabei wird aber übersprungen, dass „if one is to move

[...] from a performance-deficit statement to a treatment-deficit one, it is necessary to get into evaluation“ (ebd.). Es gibt nämlich eine Reihe an Möglichkeiten, auf zum Beispiel die fehlende Selbstwirksamkeit von Mädchen und jungen Frauen zu reagieren (wissensbasierte Ansätze, Partizipationsformate, Schulungen für Lehrpersonal, um nur einige zu nennen) und es sind unter anderem Wirkungsanalysen von schon existierenden Programmen oder Projekten, die Hinweise dazu liefern können, welche Maßnahmen wirkungsvoll auf Bedarfe reagieren (können) (Altschuld/Watkins 2014b).

Wo Wirkungsevaluationen herangezogen werden können, um in einer Programmplanung den Schritt von identifizierten Bedarfen zu daraus abzuleitenden Maßnahmen zu gehen, können Bedarfsanalysen wiederum als Baustein für Wirkungsfragen dienen. In dem Sammelband „Needs Assessment: Trends and a View Toward the Future“ gehen Altschuld und Watkins (2014a) der Frage nach, wie Bedarfsanalysen und Evaluationen zusammenhängen. Sie argumentieren, dass Evaluationen und Bedarfsanalysen zwar jeweils eigene Fragen beantworten, sie aber dennoch sowohl in den Methoden als auch in den Zielen aneinander anschließen. Denn eine Evaluation der Wirksamkeit eines Programms oder Projekts geht oft der Frage nach, ob das Programm oder Projekt auf vorher identifizierte Bedarfe der (unterschiedlichen) Zielgruppen eingegangen ist. Das macht die Identifizierung von Bedarfen zu einem wichtigen Bestandteil eines Evaluationsverfahrens. „If needs have been identified, prioritized, and their causes determined with solutions selected, these factors contribute directly to evaluation of the project or program – inevitably linking the two fields of study and practice“ (Altschuld/Watkins 2014b, S. 7).

In einem weiteren Beitrag im gleichen Sammelband zeigen Engle und Altschuld (2014) die Verknüpfung von Bedarfsanalysen und (Wirkungs-)Evaluationen anhand des Logischen Modells auf. In ihrer Darstellung des Logischen Modells befindet sich die Bedarfsanalyse ganz am Anfang und legt damit die Basis für die nächsten Schritte. Das Autorenteam argumentiert, dass ein Programm oder eine Maßnahme nur dann sinnvoll entwickelt und danach evaluiert werden kann, wenn klar ist, auf welche Problemlage oder welchen Bedarf das Programm reagiert. Ebenso argumentieren Watkins und Kavale (2014): Ohne dass klar ist, welche Bedarfe es gibt, wird es schwierig oder sogar unmöglich, Wirkungen zu evaluieren. Nach dieser Logik sollen Bedarfsanalysen also ermitteln, was gebraucht wird, während Wirkungsevaluationen dabei helfen können zu verstehen, ob und wie Maßnahmen dazu beitragen, den ermittelten Bedarf zu decken. Fragestellungen von (Wirkungs-)Evaluationen sollten somit zumindest zum Teil auf Bedarfsanalysen Bezug nehmen (Watkins/Kavale 2014).

Dass Bedarfe bei Wirkungsevaluationen eine Rolle spielen, ist erst mal nicht überraschend, denn es ist de facto nicht möglich, dass ein Programm ein nicht-existierendes Problem angeht, und folglich kann auch eine entsprechende Wirkung nicht festgestellt werden. Auch beeinflusst es die Wirkungen eines

Programms, wenn es sich nicht auf bestehende Bedarfe bezieht (nach Rossi / Freeman / Lipsey 2004). Es ist also durchaus sinnvoll, vor oder auch während eines Programms oder Projekts zu analysieren, welche Bedarfe bestehen und diese als Ausgangspunkt der Konzeption einer Intervention in Wirkungsuntersuchungen zu berücksichtigen. Generell muss dabei allerdings immer im Blick behalten werden, wie und auf welcher Ebene Bedarfe erhoben wurden und ob und wie dies mit Wirkungsevaluationen in komplexen Programmen verknüpft werden kann. So könnte der Bedarf nach mehr Medienbildung in Bezug auf Hass im Netz, der im Rahmen der Jugendbefragungen ermittelt wurde, durchaus im Rahmen eines Modellprojekts, das sich genau diese Medienbildung zum Ziel gesetzt hat, aufgegriffen und operationalisiert werden. In der Folge könnten Wirkungsevaluationen untersuchen, inwiefern es dem Modellprojekt gelingt, nicht nur neue medienpädagogische Ansätze zu Hass im Netz zu erproben, sondern dabei auch die Bedarfe von Jugendlichen, zum Beispiel nach mehr Handlungssicherheit im Umgang mit Hass im Netz, zu befriedigen. Anders sähe es aus, wenn Wirkungsuntersuchungen beispielsweise die Aktivitäten von Partnerschaften für Demokratie (als kommunalen Handlungsbereich) oder von Kompetenznetzwerken (als Handlungsbereich auf der Bundesebene) in den Blick nehmen. Diese Akteur:innen arbeiten nur indirekt zu den Bedarfen junger Menschen, im Fokus steht dort vielmehr die strukturelle Entwicklung der Angebotslandschaft. Diese strukturelle Dimension hat letztendlich auch das Ziel, Bedarfe junger Menschen in der Zukunft besser decken zu können – konkret steht jedoch im Vordergrund, wie kommunales Handeln in Partnerschaften für Demokratie oder Qualifizierungsangebote von bundesweit tätigen Kompetenznetzwerken dazu beitragen, dass pädagogische Angebote vor Ort entstehen bzw. qualifiziert sind, um junge Menschen im Umgang mit Hass im Netz befähigen zu können. Eine direkte Wirkung solcher Maßnahmen auf den vorab identifizierten Bedarf junger Menschen nach mehr Medienbildung und einer besseren Befähigung im Umgang mit Hass im Netz wäre hier kein angemessener Fokus von Wirkungsuntersuchungen. Passender wäre hier, die vorab erhobenen Bedarfe nur als eine Art Rahmen der Wirkungsevaluation zu sehen und diesen mit konkreteren Bedarfen innerhalb der Maßnahme zu erweitern.

Dabei ist es aus unserer Sicht auch wichtig zu vermeiden, dass die Durchführung einer Bedarfsanalyse Wirkungsfragen auf einen Vorher-Nachher-Vergleich reduziert, indem nur der Frage nachgegangen wird, ob sich der Bedarf nach Einführung der Maßnahme reduziert hat. Denn es gibt gute Gründe, so eine simple Verknüpfung von Bedarfen und Wirkungen kritisch zu betrachten. So sind Vorher-Nachher-Vergleiche von z. B. pädagogischen Interventionen wenig aufschlussreich, da sie keinerlei Aussage dazu erlauben, warum Bedarfe nach einer Intervention (nicht) gedeckt wurden. Ob sich die Bedarfe aufgrund gesellschaftlicher Entwicklungen geändert haben, ob die Entscheidung für die Maßnahme (nicht) richtig oder die Mittelausstattung zu gering war, lässt sich mit

einem solchen Vergleich nicht beantworten. Dazu kommt, dass Bedarfsanalysen immer eine Momentaufnahme sind.

„Bedarfe befinden sich häufig in einem diffusen, nicht reliablen Zustand und können durch vielfältige Einflussfaktoren bedingt oder gar verfälscht werden. So sind sie bspw. auch eine Reaktion auf gegenwärtige gesellschaftliche, technologische und wirtschaftliche Wandlungsprozesse“ (Denninger 2022, S. 30).

Nicht nur gesellschaftliche Veränderungen prägen Bedarfe, auch neue Erfahrungen oder der Zugang zu neuen Informationen, zum Beispiel im Rahmen von Workshops oder durch die Inanspruchnahme von Beratungsangeboten, können individuelle Bedürfnisse und Bedarfe verändern (Denninger 2022). Somit können sich auch schon während eines Programms Bedarfe verändern oder latente Bedarfe erst durch das Programm oder die Maßnahme sichtbar oder benennbar werden. Das bedeutet nicht, dass Bedarfe nicht herangezogen werden sollten, denn auch Momentaufnahmen können Hinweise auf zu berücksichtigende Bedarfe geben. Vielmehr sollten die Kontextbedingungen, die Bedarfe beeinflussen, und der Wandel, dem Bedarfe unterliegen, in die Analyse einbezogen werden.

Überdies kann eine enge Begrenzung auf vorab identifizierte Bedarfe und Bedürfnisse dazu führen, dass mögliche andere (unerwartete) Wirkungen übersehen werden. Nach dem Motto, „man findet nur, wonach man sucht“, kann ein Bedarfsbezug die Sicht einschränken. Ein Programm kann aber auch unerwartete oder unerwünschte Wirkungen haben, die losgelöst von Bedarfen entstehen oder sogar neue Hinweise auf Bedarfe liefern.

Zudem ist auch die Prozesshaftigkeit des Gegenstands im Blick zu behalten. Projekte im Bereich der Sozialen Arbeit, zu denen auch viele Projekte in den Themenfeldern der Demokratieförderung und Extremismusprävention zählen, haben oft bewusst einen prozessorientierten, offenen und flexiblen Charakter. Das steht in einem Spannungsverhältnis zu vordefinierten Wirkungsfragen oder auch Vorher-Nachher-Messungen (St Croix/Doherty 2022, S. 2).

Zusammenfassend wurden in diesem Kapitel verschiedene Verknüpfungen zwischen Bedarfs- und Wirkungsanalysen aufgezeigt. Ergebnisse einer Wirkungsanalyse können dabei helfen, Bedarfslagen sinnvoll für die konkrete Programmplanung zu nutzen. Bedarfsanalysen wiederum machen auf Bedarfe und Problemlagen aufmerksam und können somit als ein Bezugspunkt für Evaluationen dienen. Darüber hinaus ermöglichen sie es, programmbezogene Evaluationen in Bezug zu gesellschaftlichen Problemlagen zu setzen. Wichtig ist dabei, zu vermeiden, dass Bedarfsanalysen zu einem verkürzten Verständnis von Wirkungsevaluationen führen. Ein einfacher „Vorher-Nachher-Vergleich“ von Bedarfen wird der Komplexität des Gegenstands sowie der Kontextabhängigkeit von Bedarfen und Wirkungen nicht gerecht und kann die Frage nicht beantworten, ob die beobachteten Veränderungen überhaupt durch die Maßnah-

me erzeugt wurde. Um ein produktives Zusammenwirken von Bedarfsanalysen und Wirkungsanalysen zu ermöglichen, braucht es unserer Erfahrung nach eine frühe Planung, die sowohl die Austauschformate als auch die inhaltlichen Überschneidungen konkretisiert. In unserem Fall fand aufgrund der Zuordnung in unterschiedliche Forschungsprojekte anfangs relativ wenig Austausch mit den Kolleg:innen der Programmevaluation „Demokratie leben!“ statt. Vereinzelt stattgefundene Gespräche erwiesen sich allerdings als äußerst fruchtbar. So brachte die Programmevaluation „Demokratie leben!“ durch ihre Praxisnähe eigene Bedarfsfragen mit, zum Beispiel den Bedarf junger Menschen an verschiedenen Arten von Engagement, und ergänzte die „defizitorientierte“ Perspektive der Bedarfsanalysen mit Ergebnissen zu den existierenden Ressourcen im Praxisfeld. Umgekehrt gab es auch ein Interesse an unseren Daten, da diese einen Blick auf die allgemeinen Bedürfnisse von jungen Menschen, aber auch insbesondere von Fachkräften, die außerhalb der Evaluation befragt wurden, ermöglicht. Eine frühere Verzahnung der beiden Projekte hätte ermöglicht, Fragestellungen und Herangehensweisen besser aufeinander abzustimmen, um so Synergiepotenziale zu fördern und die Perspektive auf Bedarfe in den aktuell laufenden und zukünftigen Wirkungsuntersuchungen im Kontext des Bundesprogramms „Demokratie leben!“ zu stärken.

9.6 Fazit

Die in diesem Beitrag präsentierten Überlegungen stehen noch am Anfang. Mit einem wachsenden Interesse an datenbasierter Programmentwicklung auch in ähnlichen Bereichen, zum Beispiel in der politischen Bildung (siehe u. a. die Machbarkeitsstudie „Monitor politische Bildung“³), scheint es jedoch relevant, schon an dieser Stelle die Diskussion um die (Un-)Möglichkeiten einer Bedarfsorientierung in komplexen gesellschaftspolitischen Feldern zu öffnen. Beginnend mit grundsätzlichen begrifflichen und konzeptionellen Überlegungen zu Bedarfsanalysen wurde daher herausgearbeitet, welche spezifischen Merkmale Bedarfsanalysen in den Bereichen Demokratieförderung und Extremismusprävention aufweisen und welche Implikationen sich daraus für Wirkungsanalysen ergeben. Anhand unseres Forschungsprojekts haben wir überdies verschiedene Herausforderungen offengelegt, die uns in der Realisierung der Bedarfsanalysen begegnet sind, und unseren Umgang mit diesen geschildert.

Unseres Erachtens ergeben sich die spezifischen Herausforderungen einer Bedarfsanalyse in den Bereichen Demokratieförderung und Extremismusprävention vor allem aus den folgenden Merkmalen: (1) Wie auch Bedarfsanalysen

3 Siehe bericht-pb.de (Abruf 17.10.2024).

in anderen Bereichen sind normative Festlegungen darüber notwendig, welche Zustände wünschenswert sind bzw. wessen Bedürfnisse als Bedarfe anerkannt werden. (2) Das Forschungsdesign ist thematisch sowie mit Blick auf die methodische Ausrichtung und die anvisierten Zielgruppen und Akteur:innen vielfältig aufzustellen. (3) Bedarfe ergeben sich auch, aber nicht nur aus Bedürfnissen, die von Zielgruppen und beteiligten Akteur:innen geäußert werden.

Die genannten Merkmale haben nicht nur für die Konzeption und das Forschungsdesign der Bedarfsanalysen selbst Folgen. Wie gezeigt wurde, gibt es an verschiedenen Stellen Verknüpfungen zwischen Bedarfsanalysen und Wirkungsevaluationen, auf denen der thematische Fokus dieses Sammelbands liegt. Einerseits können Wirkungsanalysen Hinweise dazu liefern, welche Maßnahmen geeignet sind, um auf Bedarfslagen zu reagieren. Bedarfslagen wiederum liefern Hinweise auf Bedarfe und Problemlagen, die in einer Evaluation in den Blick genommen werden können. Die Vorstellung, dass Evaluationen prüfen, ob die Programme die identifizierten Bedarfe decken, ist jedoch aus mehrfacher Hinsicht verkürzt und lässt beiseite, dass an unterschiedlichen Stellen im Prozess von Bedarfsermittlung zur Evaluation normative, methodische, inhaltliche und politische Fragen eine zentrale Rolle spielen. Eine frühe Verzahnung von Bedarfsanalysen und Evaluationen könnte aus unser Sicht dafür sorgen, dass Synergien frühzeitig erkannt werden und gleichzeitig der Prozesshaftigkeit und Kontextabhängigkeit beider Stränge Rechnung getragen wird.

Literatur

- Altschuld, J. W. /Watkins, R. (Hrsg.) (2014a): Needs assessment: Trends and a view toward the future. In: *New Directions for Evaluation*, Number 144. New York, NY: John Wiley & Sons.
- Altschuld, J. W. /Watkins, R. (2014b): A Primer on Needs Assessment: More than 40 Years of Research and Practice. In: Altschuld, J. W. /Watkins, R. (Hrsg.): *Needs assessment: Trends and a view toward the future. New Directions for Evaluation*, Number 144. New York, NY: John Wiley & Sons, S. 5–18.
- Bradshaw, J. (1972): A Taxonomy of Social Need. In: McLachlan, G. (Hrsg.): *Problems and Progress in Medical Care: Essays on Current Research*, 7th Series. London: Oxford University Press, S. 71–82.
- Brandt, L. A. /Meysen, T. /Kadera, S. /Kindler, H. /Witte, S. (2024): Explorative Analyse zu Bedarfen der Kinder- und Jugendhilfe. Teilbericht 6 des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. Halle (Saale): DJI.
- Choi, H. J. /Park, J. H. (2024): Research Trends in Learning Needs Assessment: A Review of Publications in Selected Journals from 1997 to 2023. In: *Sustainability*, 16. Jg., H. 1, S. 382.
- Denninger, A. (2022): Unternehmensinterne Bedarfstransformationsprozesse als Herausforderung und Ansatzpunkt einer nachfrageorientierten wissenschaftlichen Weiterbildung. Dissertation. Gießen: Justus-Liebig-Universität.
- Engle, M. /Altschuld, J. W. (2014): Needs Assessment: The Perspective From the Public Sector. In: Altschuld, J. W. /Watkins, R. (Hrsg.): *Needs assessment: Trends and a view toward the future. New Directions for Evaluation*, Number 144. New York, NY: John Wiley & Sons, S. 33–45.

- Evidence-Based Policymaking Collaborative (2016): Principles of Evidence-Based Policymaking. https://www.urban.org/sites/default/files/publication/99739/principles_of_evidence-based_policymaking.pdf (Abruf 24.04.2024).
- Faulstich, P./Zeuner, C. (2011): Erwachsenenbildung. Weinheim und Basel: Beltz.
- Henschelmann, K.-P./Meijer, L./Saueremann, P. (2024): Standardisierte Online-Erhebung mit Jugendlichen und jungen Erwachsenen von 16 bis 27 Jahren. Teilbericht 4 des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. Halle (Saale): DJI.
- Hofmann-van de Poll, F./Pelzer, M. (2023): Wissensbasiert und erkenntnisgestützt? Ein Blick auf evidenzbasierte EU-Jugendpolitik. In: IJAB (Hrsg.): Forum Jugendarbeit International 2019–2022. Schwerpunktthema: Internationale Jugendarbeit zukunftsfähig gestalten. Bonn, S. 168–178.
- Jäger, C. (2017): Bildungsbedarfsanalyse. Methoden zur Ermittlung von Bildungsbedarfen. In: Team der Wissenschaftlichen Weiterbildung der Universität Rostock (Hrsg.): Weiterbildungsmanagement professionalisieren. Anregungen aus der Weiterbildungspraxis. Universität Rostock, S. 13–23.
- Kadera, S./Kindler, H./Witte, S. (2024): Einrichtungs- und Fachkräftebefragung der Kindertagesstätten und Schulen in Deutschland. Teilbericht 5 des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. München: DJI.
- Käpplinger, B. (2022): Bedarf in der Programmplanung von Volkshochschulen in ländlichen Räumen. Partizipative Erarbeitung einer Handreichung. In: Hessische Blätter für Volksbildung, 72. Jg., H. 4, S. 61–69.
- Meijer, L./Saueremann, P./Henschelmann, K.-P./Lien, S.-C. (2024): Systematischer Literaturüberblick. Teilbericht 2 des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. Halle (Saale): DJI.
- Milbradt, B./Kindler, H./Seckinger, M./Meysen, T. (2024): Zentrale Befunde und Empfehlungen des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. Halle (Saale): DJI.
- Nullmeier, F. (2017): Perspektiven auf eine Theorie der Bedarfsgerechtigkeit in zehn Thesen. Working Paper, Bd. 17. Helmut Schmidt Universität Hamburg: DFG Research Group 2104.
- Rossi, P. H./Lipsey, M. W./Freeman, H. E. (2004): Evaluation. A Systematic Approach. Thousand Oaks: Sage.
- Rottach, A./Wielath, S. (2024): Sekundäranalysen nationaler und internationaler Datenbasen. Teilbericht 3 des Projekts „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“. Halle (Saale): DJI.
- Sanderson, I. (2002): Evaluation, Policy Learning and Evidence-Based Policy Making. In: Public Administration 80. Jg., H. 1, S. 1–22.
- Schwikal, A./Steinmüller, B./Rohs, M. (2017): Bedarfsorientierte Entwicklung von Studienangeboten in der wissenschaftlichen Weiterbildung. In: Hochschulmanagement 12. Jg., H. 2+3, S. 79–88.
- Scriven, M./Roth, J. (1990): Special Feature: Needs Assessment. In: Evaluation Practice, 11. Jg., H. 2, S. 135–140.
- St Croix, T. d./Doherty, L. (2022): „Capturing the magic“: Grassroots Perspectives on Evaluating Open Youth Work. In: Journal of Youth Studies, 27. Jg., H. 4, S. 486–502.
- Sutcliffe, S./Court, J. (2005): Evidence-Based-Policymaking. What is it? How does it work? What relevance for developing countries? London: Overseas Development Institute.
- Watkins, R./Kavale, J. (2014): Needs: Defining What You Are Assessing. In: Altschuld, J. W./Watkins, R. (Hrsg.): Needs assessment: Trends and a view toward the future. New Directions for Evaluation, Number 144. New York, NY: John Wiley & Sons, S. 19–31.
- Wendt, W. R. (2023): Bedürftigkeit, Bedürfnis und Bedarf begreifen. In: Blätter der Wohlfahrtspflege, 170. Jg., H. 4, S. 123–125.

IV Statt eines Fazits

10 Anforderungen an wirkungsorientierte Evaluationen im Bereich der Demokratieförderung, politischen Bildung und Prävention von Extremismus

Ein interdisziplinäres Fachgespräch

Es diskutierten Helle Becker, Jörg Faust, Hemma Mayrhofer, Björn Milbradt und Thomas Widmer

Diskussionen um Evaluation und Wirkung in (sozial-)pädagogischen Arbeitsfeldern und in der politischen Bildung haben eine lange Tradition und sind über die Jahre von zahlreichen Lernprozessen begleitet gewesen. Auch aus der Evaluation von Maßnahmen der Demokratieförderung und Unterstützung von Governance in der Entwicklungszusammenarbeit liegen vielfältige Erfahrungen mit Wirkungsevaluationen und ihren Chancen, Risiken und Grenzen vor. Vor diesem Hintergrund veranstalteten wir am 26. März 2024 als Herausgeber:innen dieses Sammelbandes mit evaluationserfahrenen Expert:innen aus der außerschulischen politischen Bildung, der Soziologie und Sozialen Arbeit, den Erziehungs- und Politikwissenschaften sowie der Entwicklungszusammenarbeit ein Online-Fachgespräch zu Fragen der Wirkungsevaluation in den genannten Feldern mit Bezug zu Themen der Demokratieförderung, Vielfaltgestaltung und Extremismusprävention. Es diskutierten: aus einer Perspektive der außerschulischen, non-formalen politischen Bildung und als Vertreterin der Fachstelle politische Bildung des Vereins Transfer für Bildung e. V. in Essen Dr.in Helle Becker, aus Perspektive des Deutschen Evaluierungsinstituts der Entwicklungszusammenarbeit (DEval) in Bonn Prof. Dr. Jörg Faust, vor dem Hintergrund der Evaluation Sozialer Arbeit Assistenz-Professorin Dr. Hemma Mayrhofer vom Institut für angewandte Rechts- und Kriminalsoziologie der Universität Innsbruck, aus Sicht der Evaluationspraxis von Programmen der Demokratieförderung sowie Prävention von politischem Extremismus und, als Vertreter der gastgebenden Arbeitseinheit, der Fachgruppe „Politische Sozialisation und Demokratieförderung“ in der Abteilung Jugend und Jugendhilfe am Deutschen Jugendinstitut in Halle Dr. Björn Milbradt, und aus einer politikwissenschaftlichen Perspektive auf Evaluation Prof. Dr. Thomas Widmer vom Institut für Politikwissenschaft der Universität Zürich.

Das Gespräch zeichneten wir auf. Der nun vorliegende Text stellt eine transkribierte und redaktionell leicht bearbeitete Fassung des Fachgesprächs dar.

Für den Einstieg in die Diskussion baten wir die Expert:innen um ein fünfminütiges Eingangsstatement zu der Frage: Welchen fachlichen Hintergrund und welche Expertise bringen Sie in dieses Fachgespräch ein und was verbinden Sie mit dem Thema Wirkung bzw. Wirkungsevaluation in Ihrem Arbeits- bzw. Forschungsfeld? Den Auftakt machte Thomas Widmer.

Thomas Widmer: Ich befasse mich mit Evaluation in ganz verschiedenen Zusammenhängen. Dabei habe ich mich mit zwei Themenfeldern besonders intensiv beschäftigt. Die eine Thematik ist die Qualität von Evaluation. Das zweite Thema ist Forschung über Evaluation. Was die *Qualität der Evaluation* angeht, beobachte ich in der Evaluationspraxis in vielen Feldern drei Verhaltensweisen, wenn man damit konfrontiert ist, Wirkungsevaluationen durchzuführen. Die eine Verhaltensweise ist die des Rückzugs, des Verzichts auf die Untersuchung und damit auch den Nachweis von Wirkungen. Man befasst sich vorzugsweise mit Prozessfragen und geht nicht auf die Thematik des Wirkungsnachweises ein. Die zweite Verhaltensweise ist jene der Atomisierung. Dabei wird versucht, die Fragestellung so klein zu machen, dass sie sich auch für einen Wirkungsnachweis eignet. Das heißt bei einer Programmevaluation beispielsweise, dass man das Programm in Mikrointerventionen zerlegt, die sich dann auch eher für einen kausalen, empirischen Wirkungsnachweis eignen. Die dritte Verhaltensweise – leider die häufigste Reaktion – ist die des Scheiterns. Dabei werden wenig robuste Befunde produziert, die oft mit überzogenen Geltungsansprüchen verbunden sind. Selbstverständlich gibt es auch Mischformen der drei Reaktionen: Man kann auch mit dem Atomisieren scheitern oder sich auf Prozessanalysen zu Mikrointervention zurückziehen. Das heißt, die Wirkungsevaluation stellt für die Realisierung einer Evaluation eine sehr anspruchsvolle Herausforderung dar. Und in der Regel sind die Voraussetzungen nicht gegeben, um diese Herausforderung auch mit der nötigen Eleganz, Schönheit und Wahrhaftigkeit zu bewältigen.

Zu meinem zweiten Thema, der Forschung über Evaluation: Hier bewege ich mich vor allem im Kontext von politischen Entscheidungsprozessen. Was dabei auffällt ist, dass wir eine Widersprüchlichkeit haben zwischen der politischen Debatte auf der einen Seite und der vorliegenden, empirischen Evidenz auf der anderen Seite. Das Grundproblem besteht darin, dass in der politischen Debatte häufig aufgrund einer sich verändernden Symptomatik darauf geschlossen wird, ob Maßnahmen wirksam sind oder auch nicht. Das heißt, wenn ein Problem größer wird, so sind die dafür geschaffenen Maßnahmen vermeintlich gescheitert. Ein Beispiel hierfür ist PISA. Man könnte aber auch die Demokratieförderung nennen, wo das Erstarken von zum Beispiel rechtspopulistischen Parteien und Bewegungen als Nachweis dafür gilt, dass Demokratieförderungsmaßnahmen nicht wirksam seien. Diese Argumentation fußt auf zwei Grundproblemen: Das

eine ist, dass man keine Informationen über die kontrafaktische Entwicklung hat. Das heißt, man weiß nicht, wie sich die Entwicklung ohne die entsprechenden Maßnahmen präsentiert hätte. Hätte man keine Maßnahmen ergriffen, hätte es ja noch schlimmer kommen können. Wir haben also keine Möglichkeit eines Policy-on-/Policy-off-Vergleichs. Zum zweiten sind wir in der Regel mit multikausal bedingten Phänomenen konfrontiert und die zu evaluierende Maßnahme ist in der Regel marginal von ihrer Bedeutung in Bezug auf das multikausale Problem bzw. in ihrer „Dosierung“ so gering, dass eine messbare Wirkung kaum erkennbar wird.

Das führt mich zu einer, in der politischen Welt (zumindest auf deklaratorischer Ebene) stark verbreiteten Illusion, dass evidenzbasierte Politikgestaltung funktioniert. Aus verschiedenen Gründen ist das ein Irrglaube. Einmal ist es erkenntnistheoretisch problematisch, weil wir sehr häufig gar keine entsprechende Evidenz vorliegen haben, die den Wirkungsnachweis erbringen kann. Wir haben auch das Problem, dass wir vielfach mit Interventionen konfrontiert sind, die sich einem Zugriff mit experimentellen Methoden („randomized controlled trials“) verschließen, weil sie etwa nicht uniform sind, sondern sich je nach Kontext auch verändern. Das dritte Problem ist ein konzeptionelles und betrifft die Vorstellung darüber, wie politische Entscheidungsprozesse ablaufen. Diese Prozesse erfolgen nicht wie fachliche Entscheidungen, sondern sie sind eben politisch geprägt. Wie dies Carol Weiss betont hat, ist es in der Politik vor allem auch legitim, ideologiegetrieben und/oder interessenbasiert zu entscheiden.¹ Schließlich gibt es auch noch weitere Probleme, die Folgeerscheinungen dieser Bewegung der evidenzbasierten Politik sind. So ist es beispielweise problematisch, dass eine exogene Steuerung der Forschung dadurch erfolgen kann, dass die politische Nachfrage nach entsprechender Evidenz handlungsleitend wird für die Forschenden oder Evaluierenden. Und wir haben auch demokratietheoretische Vorbehalte. Denn streng genommen nähert man sich mit evidenzbasierter Politikgestaltung einer Expertokratie an. Man delegitimiert demokratische Entscheidungsprozesse und setzt ein Primat der fachlichen Entscheidungsprozesse. Das ist demokratietheoretisch problematisch und sollte gerade im Feld der Demokratieförderung spezielle Beachtung finden.

Besten Dank!

Helle Becker: Ich beginne etwas anders als Herr Widmer und erkläre zuerst einmal ganz kurz, was wir eigentlich bei „Transfer für Bildung“ tun. Wir haben es uns zur Aufgabe gemacht, in den Bereichen kulturelle, internationale und politische Bildung zwischen Wissenschaft und Praxis Transferarbeit zu leisten. Der Grund

1 Siehe hierzu Weiss, C. H. (1983): Ideology, interests, and information. The basis of policy positions. In: Callahan, D. / Jennings, B. (Hrsg.): Ethics, the social sciences, and policy analysis. New York: Plenum Press, S. 213–245.

dafür ist, dass es in diesen Bereichen sehr wenig empirische Forschung gibt, vor allen Dingen, wenn man sich den non-formalen Bereich anschaut, also die Jugend- und Erwachsenenbildung. Wir sind eine Art Wissensnetzwerk und sorgen für Vernetzung, Verknüpfung, Überblick, Transfer, begleiten Modellprojekte und machen auch eigene Forschung. Wir haben eine sogenannte Topografie der Praxis der politischen Bildung entwickelt, eine interaktive Karte der verschiedenen Praxisfelder formaler und non-formaler politischer Bildung. Und wir haben unter anderem auch eine umfangreiche Literaturdatenbank, wo wir versuchen, die empirische Literatur zur politischen Bildung abzubilden. Ziel ist immer eine Gesamtdarstellung der Diversität innerhalb der politischen Bildung, und zwar sowohl innerhalb des Wissenschaftsdiskurses als auch in der Praxis. Wir glauben, dass das auch von Nutzen sein kann, um ein Professionsbewusstsein in der politischen Bildung zu unterstützen.

Was sind unsere Erfahrungen mit Evaluation und Wirkungsforschung? Neben unseren eigenen Forschungen, die auch in diese Richtung gehen, haben wir 35 Jahre Erfahrung im Bereich der non-formalen politischen Bildung. Ich habe 2011 eine Literaturstudie über die davorliegenden zehn Jahre durchgeführt, also bis zum Jahr 2000 zurückgehend, und unter anderem untersucht, welche Studien zur Wirkungsforschung es zur politischen Bildung gibt. Für diese zehn Jahre habe ich insgesamt einhundert empirische Studien gefunden. Im Vergleich dazu sind im Bereich der Konzeptbildung und der Theorie vierhundert bis fünfhundert Titel pro Jahr veröffentlicht worden. Daran kann man sehen, wie wenig die non-formale politische Bildung im Fokus empirischer Forschung steht. Dass es da so wenig gibt, hat zum einen mit der Wissenschaftslandschaft zu tun. Auf der anderen Seite aber auch damit, dass es kaum finanzielle Unterstützung für entsprechende Studien gibt und dass die Studien, die vorhanden sind, entweder von großen und sehr stark interessegeleiteten Stiftungen oder Ministerien finanziert worden sind. Es gab zwei große Evaluationen, die das Praxisfeld geradezu traumatisiert haben, mit Auswirkungen bis heute. Das waren zum einen die Evaluationsstudie von Achim Schröder zur Jugendbildung und zum anderen die von Lothar Böhnisch zur Erwachsenenbildung.² Beide führten zum bisher letzten großen Aufschrei innerhalb der Fachpraxis. Man hat die Entwicklung hin zu diesen Evaluationen, die Fragestellung, das Forschungsdesign einer sehr, sehr kritischen Rezeption unterzogen. Die Studien waren von BMBF und BMFSFJ in Auftrag gegeben worden, mit einer ihnen entsprechenden Fragestellung. Die Praxis fühlte sich nicht adäquat befragt und kritisierte dann auch etwas überzogen, dass es hier nicht darum gehen könne, so eine Art Gesinnungsprüfung durch Wirkungsfors-

2 Gemeint sind: Schröder, A. / Balzter, N. / Schroedter, T. (2004): Politische Jugendbildung auf dem Prüfstand. Weinheim und München: Juventa; sowie Fritz, K. / Maier, K. / Böhnisch, L. (2006): Politische Erwachsenenbildung. Trendbericht zur empirischen Wirklichkeit der politischen Bildungsarbeit in Deutschland. Weinheim und München: Juventa.

sung einzuführen. Den Studien wurde aus fachlicher Sicht ein zweifacher Vorwurf gemacht: Wie kann man non-formale politische Bildung überhaupt auf ihre Wirkungen hin befragen? Und wo kommen eigentlich unsere Ideen zu Wirkungserwartungen her, wenn sie nicht von der Politik gesetzt werden? Es gab damals kleine Ansätze, die Frage zu beantworten, wie denn adäquate Forschungsdesigns aussehen könnten. Es gab zum Beispiel ein Modell partizipativer Forschung und erstmalig Ansätze einer Praxisforschung. Es hat seitdem, bis auf die letzten Jahre und wieder über die Politik, also wieder von außen angestoßen, kaum Initiativen gegeben, diesen ganzen Fragekomplex weiter zu behandeln. Aus unserer Sicht fehlt es aber nicht nur an Wirkungsforschung und Evaluation, es fehlt vor allen Dingen an Praxis- und Feldforschung. Diese müsste eigentlich erst einmal vorgeschaltet sein, weil wir es in diesem Feld mit teilweise völlig überzogenen oder von außen herangetragenem Wirkungserwartungen zu tun haben. Beispielsweise wird die politische Bildung immer wieder adressiert, sie möge dafür sorgen, dass es eine höhere Wahlbeteiligung gibt, dass mehr Menschen in Parteien eintreten und dass das Vertrauen in den Staat bzw. die Demokratie gestärkt wird. Das sind Wirkungserwartungen, die sich die non-formale Praxis der politischen Jugend- und Erwachsenenbildung selbst nicht vorrangig auf die Fahne schreiben würde.

Auch die Entwicklungen im Rahmen des Forschungsverbundprojekts PrEval³ haben in der Szene der politischen Bildung für große Aufregung gesorgt und auch die Evaluation von „Demokratie leben!“⁴ Wir als Transferstelle haben versucht, diese Aufregung produktiv zu nutzen und noch einmal betont, dass wir selbst uns darüber Gedanken machen müssen, was wir als adäquate Fragestellungen, adäquate Forschung zur non-formalen politischen Bildung verstehen. Das ist nicht ganz einfach, würde ich als Erziehungs- und Bildungswissenschaftlerin sagen. Es gibt aber auch Betrachtungsweisen, die das anders sehen, die zum Beispiel sehr stark politikdidaktisch oder politikwissenschaftlich motiviert sind.

Wir haben es mit einer Skepsis zu tun aus der Trägerlandschaft, die, deswegen habe ich das gerade so dramatisch „traumatisiert“ genannt, natürlich aufgrund ihrer Erfahrungen Evaluation als eine Fremdkontrolle empfindet, obwohl, wenn

3 Das Verbundprojekt PrEval (Zukunftswerkstätten Evaluation und Qualitätssicherung in der Extremismusprävention, Demokratieförderung und politischen Bildung: Analyse, Monitoring, Dialog) versteht sich als Forschungs- und Transfervorhaben und entwickelt Formate sowie Strukturen zur Stärkung von Evaluation und Qualitätssicherung in der Extremismusprävention, Demokratieförderung und politischen Bildung in Deutschland weiter. Es wird vom Bundesministerium des Innern und für Heimat (BMI) mit einer Laufzeit von Oktober 2022 bis September 2025 gefördert (preval.hsfk.de, Abruf 16.05.2024).

4 Das Bundesprogramm „Demokratie leben!“ wird von einem Evaluationsverbund wissenschaftlich begleitet. Informationen zur Evaluation der Förderperiode 2020 bis 2024 siehe <https://www.demokratie-leben.de/das-programm/programmevaluation> (Abruf 16.05.2024) und in der Einleitung zu diesem Band. Informationen zur Evaluation der Förderperiode 2015 bis 2019 siehe <https://www.demokratie-leben.de/das-programm/foerderperiode-2015-2019> (Abruf 16.05.2024).

man nachfragt, auch ein großes Interesse daran besteht, über die eigene Praxis aufgeklärt zu werden, vor allen Dingen über Bedingungsgefüge, die es da gibt. Also: Man hätte ganz viel zu tun. Wir haben das als Transferstelle immer wieder auf die Agenda gesetzt und dann im letzten Sommer mit der Bundeszentrale für politische Bildung einen allerersten kleinen Aufschlag gemacht. Wir haben Personen aus unterschiedlichen Feldern der Wissenschaft angesprochen und an einen Tisch gebracht, die aus ganz unterschiedlichen Perspektiven empirisch zu politischer Bildung forschen. Ich lasse jetzt mal beiseite, dass das vielleicht auch Demokratiepädagogik, Demokratiebildung et cetera heißen kann. Und wir waren doch über die große Resonanz überrascht. Es war das allererste Mal, dass unterschiedliche Wissenschaftler:innen aus verschiedenen Diskurszusammenhängen zu diesem Thema gemeinsam an einem Tisch saßen.

Wir haben daraus jetzt die Schlussfolgerung gezogen, zum Thema Feldforschung – im weitesten Sinne Evaluation und Wirkungsforschung – einen eigenen Wissens-Hub aufzubauen. Da sind wir gerade dabei. Deswegen interessiert uns das Thema besonders. Wir glauben, dass wir da, wenn wir vierteiliger und auch mit einer anderen Perspektive und mit einer eigenen Agenda schauen, erfolgreich sein können.

Jörg Faust: Ich möchte mich der Eingangsfrage aus drei Perspektiven nähern: aus einer institutionellen, einer thematischen und einer methodischen Perspektive. Die institutionelle Perspektive stammt aus meiner Tätigkeit in einem Politikfeld, das durch eine besondere Institution, das DEval, geprägt ist. In diesem Kontext durfte ich in den letzten Jahren erleben und mitgestalten, wie wichtig es für die Nützlichkeit von (Wirkungs-)Evaluation ist, in einem angemessenen institutionellen Umfeld zu arbeiten. Aus dieser Arbeitserfahrung und dem Blick auf andere Politikfelder sehe ich eine Vorreiterrolle der Entwicklungszusammenarbeit bei der Institutionalisierung von Evaluierung in einem Politikfeld. Denn hier wurde über die vergangenen Jahrzehnte ein vergleichsweise fortschrittliches Evaluierungssystem aufgebaut, das wichtige Elemente der Evaluation gut verregelt hat. Dies betrifft sowohl die nationale Ebene hier in Deutschland als auch die internationale Ebene im Rahmen der OECD.

Um auf Dauer mit Evaluierungen tatsächlich Wirkungen zu erzeugen bzw. deren Kernfunktionen der Rechenschaftslegung und des Lernens strukturell zu erfüllen und vielleicht – zumindest in begrenztem Maße – damit zu evidenzbasierter Politikgestaltung beizutragen, bedarf es der Institutionalisierung: Dazu gehört die Bereitstellung der notwendigen finanziellen und personellen Mittel, die Festlegung von Evaluierungsmandaten und die Festschreibung von übergeordneten Prinzipien und Standards guter Evaluation, insbesondere was die Kernprinzipien Unabhängigkeit, Nachvollziehbarkeit und Nützlichkeit angeht. Zudem sollten die Kernelemente von Evaluationsprozessen kodifiziert werden sowie Fragen der Veröffentlichung und des Umsetzungsmonitorings. Letzteres betrifft die

Frage, wie die Umsetzung der ausgesprochenen Empfehlungen verfolgt wird. Mit seinen 2021 veröffentlichten Evaluierungsleitlinien hat das Entwicklungsministerium (BMZ) hier ein im Vergleich zu den meisten anderen Ministerien sehr fortschrittliches Regelwerk vorgelegt. Gerade vor dem Hintergrund der Diskussionen um wirkungsorientierte Haushaltsführung könnten andere Ministerien davon lernen.

Zu meiner zweiten, der thematischen Perspektive: Auch in der Entwicklungszusammenarbeit spielt Demokratieförderung eine größere, gegenwärtig vermutlich wieder steigende Rolle. Es gehört zu den Zielsetzungen vieler Geberländer in der Entwicklungszusammenarbeit, demokratische Prozesse in Entwicklungs- und Schwellenländern zu stärken. Zu diesem Zwecke werden eine ganze Reihe von Maßnahmen finanziert, beispielsweise die Aktivitäten der politischen Stiftungen im Ausland, Wahlbeobachtung in Entwicklungs- und Schwellenländern oder Maßnahmen der politischen Bildung oder Maßnahmen des Capacity Building in Parlamenten. Auch wurden unter anderem Mittel bereitgestellt für partizipative Haushaltsführung auf kommunaler Ebene. Auch hierbei stellt sich natürlich die Frage, wie die Wirkungen dieser Maßnahmen identifiziert werden können. Dies führt mich zu der methodischen Perspektive, wo ich für die Stärkung quantitativer Verfahren der Wirkungsmessung plädiere. In dem Themenfeld der „Demokratiehilfe“ mit Mitteln der Entwicklungszusammenarbeit gibt es auf der Makroebene den Vorteil, dass länderübergreifend über die OECD erfasst wird, wie viele Gelder für Demokratieförderung eingesetzt werden. Dadurch sind Evaluierung und Forschung in der Lage, länderübergreifende statistische Wirkungsuntersuchungen durchzuführen. Auf der Mikroebene einzelner Projekte ist es ebenfalls anspruchsvoll mit der Wirkungsattribution. Aber auch dort beobachten wir in den letzten zwei Jahrzehnten eine Zunahme quasi-experimenteller und in einigen Fällen auch experimenteller Untersuchungen, die der Frage nachgehen, welche Wirkungen solche Interventionen, etwa im Bereich der politischen Bildung oder im Bereich der Wahlbeobachtung nach sich ziehen. Die auf den Bereich der rigorosen Wirkungsevaluierung in der Entwicklungszusammenarbeit spezialisierte Organisation 3ie hat hierzu kürzlich einen Überblick über rigoros evaluierte Maßnahmen gegen „democratic backsliding“ vorgelegt, der auch für die hiesige Diskussion von Interesse sein dürfte.⁵ Gleichzeitig bedeutet die Stärkung quantitativer Verfahren nicht den Verzicht auf qualitative Elemente. In hochwertigen Evaluationen sind quantitative Verfahren mit qualitativen kombiniert, weil gemeinhin ein guter Mix aus qualitativen und quantitativen Verfahren die besten Ergebnisse zeitigt.

5 Für weiterführende Informationen siehe <https://www.3ieimpact.org> (Abruf 01.07.2024).

Björn Milbradt: Ich würde eingangs kurz etwas zur Perspektive sagen, die wir hier am Standort Halle des Deutschen Jugendinstituts vertreten, dann etwas zur Entwicklung der Frage nach Wirkungsuntersuchung in unseren Feldern und ein, zwei Einzelaspekte herausgreifen. Ich habe heute noch einmal in einige der vorliegenden Sammelbände hineingeschaut, die sich um die Evaluation von Maßnahmen und Programmen der Demokratieförderung und Extremismusprävention drehen und bin dabei auf einen Aufsatz von Andreas Hirsland, Oliver Dimbath, Julia von Hayek und Werner Schneider aus dem Jahr 2004⁶ gestoßen, der sich um die Wirkungsevaluation eines Seminarprogramms mit dem Titel „Achtung plus Toleranz“ dreht. Und die Autor:innen haben bereits damals eine Fragestellung aufgeworfen, die uns teilweise immer noch bewegt. Sie haben schlicht und einfach gefragt, wie die Entwicklung von Toleranz zu evaluieren ist und wie auch Wirkung zu evaluieren ist und sind da natürlich auf die ganzen Fallstricke gekommen, die man bei solchen komplexen Konzepten hat. Zum Beispiel kritisieren sie die bloße Abfrage von kanonisiertem Faktenwissen und schlagen stattdessen ein eher rekonstruktives Evaluationsdesign vor. Auch wird hier ein sehr kritischer Blick auf das Thema Wirkungsevaluationen geworfen. Gleichzeitig sieht man an den Publikationen, die seitdem entstanden sind, dass der Wissenszuwachs in dem Bereich sehr groß ist und inzwischen relativ viel vorliegt. Trotzdem habe ich teilweise den Eindruck, dass bestimmte Diskussionen nicht weitergeführt wurden.

In der Fachgruppe „Politische Sozialisation und Demokratieförderung“ am Deutschen Jugendinstitut in Halle versuchen wir immer auch die pädagogische Fachpraxis in ihrer Eigenlogik zu verstehen, wenn wir sie evaluieren. Das bedeutet meiner Ansicht nach, Evaluationsdesigns gegenstandsangemessen anzulegen, das heißt so, dass man sie nicht abstrakt dem Evaluationsgegenstand überstülpt. Das ist, glaube ich, ein relativ zentraler Punkt und bedingt überhaupt noch keine Priorisierung von quantitativen oder qualitativen Designs, sondern beides hat gleichermaßen seine Berechtigung in Wirkungsstudien. Ich würde sagen, dass Gegenstandsangemessenheit eine sehr wichtige Voraussetzung auch für Wirkungsevaluation ist. Ich bin seit 2016 mit diesen Themen befasst und ich beobachte seitdem eine starke Ausweitung der Frage nach Wirkung, ein Stück weit vorangetrieben durch Ministerien. Das ist teilweise mit einer starken Forderung nach sehr eingängigen, eindimensionalen und leicht verwertbaren Kennzahlen verbunden. Das stellt, wie in den Statements meiner Vorgänger:innen schon deutlich geworden ist, ein Stück weit eine Schwierigkeit dar. Gleichzeitig nehme ich wahr, dass Stück für Stück Wirkungsevaluationen entstehen. Bei so einem großen Programm wie „Demokratie leben!“ ist das

6 Hirsland, A./Dimbath, O./Hayek, J. von/Schneider, W. (2004): Evaluation des Programms „Achtung (+) Toleranz“ – ein Praxisbericht. In: Uhl, K./Ulrich, S./Wenzel, F. M. (2004): Evaluation politischer Bildung. Ist Wirkung messbar? Gütersloh: Bertelsmann Stiftung, S. 57–82.

eine Herausforderung. Natürlich ist es auch für einzelne Projekte eine Herausforderung. Für große Programme aber umso mehr, weil wir es mit einem Mehr-Ebenen-Programm zu tun haben, was alle föderalen Ebenen umfasst, was teilweise einen Innovationsauftrag hat, teilweise einen Vernetzungsauftrag, teilweise einen Bildungsauftrag und einen Förderauftrag. Und insofern muss man da jeweils ganz unterschiedliche methodische Designs anwenden.

Mittlerweile führen wir ebenfalls Wirkungsforschung durch. Die Herausforderungen, beispielsweise bei der Ausstiegs- und Distanzierungsberatung, sind dabei denen von Bildungsprozessen ähnlich, würde ich sagen. Wir haben es hierbei mit langfristigen, nicht standardisierbaren Prozessen zu tun, wenn Menschen gewaltbereite, extremistische Szenen verlassen. Wir fangen da ganz unterschiedliche Perspektiven ein, oft die von Beratenden, von Adressatinnen und Adressaten oder teilweise eben auch von externen Stakeholdern, um beurteilen zu können, wie sich Wirkungen herstellen (siehe Beitrag von Faglestahler/Haase/Schau in diesem Band). Zusammenfassend würde ich sagen, sehen wir eine positive Entwicklung. Zugleich müssen wir aber auch kritisch auf eine starke Wirkungsorientierung bzw. die Anforderungen blicken, die teilweise an Wirkungsevaluationen herangetragen werden. In methodischer Hinsicht plädiere ich für Mixed Methods Designs und Gegenstandsangemessenheit in der Wirkungsevaluation.

Hemma Mayrhofer: Ich setze seit ungefähr 20 Jahren Evaluationsforschungen um, seit zehn Jahren beschäftige ich mich mit der Konzeption, methodischen Umsetzung und Leitung von Projekten der Wirkungsevaluation in der Sozialen Arbeit. Aufbauend auf meiner Dissertation, einer soziologischen Grundlagenforschung zum Thema Niederschwelligkeit in der Sozialen Arbeit, beschäftigt mich in verschiedenen Forschungsschwerpunkten die Frage: Was heißt es eigentlich für niederschwellig arbeitende Tätigkeitsfelder in der Sozialen Arbeit und in angrenzenden Feldern, wenn an sie vermehrt das Ansinnen herangetragen wird, ihre Wirksamkeit, also die Wirksamkeit ihrer Interventionen nachzuweisen und damit auch die Legitimität des Ressourceneinsatzes zu rechtfertigen? Was bedeutet das für solche Felder, die sehr niederschwellig arbeiten, die ganzheitlich Problemfelder bearbeiten, die situativ arbeiten, unverbindliche Angebote offerieren und so weiter? Diese Bereiche können wesentlich schwerer ihre Wirkungen nachweisen. Wie ist in so einem Feld Wirkungsevaluation möglich? Um mich näher mit diesen Fragen auseinandersetzen zu können, beantragte ich eine Forschungsförderung für ein Projekt, das sich mit Wirkungsevaluation mobiler Jugendarbeit beschäftigen sollte. Und diese Forschung prägt mich bis heute. Wir hatten sehr günstige Bedingungen für das Forschungsprojekt: Es gab keinen direkten Auftraggeber. Das heißt, wir konnten unsere Forschungsfragen selbst formulieren, unsere Zugänge und Methoden ganz uneingeschränkt auswählen. Und wir hatten in Bezug auf die verfügbaren Ressourcen relativ günstige Bedingungen. Sprich, wir haben uns zum Beispiel am Anfang eine explorative Phase

gegönnt, wo wir überlegt haben: Wie könnten denn Wirkzusammenhänge, Wirkweisen, Wirkungsziele, Dimensionen und Indikatoren in diesem Tätigkeitsfeld überhaupt beschaffen sein? Wie könnten sie zu fassen sein? Und dazu haben wir uns nicht nur den Forschungsstand zum Thema und Selbstbeschreibungen der Institutionen wie Projektconcept und Ähnliches angeschaut, sondern auch explorative Gespräche, Gruppendiskussionen geführt mit Jugendarbeiter:innen. Und auf dieser Basis haben wir erst verschiedene Methoden der Erforschung von Wirkungen konkretisiert und dann im Zusammenspiel verschiedener Perspektiven, methodischer Ansätze und Vorgehensweisen Wirkungen erforscht. Das war ein multiperspektivisches und auch ein methodenplurales Vorgehen, in dem wir verschiedene quantitative und verschiedene qualitative Methoden integriert haben. Außerdem haben wir die Fachpraxis integriert. Wir standen mit Praxispartner:innen in engem Austausch. Das passierte bis in die Disseminationsphase hinein, wo es um die Frage ging, was bedeuten diese Ergebnisse überhaupt für die Praxis. Das prägt mich bis heute, diese Vorgehensweise und diese Erfahrungen, die ich bzw. die wir in diesem Zusammenhang machen durften.

Ich würde jetzt gern zwei, drei Aspekte kurz einbringen, die für mich in meiner Auseinandersetzung mit dem Thema bis heute sehr wichtig sind. Der eine Punkt sind die Ziele der Einrichtungen bzw. der Programme, der Projekte, der Maßnahmen. Da hatten wir in unserem Forschungsprojekt beispielsweise die Möglichkeit, sehr genau hinzuschauen: Welche Ziele gibt sich mobile Jugendarbeit selbst? Was bedeutet das dann in der Folge auch für erzielbare Wirkungen und für die Wirkungsevaluation? Im Bereich sozialer Interventionen hatten wir es mit einem Forschungsfeld zu tun, an das recht verschiedene Zielsetzungen und Wirkungserwartungen herangetragen werden. Sowohl die Praxisakteur:innen setzen sich selbst bestimmte Ziele als auch die Politik, die Fördermittelgeber:innen, die Adressat:innen, alle tragen verschiedene Ziele an die Praxisakteur:innen heran. Diese Ziele sind oft nicht „SMART“. Sie sind nicht immer spezifisch und sie sind schwer messbar. In der Regel sind sie zumindest teilweise akzeptiert, wenngleich manchmal unrealistisch und auch nicht immer terminierbar. Und das aus gutem Grund! Das ist dem spezifischen Arbeitsfeld und seinem Vorgehen geschuldet. Hier lassen sich Ziele oft weder überindividuell konkretisieren noch können sie bei einer stark flexiblen, Adressat:innen-orientierten Arbeitsweise gut im Vorhinein festgelegt werden. Zugleich, das greift, denke ich, auch einige Aspekte meiner Vorredner:innen auf, ist es wichtig zu beachten, dass die Einrichtungen mit den verschiedenen, zum Teil widersprüchlichen Zielsetzungen umgehen müssen. Eine Möglichkeit, damit zurechtzukommen, ist, solche Ziele bzw. die Zielfolgen latent zu halten, und zwar dadurch, dass ich sie möglichst allgemein benenne und nicht zu konkret mache. Denn je stärker ich es konkretisiere, desto stärker treten Widersprüche zutage. Dessen muss sich Evaluationsforschung, Wirkungsevaluation im Speziellen, die nicht nur zeremoniell ist, nicht nur zum Fassadenmanagement dient, bewusst sein, damit sie dann

tatsächlich auch zu aussagekräftigen Ergebnissen kommt in Bezug darauf, was auf der Aktivitätsebene passiert und was damit erreicht werden kann.

Ein zweiter Punkt, der mir wichtig ist, ist die Methodenfrage. Das ist ein extrem bedeutsamer und zugleich herausfordernder Aspekt in der Wirkungsevaluation. Die zentrale Frage ist: Wie kann in der Wirkungsevaluation mit der enormen Komplexität nicht nur des Untersuchungsgegenstands, sondern auch der Aufgabe, Ursache-Wirkungs-Zusammenhänge zu erschließen, in einer Weise umgegangen werden, die die Komplexität bewältigbar macht und sie trotzdem nicht unangemessen reduziert? Und diese Herausforderung, das wurde schon mehrfach angesprochen, kann meiner Erfahrung nach in besonderer Weise in einem methodenpluralen Forschungsdesign bzw. mit Mixed Methods Ansätzen umgesetzt werden, und zwar in solchen Designs, die auch elaborierte, qualitative Forschungsansätze integrieren.

In unsere Forschung hatten wir ein ethnografisches Vorgehen integriert, das unter anderem teilnehmende Beobachtung vorsah und ethnografische Interviews. Damit begleiteten wir Projekte, Ansätze und Interventionen auch über eine längere Phase. Zudem waren uns biografische Fallrekonstruktionen, also die retrospektive Rekonstruktion von Deutungen und Erzählungen, möglich. Damit lassen sich individuelle Verarbeitungsweisen von Ereignissen und Interventionen rekonstruieren und auch Wirkweisen auf individueller Basis erschließen, die zugleich eingebettet sind in den komplexen lebensgeschichtlichen und lebensweltlichen Kontext der Biograf:innen. Mit qualitativen Methoden lässt sich besonders gut erschließen, wie etwas wirkt, wie sich komplexe Wirkzusammenhänge gestalten können, was womit wie zusammenwirkt. Der dritte Punkt, den ich auch für sehr, sehr wichtig halte, aber auf den ich später noch mal gesondert eingehe, ist der Aspekt des Ergebnistransfers, der kooperativen Wissensbildung mit der Praxis.

Herausgeber:innen: Wir haben nun fünferte Statements mit Überlegungen zum Thema Wirkungsuntersuchungen gehört. Herzlichen Dank dafür! Wer von Ihnen möchte auf einen oder mehrere seiner Vorredner:innen eingehen? Welche Überlegungen wollen Sie aufgreifen, was möchten Sie ergänzen oder wo vielleicht auch widersprechen?

Jörg Faust: Ich stimme Herrn Widmer zu, dass wir nicht der Illusion verfallen sollten, mit Wirkungsevaluation evidenzbasierte Politikgestaltung im Sinne einer Expertokratie zu etablieren. Dies sollte auch gar nicht das Ziel sein in einer Demokratie, die ihre Performanzüberlegenheit ja aus einem offenen und inklusiven Wettbewerbsprozess zieht, in dem auch Raum für ideologisch und politisch motivierte Entscheidungen besteht. Gleichzeitig bin ich überzeugt, dass Evidenz, die über unabhängige Evaluation an Politik und an politische Entscheider:innen herangetragen wird, deren Politikgestaltung zum Besseren beeinflussen kann. Zwar

wird solche Evidenz sicherlich nicht die ideologischen Grundhaltungen demokratisch-legitimierter Regierungen verändern, wohl aber wird sie in Bezug auf Instrumentenwahl und Instrumentengestaltung auf Dauer eine positive Wirkung haben. Insofern würde ich auf der einen Seite keine überzogenen Erwartungen an Wirkungsevaluation formulieren, auf der anderen Seite bin ich davon überzeugt, dass im Habermas'schen Sinne zumindest demokratisch-legitimierte Entscheidungstragende offen für einen fach- und evidenzbasierten, deliberativen Diskurs sind. Voraussetzung dafür, dass Evaluation so wirkt, ist allerdings eine gute institutionelle Aufstellung der Evaluierung, sprich die Existenz eines guten Evaluierungssystems.

Thomas Widmer: Die Problematik der evidenzbasierten Politikgestaltung liegt ja vor allem auch darin, dass die Politik in der Regel entsprechende Evidenz nicht nur nach Güte auswählt, sondern diese auch nach politischer Opportunität selektioniert, wie dies auch schon früh die bereits erwähnte Carol Weiss zusammen mit dem Kollegen dargelegt hatte.⁷ Das ist das Vorgehen des politisch denkenden Menschen. Ich möchte das nicht abwerten, sondern ich konstatiere das lediglich. Und das führt dazu, dass politische Akteure die ihnen politisch opportun erscheinenden Evidenzen nutzen, um die eigene Entscheidung zu legitimieren. Sehr häufig werden dann auch Evidenzen beigezogen, die nicht von höchstehender Qualität oder möglicherweise sogar falsch sind. Und es ist auch so, dass die Politik die vorliegende Evidenz häufig im eigenen Sinne interpretiert, das heißt, nicht nur selektioniert, sondern die Befunde teilweise auch auf die entsprechenden persönlichen Präferenzen anpasst. Und diese Phänomene sind meines Erachtens problematisch. Sie führen dazu, dass vorliegende Evidenz, die teilweise, ich möchte das gar nicht infrage stellen, robust und aussagekräftig und durchaus auch entscheidungsrelevant ist, nicht oder nur partiell in den Evidenzfundus einfließt. Stattdessen wird dieser Fundus vielfach durch Evidenzen minderer Güte versorgt oder durch Evidenz, die gar nicht geschaffen ist, um entsprechende Aussagen zu tätigen. Die im Konzept der evidenzbasierten Politikgestaltung enthaltene Mechanik vermittelt ein verzerrtes Bild politischer Entscheidungsfindung.⁸

Helle Becker: Ich würde da gerne mit zwei kritischen Anmerkungen dezidiert aus der Sicht der politischen Bildung einhaken, als einem Bereich, um den herum sich in letzter Zeit sehr viel tut. Damit meine ich den Bereich der Demokratieförderung im weitesten Sinne, der aber nicht unbedingt politische Bildung und vor allen Dingen nicht unbedingt Bildungsprozesse betrifft. Während für Demokratieförderung allgemein die Aufmerksamkeit hoch ist, kann man andererseits

7 Siehe Weiss, C. H. / Bucuvalas, M. J. (1980): Truth tests and utility tests: Decision-makers frames of reference for social science research. *American Sociological Review*, 45. Jg., H. 2, S. 302–313.

8 Dazu lesenswert: Pawson, R. (2006): *Evidence-based Policy. A Realist perspective*. London: Sage.

– wie Thomas Krüger [Präsident der Bundeszentrale für politische Bildung; Anmerkung d. Hrsg.] es nannte – von einer „Verzweigung“ der politischen Bildung sprechen. Damit gerät der politische Bildungsbereich, der ohnehin schon wissenschaftlich marginalisiert ist, wie ich eingangs versucht habe zu verdeutlichen, noch mehr ins Hintertreffen, steht faktisch mit dem Rücken an der Wand. Vor diesem Hintergrund will ich noch einmal auf zwei Punkte hinweisen, mit denen die Praxis und die Wissenschaft konfrontiert sind, wenn es beispielsweise um die Finanzierung von Forschungsprojekten zu politischer Bildung geht: Da werden nicht nur inhaltliche Wirkungserwartungen – ganz real aus Politik- und Ministeriumsrunden – an uns als Forschende herangetragen, sondern es werden auch Erwartungen an das Bedingungsgefüge oder an die Wirkzusammenhänge formuliert. Mit anderen Worten: Für die politische Bildung ist es angesichts dieser Erwartungen schwierig, zu argumentieren. Sie hat vor allem ein Problem damit, zu vermitteln, dass politische Bildung wirkt, aber diese Wirkungen nicht zielgerichtet herbeizuführen sind. Das ist die eine große Schwierigkeit. Eine weitere Herausforderung ist, dass die Wirkungsfantasien – bezogen auf die Praxis – ins Kraut schießen, bildlich gesprochen, solange es keine solide Praxisforschung gibt. Außerdem ist aus Sicht der politischen Bildung im Moment die Gefahr groß, dass diejenigen Forschungsdesigns und Forschungsstudien herangezogen werden, die aus Sicht der Politik, also nicht aus Sicht der Wissenschaft, besondere Evidenz versprechen. Aus Sicht der Politik wird vor allen Dingen quantitativen Studien eine größere Evidenz zugesprochen als qualitativen. Je standardisierter die untersuchten Settings sind, umso einfacher ist es, quantitativ zu forschen und für die Politik, daraus greifbare Ergebnisse abzuleiten. Damit kann dann wieder ein stärkerer Wirkungszusammenhang behauptet werden, der scheinbar wiederholbar und damit belastbar ist. In der non-formalen politischen Bildung haben wir es jedoch mit einer ganzen Reihe von sehr unterschiedlichen Settings zu tun. Frau Mayrhofer hatte das eben in ihren Forschungserfahrungen zur Jugendarbeit angesprochen. Und wenn wir uns anschauen, wie die offene Jugendarbeit oder offenere Settings von politischer Jugendbildung arbeiten, dann ergeben sich wissenschaftstheoretische und forschungspraktische Herausforderungen für Wirkungsuntersuchungen. Daher noch einmal mein Plädoyer: Solange es keine Praxisforschung gibt, die überhaupt erst einmal einen Einblick gibt oder vielleicht auch eine Systematisierung vornimmt, welche unterschiedlichen Formen und Settings und Wirkzusammenhänge es in der politischen Bildung gibt, so lange bleibt – zugespitzt formuliert – zumindest die Praxis und eben oft auch die Forschung den eingangs genannten Anforderungen vonseiten der Politik relativ hilflos ausgesetzt.

Björn Milbradt: Ich möchte hier direkt anschließen und bringe noch einmal das Beispiel der Ausstiegs- und Distanzierungsarbeit und Beratung an, weil das mir nah an jenen Fragen zu sein scheint, die auch in der politischen Bildung eine Rol-

le spielen. Will man junge Menschen oder Menschen überhaupt aus extremistischen Szenen bzw. Gruppen herausholen, so handelt es sich hierbei um langfristige Prozesse, die einerseits biografische Distanzierungsprozesse sind und andererseits natürlich Bildungsprozesse, in denen alternative Weltansichten oder Deutungsmuster langfristig aufgebaut werden müssen. Und es gibt häufig eine wenig angemessene Vorstellung von solchen Distanzierungsprozessen. Da geht man dann davon aus, dass man mit den Leuten ideologiekritische oder ideologiebezogene Workshops macht und dann nach ein paar Monaten deradikalisierte Jugendliche vorfindet. Und jeder sozialpädagogische Praktiker, jeder Ausstiegsberater weiß, dass das so nicht funktioniert. Aber es entspricht der Wunschvorstellung von Politik, dass es so funktionieren würde.

Im Rahmen der wissenschaftlichen Begleitung der Ausstiegs- und Distanzierungsarbeit im Bundesprogramm „Demokratie leben!“ haben wir in diesem Feld auch Wirkungsuntersuchungen durchgeführt. Dort trifft man dann auf Schwierigkeiten, die beim Feldzugang anfangen. In einem hochsensiblen Bereich wie diesem ist es relativ schwer, Pädagoginnen und Pädagogen, aber auch Klient:innen zu finden, die da überhaupt mitmachen. Große Stichproben wird man dort eh nicht zusammenbekommen. Dann ist es so, dass die Beteiligten, wenn man sie dann befragen kann, ganz unterschiedliche Perspektiven auf Wirkungen haben. Die Berater:innen beispielsweise schreiben die Wirkung solcher Distanzierungsprozesse teilweise den Klient:innen zu, die aus sich heraus diesen Prozess gestalten würden und denen die Beratung nur den Anstoß gibt. Die Klientinnen und Klienten wiederum schreiben die Wirkung vor allem den Berater:innen zu. Insofern ergeben sich da schon zwei unterschiedliche Sichtweisen darauf, wie Wirkungen zustande kommen. Das ist für sich genommen erst einmal unproblematisch. Wir sehen hier, dass Wirkungen immer auch ein Stück weit perspektivabhängig sind.

Daher ist es nicht so einfach, solche Beratungsprozesse eindeutig auf einen bestimmten Wirkmechanismus zu reduzieren. Wir haben hier tatsächlich mehrere Mechanismen rekonstruieren können, beispielsweise, dass es erst einmal eine vertrauensvolle Beratungsbeziehung braucht und dass deren Zustandekommen schon für sich genommen ein langwieriger Prozess ist. So ein Beratungsprozess dreht sich teilweise längere Zeit erst einmal nicht um Politik oder Rechtsextremismus oder damit verbundene Weltansichten, sondern zunächst bauen Berater:innen Vertrauen auf. Zudem sind solche Beratungsprozesse hoch individuell, das heißt kaum standardisierbar. Wenn man sie manualisieren wollte, würde man sie in einer Weise überformen, die ihnen nicht entspräche. Und das ist einfach eine Herausforderung. Ich sage das nicht, um Wirkungsforschung oder Evidenzbasierung aus diesem Feld herausnehmen zu wollen, sondern um die Herausforderung darzustellen. Ich wünsche mir mehr Verständnis für solche komplexen Prozesse und die damit verbundenen Herausforderungen zur Erlangung von Evidenz.

Hemma Mayrhofer: Ich möchte noch ein paar Gedanken zu den vielen spannenden Eindrücken eben beisteuern. Wir haben in unserem Institut für angewandte Rechts- und Kriminalsoziologie auch viel mit Politik und mit politischer Verwaltung auf Bundes-, Landes- und kommunaler Ebene zu tun. Und diese Zusammenarbeit ist und bleibt herausfordernd, gerade auch zu Themenbereichen wie Sicherheit, Kriminalität, Extremismus und so weiter. Das sind alles polarisierende gesellschaftlichen Themen, die Forschung herausfordern. Dabei sollten idealerweise nicht die Legitimationsinteressen von Politik und Verwaltung im Vordergrund stehen, sondern Lern- und Entwicklungsfacetten in den Blick kommen bzw. gestärkt werden. Politik ist immer auch daran interessiert, wiedergewählt zu werden, und insofern ist es völlig logisch, dass sie nicht nur fach- und evidenzbasiert entscheiden kann, sondern manchmal auch machtpolitisch entscheiden muss. Uns beschäftigt vor diesem Hintergrund immer die Frage: Was können wir tun, um unsere Forschungserkenntnisse möglichst gut an die Politik und Praxis zu vermitteln? Und da erweist es sich meiner Meinung nach als sinnvoll, von vornherein mit zu bedenken: Wie kann ich Formate der kooperativen Wissensbildung bereits in der Endphase eines Forschungsprojekts implementieren? Wie kann ich etwa in Workshops gemeinsam erarbeiten, was von diesen Erkenntnissen hilfreich und nützlich ist? Wie kann ich mit Blick auf politische Entscheidungen, etwa über Enqueten⁹ oder ähnliche Formate, auch auf einer breiteren Ebene wirken? Wie kann ich im Arbeitsbündnis mit den untersuchten Projekten und Einrichtungen nicht nur Legitimationserfordernisse in den Vordergrund stellen, sondern auch gemeinsames Lernen und Erweitern von Wissen ermöglichen? Was uns auch immer sehr wichtig ist: Wie kann ich den Auftraggeber:innen der Evaluierung vermitteln, dass ein komplexer Forschungsgegenstand auch eine komplexe wissenschaftliche Herangehensweise benötigt? Ich erinnere mich da an ein Gespräch mit der Geschäftsführung eines großen, Sozialen Dienstes in Österreich, in dem auf unser Forschungsdesign mit dem Spruch reagiert wurde: „Wir wollen keinen Rolls Royce, uns reicht auch ein Dacia“. Dann haben wir in längeren Gesprächen zu verdeutlichen versucht, dass es nicht darum gehe, welche Marke das Auto hat, sondern ob es überhaupt fährt oder nicht. Durch diese vermittelnden Prozesse konnten wir schlussendlich die Wirkungsevaluation zu einer wichtigen Dienstleistung realisieren, sie wird im Rahmen einer Forschungsförderung finanziert und in Kooperation mit dem Praxispartner umgesetzt. Ein letzter Punkt noch, eine andere Facette, die mich als Leiterin nicht nur eines Instituts, sondern auch der konkreten Forschungsprojekte immer sehr beschäftigt: Wie schaf-

9 In Österreich bedeutet das Wort Enquete „eine Untersuchung, die von Abgeordneten eines Parlaments durchgeführt wird. Häufig geht es dabei um konkrete Missstände (Untersuchungsausschuss) oder aber um bestimmte Gesetzesvorschläge, die noch näher diskutiert werden sollen. Zu parlamentarischen Enqueten werden oft auch Experten und Expertinnen eingeladen, die nicht dem Parlament angehören“ (m.politik-lexikon.at/enquete, Abruf 16.05.2024).

fen wir es in unseren Projektteams, dass diese vielfältigen, methodischen Kompetenzen, die anspruchsvolle, multiperspektivische und methodenplurale Wirkungsevaluationen benötigen, vorhanden sind? Wie schaffen wir es, mit überschaubaren Ressourcen Forschungsteams zusammenzustellen, die sowohl quantitative Methoden beherrschen und multivariate komplexe Berechnungen durchführen sowie mit großen Datenmengen arbeiten können, aber auch über elaborsiertes, qualitatives Know-how verfügen und zum Beispiel rekonstruktive Verfahren anwenden können? Diese Kompetenzen in einem Projektteam zu vereinen, ist äußerst schwer. Ich denke, das setzt auch fundierten Wirkungsevaluationen deutliche Grenzen.

Jörg Faust: Natürlich haben Entscheidungstragende große Anreize, politisch opportunistisch Evidenz zu selektieren. Das ist eine große Herausforderung und wir werden diese Anreize nicht beseitigen können, sie sind in Demokratien Teil des politischen Wettbewerbs. Gleichzeitig sind Entscheidungstragende in Demokratien auch an einer Verbesserung der Kollektivgutversorgung interessiert, weil sie in Wahlen umfassende Mehrheiten hinter sich scharen müssen. Die Existenz eines robusten Evaluierungssystems, das Entscheidungstragende dazu veranlasst, auf Evidenz aus Evaluierung zu reagieren und sich zu dieser transparent zu verhalten, kann die Anreize zur opportunistischen Evidenzselektion mindern und diejenigen zur Politikverbesserung stärken. Darin sehe ich den großen Vorteil eines verregelten Evaluierungssystems, in dem Entscheidungstragende darauf verpflichtet werden, sich zu qualitativ hochwertiger Evidenz aus Evaluierungen auch zu verhalten. Eine methodische Anmerkung zur Evaluation von komplexen Interventionen und geringer Fallzahl, Herr Milbradt hatte das gerade illustriert. Als Evaluierende stehen wir diesbezüglich oft vor der Herausforderung, dass wir einerseits die Nachfrage nach einfach vermittelbarer, generalisierbarer und in ihrer Komplexität reduzierter Information bedienen sollen. Und auf der anderen Seite haben Sie das Problem, aufgrund der Komplexität von Evaluierungsgegenständen und geringer externer Validität nur schwer dieser legitimen Anforderung nachzukommen. Nach meiner Erfahrung besteht aufseiten von Entscheidungstragenden nur eine begrenzte Kapazität, sich mit der Komplexität von Interventionen und Methoden auseinanderzusetzen. Es wird dann zu unserer Herausforderung als diejenigen, die Evidenz an diese Entscheidungstragenden herantragen, die Komplexität entsprechend zu reduzieren, auch wenn wir es mit komplexen Interventionen in komplexen Kontexten zu tun haben.

Herausgeber:innen: Wir sind gerade bei der Frage, wie sich Anforderungen, die politische Entscheidungsträger:innen an Evaluation bzw. an Evidenz beispielsweise eines Bundesprogramms wie „Demokratie leben!“ stellen, gut ins Verhältnis setzen lassen zu den Anforderungen, die beispielsweise Praktiker:innen formu-

lieren, wenn es um die Evaluation ihres Tuns geht. Welche Lösungsansätze sehen Sie hierfür?

Thomas Widmer: Sie fragen nach Möglichkeiten und Instrumenten, um die entsprechenden Herausforderungen zu überwinden. Ich würde gerne einen Schritt zurückgehen und noch etwas bei der Diagnose bleiben, weil sich ja aus der Diagnose dann auch die entsprechenden Therapien ableiten lassen. Ich habe neun Punkte, aber eine Vorbemerkung muss ich noch machen. Grundsätzlich ist ja alles Wirkung, alles, was wir beobachten können, ist grundsätzlich Wirkung. Von daher sollten wir präzise sein bei der Abgrenzung, was wir eigentlich als Wirkung bezeichnen. Die Diskussion dreht sich in der Regel nicht um Wirkungsevaluation, sondern eigentlich um Wirksamkeitsevaluation. In der Regel befassen wir uns mit den intendierten Wirkungen, die eintreten, also mit Wirksamkeit. Und wir befassen uns weniger mit den nicht intendierten Wirkungen. Aber auch diese sollten unter dem Titel Wirkungsevaluation meines Erachtens mit beachtet werden, was aber selten möglich ist. Nach diesen Vorbemerkungen nun die neun Punkte: Das erste Problem ist, dass wir eigentlich gar nicht so recht wissen, um was es geht. Phänomene wie Demokratie oder Extremismus sind massiv unterdefiniert bzw. wir haben sehr unterschiedliche Definitionen dieser Phänomene. Meines Erachtens ist es wichtig, daran zu arbeiten, das zu präzisieren. Wenn ich die entsprechenden Programme anschau, ist beispielsweise zum Teil diffus, was mit Demokratie bezeichnet wird. Zweitens: Wir befassen uns hier mit Themen, die sowohl manifeste Anteile haben, aber auch solche, die gar nicht direkt sichtbar und erkennbar, messbar, erfassbar sind. Drittens: Es handelt sich hier um Phänomene, die recht vielfältig sind. Wenn wir uns beispielsweise mit Extremismusprävention befassen, kann dieser Extremismus sehr unterschiedliche Ausprägungen haben. Ich meine damit nicht nur politische Ausprägungen wie links oder rechts, sondern Vieles mehr. Beispielsweise kann Extremismus gewalttätig sein oder nicht. Wenn wir uns mit der Extremismusfrage befassen, haben wir Verhaltensweisen, die zumindest potenziell illegal sind. Sie sind sozial unerwünscht bzw. werden als sozial unerwünscht empfunden. Deswegen sind sie teilweise schwierig zu fassen, weil sozial unerwünschtes Verhalten bzw. sozial unerwünschte Einstellungen eher im Verdeckten ausgelebt werden. Viertens: Wir haben potenziell, was sehr unterbelichtet ist in dieser Diskussion, eine adressatenspezifische Wirksamkeit. Wir können nicht davon ausgehen, dass eine gewisse Intervention bei verschiedenen Personen die gleiche Wirkung entfalten wird, sondern wir müssen damit rechnen, dass diese Wirkungen heterogen ausfallen. Fünftens: Wir bewegen uns in einem Bereich stark emotionalisierter, moralisch aufgeladener Phänomene. Sechstens: Wir haben dann zumeist Angebote non-formaler oder informeller Bildung, Coaching, Sozialberatung, also Angebote, die zumeist unter der Kategorie soziale Dienstleistungen zusammengefasst werden können. Diese zeichnen sich dadurch aus, dass sie sich jeweils unterschiedlich präsentieren, dass die kon-

krete Ausgestaltung und Umsetzung im spezifischen Kontext relevant sind. Siebentens: Wir können daher nicht von einem allgemeinen Interventionsmodell ausgehen, das wir dann uniform in unterschiedlichen Kontexten unverändert einsetzen können. Achtens: Wir haben zudem das Problem, dass diese Maßnahmen in der Regel nicht kontinuierlich, sondern punktuell ein- und umgesetzt werden, dass sie also keine Dauerhaftigkeit haben, was wiederum Wirkungsnachweise erschwert. Neuntens: schließlich, und das ist für mich der wichtigste Punkt: Wir wissen relativ wenig über die kausale Herleitung beispielsweise von Dingen wie Demokratieverständnis oder extremistischen Haltungen. Hier besteht also ein Forschungsbedarf, der es uns erst sekundär ermöglichen würde, an dafür ursächlichen Faktoren zu arbeiten, sofern wir hier Veränderungen herbeiführen wollen.¹⁰

Herausgeber:innen: Wir haben jetzt von Thomas Widmer einen Neun-Punkte-Problemaufriss über das zu evaluierende Feld gehört. Welche Anregungen haben Sie aus Ihrer Perspektive: Wie lassen sich diese Herausforderungen in einer wirkungsorientierten Evaluation bewältigen?

Hemma Mayrhofer: Ich halte es zunächst einmal für essenziell, mich von diesen Herausforderungen nicht grundlegend beirren oder einschüchtern zu lassen, sondern eine allgemeine ethnografische Forschungshaltung einzunehmen. Das heißt, ich lasse mich auf das Forschungsfeld ein, ich lasse mich auch von diesem überraschen, ich lerne sehr viel von den Praxisakteur:innen im Feld und ich sehe deren Kompetenzen. Zudem finde ich es wichtig, die Handlungszwänge der Politik zu sehen und anzuerkennen: Die sind gegeben und ich kann und werde sie nicht verändern. Zugleich und gerade vor diesem Hintergrund kann ich als Wissenschaftlerin, als Evaluatorin nach Möglichkeiten suchen, die Dialog ermöglichen und eröffnen. Das heißt auch, dass ich proaktiv auf Politik und Praxis zugehen und mir überlegen muss: Wie schaffe ich es, Zugang zu finden? Wie schaffe ich es, gehört zu werden? Und wie gut höre ich gleichzeitig auch selbst zu? Das finde ich sehr zentral.

Einen anderen Punkt halte ich ebenfalls für sehr bedeutsam: Wir haben in unserer Forschung zum Teil mit schwer erreichbaren Adressat:innen-Gruppen zu tun. Ich halte es für äußerst wichtig, zu überlegen, wie wir auch über Wirkungsevaluation hinaus methodische Zugänge zu diesen finden können. Wie kann ich zur Unterstützungssituation von Menschen forschen, die kognitive Behinderungen haben, die an Demenz erkrankt sind oder ähnliche Einschränkungen haben? Wie kann ich zu jungen Menschen und ihren Lebensgeschichten forschen, mit

10 Die neun Punkte mit Lösungsvorschlägen sind ausführlicher dargelegt in: Widmer, T./Hirsch, C. (2007): Herausforderungen und Konsequenzen der Evaluation von Massnahmen gegen Rechtsextremismus. LeGes – Gesetzgebung & Evaluation, 18. Jg., H. 2, S. 255–274.

jungen Menschen forschen, die straffällig geworden sind und zugleich auch nicht gerne darüber reden? Wir veröffentlichen demnächst eine Studie zu Resilienz-faktoren gegen Extremismus. Hierfür haben wir zu biografischen Erzählungen Fallrekonstruktionen durchgeführt und wir hatten das große Problem, dass es sehr schwer war, junge Menschen zu finden, die überhaupt bereit waren, mit uns darüber zu reden und ihre Lebensgeschichte zu erzählen. Folglich haben wir eine einseitige Fallauswahl, das muss man reflektieren. Das verweist auf die Frage: Wie können wir jene Bereiche erforschen, die methodisch schwer zugänglich sind? Da ist viel Methodenentwicklung gefragt, für die in der Umsetzung von Wirkungsevaluation bzw. in der empirischen Forschung allgemein auch Freiräume zu schaffen sind. Um abschließend an einen Aspekt vom Anfang anzuschließen: Wir hatten in unserer bereits erwähnten Studie zur Wirkungsevaluation mobiler Jugendarbeit die Möglichkeit zu scheitern. Das war wunderbar. Das macht im Forschen mutig und eröffnet Spielräume für Neues, dazu kann ich nur einladen. Also: Mut zum Scheitern. Aber dafür braucht es entsprechende Rahmenbedingungen.

Helle Becker: Ich würde, ganz ähnlich wie Frau Mayrhofer, dafür plädieren, das Forschungsfeld mitzunehmen. Das heißt, die Interessen und Bedarfe der Praxis stärker zu berücksichtigen und auch die Sensibilitäten des Feldes. Das heißt zum Beispiel, dass die Frage immer mitgedacht werden sollte, wie die Forschung selbst die Praxis beeinflusst. Zugleich darf man nicht vergessen, dass es in jedem Feld auch Routinen der Berichterstattung, also der Selbstevaluation, gibt. Im Bereich der politischen Bildung zum Beispiel gibt es einen hochgradigen Zwang zur Berichterstattung, für die die Träger Daten erheben müssen. Das ist für sich schon eine Belastung. Daran schließt sich die Frage an: Was mutet man der Praxis eigentlich zu, wenn man sie beforscht? Aus der Sicht der politischen Bildung gibt es zumindest in diesem Bereich zudem zu viel Rechtfertigungsforschung und viel zu wenig Forschung, die Praxisentwicklung betreibt. Das schließt an das an, was Frau Mayrhofer eben sagte. Solche Rechtfertigungsforschung, die vor allem Legitimationszwecken dient, lässt wenig Raum für Fehlerfreundlichkeit oder Scheitern. Hingegen gibt es wenig Forschung, die mithilft, die Praxis weiterzuentwickeln, und die dann auch entsprechende Designs vorweisen kann, die die benannten Sensibilitäten berücksichtigen. Es gibt zum Beispiel ein Modell der partizipativen Evaluation in der politischen Bildung. Ich selbst habe ein größeres Forschungsprojekt mit einer 360-Grad-Forschung gemacht – hochgradig partizipativ. Das zielt darauf, die Praxis selbst an ihrer Entwicklung durch die Evaluation zu beteiligen. Der nächste Punkt betrifft diese Sensibilitäten: Aus Sicht der politischen Bildung gibt es demokratietheoretische Grundfragen zu klären, sowohl hinsichtlich der Frage, welche Wirkungen eigentlich durch politische Bildung erzielt werden sollen, als auch in konzeptioneller Hinsicht, welche Art oder welche Formen politischer Bildung zu den entsprechenden Wirkungen beitragen können. Und was bedeutet das für das Praxisfeld? Entsprechend scharf werden die

Debatten geführt, was denn eine „richtige“ bzw. eine positive Wirkung politischer Bildung ist und was nicht. Last but not least, Frau Mayrhofer hatte es ebenfalls bereits angesprochen: Es wäre wünschenswert, wenn es hier eine größere Allianz und mehr Kommunikation innerhalb der Wissenschaften und natürlich auch mit der Praxis und mit der Politik gäbe. Aber mir würde schon mehr Verständigung innerhalb der Wissenschaften als ersten Schritt reichen, zum Beispiel im Hinblick auf die Frage: Wie kann man Einfluss nehmen auf die Forschungsförderung? Es gibt kaum Finanzquellen, um in diesem Bereich zu forschen, wenn die Forschung nicht gebunden ist, zum Beispiel an große Bundesprogramme. Das, finde ich, geht die Forschungscommunity an und da würde ich mir ein bisschen mehr Austausch und gegenseitige Unterstützung wünschen.

Jörg Faust: Wenn ich die Diskussion heute betrachte und auch andere, die ich ab und zu eher als externer Beobachter im Feld der Demokratieförderung und Extremismusprävention in Deutschland begleiten darf, dann erlebe ich immer wieder, dass die Komplexität von Interventionen und Evaluierungen betont wird. Gleichzeitig ist das alles aus meiner Perspektive nicht neu und singulär, da dieses Arbeitsfeld ja seit Jahrzehnten besteht und auch in anderen Ländern gibt es Forschung und Evaluation zur politischen Bildung. Ich denke, die im Feld tätige Forschung und Evaluation sollte in die Lage kommen, nun auch Synthesewissen herauszuarbeiten und anzubieten, was funktioniert und was nicht. Welches gesättigte Wissen über die Wirkungen von politischer Bildung oder von Maßnahmen der Extremismusprävention existiert, welches sind grundlegende Handlungsempfehlungen an die Politik? Was sind die fünf oder drei oder auch sieben Wirkungsbefunde, die sich – bei aller Komplexität – vielleicht verdichten und synthetisieren lassen? Das ist eine Frage, die ich an das Feld stellen würde. Und ich glaube, wenn Sie sich stärker auf diese Frage fokussieren würden, dann könnten Sie mit einer entsprechend komprimierten Ergebnisvermittlung vielleicht auch andere erreichen und besser adressieren. Dass das aus meiner Sicht noch nicht ausreichend passiert, kann auch ein Problem kollektiven Handelns in der Evaluations- und Forschungslandschaft sein. Erschwerend kommt hinzu, dass die Politik in diesem Feld auch nicht kohärent als Nachfrager auftritt und letztlich auch kein funktionales Evaluierungssystem errichtet hat. Auf Angebots- und Nachfrageseite von Evaluation gibt es ein Collective Action Problem.

Herausgeber:innen: Mit dem Bundesprogramm „Demokratie leben!“ haben wir ein Programm vor uns, das sehr verschiedene Arbeitsfelder berührt, unterschiedliche Zielgruppen adressiert und versucht, das Ganze unter einem Dach zusammenzuführen. Herr Widmer hatte eingangs gesagt, im Umgang mit der Wirkungsfrage sei unter anderem eine Atomisierung zu beobachten. Man könnte sagen, dass dies eine Strategie ist, die – von der Konstruktion des Bundesprogramms her – geradezu nahelegt wird. Was wäre die Gegenstrategie

zur Atomisierung? Oder ist eine kleinteilige Zerlegung das Mittel der Wahl, um einem komplizierten bis komplexen Evaluationsgegenstand entsprechen zu können?

Thomas Widmer: Aus meiner Perspektive wäre die Problematik dadurch zu bewältigen, dass darüber diskutiert wird, welche Evaluationsfrage man letztlich bearbeitet. Ich habe den Eindruck, dass es zu hoch gegriffen ist, danach zu fragen: Welche Wirkungen hat das Programm „Demokratie leben!“? Für mich ist es realistischer zu fragen, welche Wirkungen einzelne Elemente des Programms haben. Meines Erachtens kann Evaluation keine empirisch abgestützte Aussage darüber treffen, ob ein solches Programm wirksam ist und damit eine Entscheidung darüber vorbereiten, ob es ein Nachfolgeprogramm geben soll oder nicht. Solche generellen Entscheidungen sind meines Erachtens politische Entscheidungen, die man nicht alleine evidenzbegründet treffen kann. Was man aber sehr wohl kann als Evaluation ist, den fachlichen Diskurs über ein solches Programm mit entsprechender Evidenz zu versorgen. Und dieser fachliche Diskurs bewegt sich zumeist nicht auf der Ebene des Gesamtprogramms, sondern befasst sich in der Regel mit einer spezifischen Form der Intervention. Beispielsweise wären dann die Diskurse in der Pädagogik oder der Sozialarbeit das Ziel einer solchen Evidenzvermittlung. Das hat auch mehr Potenzial im Hinblick auf eine sinnvolle und angemessene Evidenznutzung, als zu versuchen, das Gesamtprogramm durch entsprechende Evidenz legitimatorisch zu stärken. Ich glaube, das ist ein unerfüllbares Ziel. Das heißt in der Konsequenz, die Atomisierung ist nur dann ein Problem, wenn sie dazu dienen soll, eine übergreifende Fragestellung zu bearbeiten. Es wird ja oft argumentiert, dass das Programm mehr ist als die Summe der einzelnen Interventionen, die darin enthalten sind. Ich habe da meine Zweifel, ob das wirklich so ist. Das bezieht sich weniger auf die verschiedenen Supportmaßnahmen, die oft auf der Gesamtprogrammebene stattfinden. Da sehe ich gewisse Effekte. Aber was die Evaluation eines solchen Programms insgesamt angeht, habe ich große Schwierigkeiten damit. Wir haben uns vor längerer Zeit unter dem Titel „Evaluation von Mehrebenen-Netzwerkstrategien“¹¹ einmal mit der Frage der unterschiedlichen Ebenen von Programmen, Strategien, Projekten und so weiter befasst. Das könnte in diesem Zusammenhang vielleicht noch interessant sein.

Hemma Mayrhofer: Ich möchte gerne einen Aspekt ergänzen oder vielmehr noch einmal hervorheben: Ich halte es für zentral, dass nicht der Versuch unternommen wird, einzelne Wirkfaktoren zu isolieren bzw. die Wirkungen verschiedener Faktoren getrennt herauszuarbeiten. Stattdessen muss das Augenmerk insbesondere auch auf das Zusammenwirken der verschiedenen Faktoren gerichtet

11 Widmer, T./Frey, K. (2006): Evaluation von Mehrebenen-Netzwerkstrategien. In: Zeitschrift für Evaluation, 5. Jg., H. 2, S. 287–316.

werden. Keine Intervention erzeugt alleine eine Wirkung, sondern sie trifft auf bestimmte Umstände, Lebenssituationen, Vorgeschichten usw. und wirkt mit anderen Faktoren komplex zusammen. Das bedeutet nicht nur Nicht-Linearität und Multikausalität, sondern verweist auch auf die Pfadabhängigkeit von Entwicklungen und Wirkungen. Auch das muss eine Wirkungsevaluation hinreichend erfassen können. Erst dadurch erkenne ich, warum etwas manchmal wirkt und ein anderes Mal nicht und was den Unterschied ausmacht. Das ist natürlich methodisch anspruchsvoll, soll aber bei der Realisierung von komplexen Wirkungsevaluationen im Mittelpunkt stehen.

Thomas Widmer: Ich möchte in diesem Zusammenhang auch noch einen Punkt ergänzen. Wir haben uns jetzt stark auf Wirkungsevaluation orientiert. Was nicht vergessen werden darf, ist, dass wir auch andere, anders ausgerichtete Evaluationen antreffen, die sich beispielweise mit Prozessen, mit Umsetzung und so weiter befassen. Dort ist die Situation eine etwas andere. Es erscheint durchaus sinnvoll und auch zweckmäßig, eine Evaluation durchzuführen, die sich mit Prozessqualität befasst. Je nach Definition lässt sich die Implementation eines Programms durch verschiedene einzelne Projekte, das heißt die Projektumsetzung, als Wirkung der entsprechenden Aktivitäten auf der Programmebene verstehen. Das verweist auf einen definitorischen Aspekt. Davon unbenommen: Grundsätzlich erscheint alles, was prozessbezogen ist, also Leistungserbringung, Leistungserstellung, Interaktion zwischen Leistungserbringern und Leistungsempfängern, als handhabbarer im Vergleich zur Frage nach den entsprechenden Auswirkungen bei den Zielgruppen oder darüber hinaus.

Jörg Faust: Ich würde hier nicht nur die Prozessevaluation ergänzen, die Herr Widmer gerade angesprochen hat, sondern auch Allokations- und Portfolioanalysen als wichtiges evaluatorisches Produkt auf einer Gesamtprogrammebene oder im Feld insgesamt: Meines Wissens gibt es keinen Portfolioüberblick, wie das Portfolio der Demokratieförderung und der Extremismusprävention in Deutschland aussieht, wo welche Projekte mit welchen Zielsetzungen durchgeführt wurden und nach welchen Kriterien staatliche Mittel wohin verteilt wurden. Das ist nicht gut und sollte durch eine übergreifende, strategische Portfolio- und Allokationsanalyse aufgearbeitet werden. Politischer Extremismus und Demokratieverdrossenheit sind elementare Herausforderungen für unsere Gesellschaft. Entsprechend sollten die verantwortlichen Ministerien auf unterschiedlichen Ebenen die Möglichkeiten schaffen, das Portfolio staatlich finanzierter Aktivitäten zu identifizieren und zu analysieren. Neben der Frage der Wirksamkeit ist doch auch wichtig zu wissen, wie sich die Allokation der Ressourcen auf Interventionen, auf unterschiedliche Teilprojekte, vielleicht auch räumlich, geografisch über die Jahre verändert hat und was für Kriterien diesen Allokations- oder Portfolioentscheidungen zugrunde liegen?

Herausgeber:innen: Mit Blick auf die von Thomas Widmer geäußerte Skepsis, Wirkungsevaluation eines solchen Gesamtgebildes wie „Demokratie leben!“ zu realisieren: Jörg Faust, was sind denn Ihre Erfahrungen in der Entwicklungszusammenarbeit damit? Dort finden wir ja ebenfalls Programme vor, die eine gewisse Komplexität aufweisen.

Jörg Faust: Ja, Programme in der Entwicklungszusammenarbeit sind komplex, manchmal sogar überkomplex, weil zu viele Ziele an das Politikfeld herangetragen werden. Gerade deshalb will ich das, was Herr Widmer sagte, auch noch mal unterstreichen. Die Analyse sollte auf unterschiedlich relevante Wirkungsfragen zielen, die ein Programm zum Gegenstand hat. Oder es können gegebenenfalls nur Teilprojekte eines größeren Programms evaluiert werden. Viele Ansätze, Instrumentenkombinationen oder Programme in der Entwicklungszusammenarbeit sind derart komplex, dass ein wichtiger Beitrag der Evaluierenden darin liegt, für die Praxis eine Rekonstruktion einer Wirkungs- oder Interventionslogik vorzunehmen. Bei den strategischen Evaluierungen des DEval rekonstruieren bzw. erarbeiten wir oft die Wirkungslogik eines Programms oder unterschiedlicher Interventionen in einem Themenfeld möglichst in Zusammenarbeit mit den Praktiker:innen. Die Erarbeitung solch übergreifender Wirkungslogiken und deren Vermittlung an Durchführungsorganisationen und Ministerium ist ein Mehrwert strategischer Evaluierungen, der manchmal unterschätzt wird.

Helle Becker: Ich frage mich, wer definiert eigentlich, welche Wirkungen in den Blick genommen werden sollen? Wenn ich ein solches Bundesprogramm wie „Demokratie leben!“ betrachte, dann zeitigt dieses zum Beispiel ja auch Wirkungen auf einer strukturellen Ebene, die ganz erheblich sind. Diese sind etwas völlig anderes als individuelle Bildungswirkungen aufseiten von Adressat:innen, sie sind auf einer ganz anderen Ebene angesiedelt. Ich würde daher mindestens unterscheiden wollen in Mikro-, Meso-, Makroebenen der Wirkungen von solchen Förderprogrammen.

Björn Milbradt: Ich würde diesen Aspekt aufgreifen wollen. In der Programmevaluation „Demokratie leben!“ sind wir durchaus damit konfrontiert, dass es Wirkungserwartungen auf einer Gesamtprogrammebene gibt. Diese sind teilweise auch sehr umfassend, zum Beispiel in Bezug auf Haltungen und Einstellungen von Zielgruppen, teilweise werden sogar Wahlergebnisse in Zusammenhänge mit Programmaktivitäten gestellt. Hier ist es wichtig, als Evaluation darauf hinzuweisen, dass ein solches Programm nicht primär daraufhin angelegt ist, solche Wirkungen zu erreichen, sondern das Programm in seinen verschiedenen Programmbereichen sehr spezifische Wirklogiken verfolgt. Unbenommen dessen wäre es natürlich extrem schwierig, Kausalnachweise auf einer solchen Aggregatebene zu führen. Ein Weg, den wir im Rahmen der Programm-

evaluation versuchsweise beschritten haben, ist es, die Wirkungen solch eines großen Programmes auf konkrete Wirkungen auf eine regionale Ebene herunterzubrechen und dort empirisch zu rekonstruieren. So ein Programm schwebt natürlich nicht abstrakt über allem, sondern prägt regionale Ausformungen aus. Das heißt, auf regionaler Ebene arbeiten Programmakteure zusammen, erreichen unterschiedliche Wirkungen, arbeiten mit externen Akteuren zusammen und werden von externen Akteuren unterschiedlich beurteilt (siehe Beitrag von König/Stärck/Zierold in diesem Band). Ergebnisse, die Wirkungen auf einer solchen regionalen Ebene zeigen können, stehen dann jedoch vor dem Problem der Verallgemeinerbarkeit: Was sagen die Erkenntnisse aus einer bestimmten Region für das Gesamtprogramm aus? Auch über diese Ebene hinaus ist die Frage interessant, was exemplarische Erkenntnisse, zum Beispiel aus Regionalanalysen, bedeuten. Beispielsweise ist eine Erkenntnis aus solchen Regionalanalysen in „Demokratie leben!“, dass die programmgeförderten Akteur:innen in einer Region andere Akteur:innen, die nicht über das Programm gefördert werden, ver- oder zumindest in ihrem Engagement zurückdrängen können, weil dann – gewollt oder nicht – Verantwortungszuschreibungen für Demokratieförderung und Extremismusprävention an „Demokratie leben!“-Akteur:innen erfolgen. Solche Nebenwirkungen, die wir empirisch durchaus belegen können, sind dann zu befragen darauf, ob sie sich verallgemeinern ließen. Das ist also eine Frage des Geltungsanspruches solcher Befunde.

Thomas Widmer: Ich möchte nun doch noch einen positiven Punkt einbringen: Meines Erachtens leistet die Evaluation im Allgemeinen und die Wirkungsevaluation im Speziellen einen wichtigen Beitrag zur Klärung der Programmatik eines Programms. Beispielsweise werden die Schwachstellen von Programmüberlegungen aufgedeckt, wenn man sich Gedanken über die Wirkungsevaluation macht. Bereits das Nachdenken über eine Wirkungsevaluation leistet einen wichtigen Beitrag zur Entwicklung eines Programms. Meines Erachtens wäre es wichtig, solche Überlegungen schon zu einem frühen Zeitpunkt der Programmentwicklung anzustellen und nicht erst zu einem Zeitpunkt, wenn die Konzeption des Programms bereits abgeschlossen ist. Hierbei kann es nicht nur darum gehen, retrospektiv die Wirkungsevaluationsbefunde aus der vorhergehenden Förderperiode zu berücksichtigen. Stattdessen ist es sinnvoll, sich prospektiv Gedanken zu zukünftigen Wirkungsevaluationen zu machen. Dies leitet dann die entsprechenden programmgestaltenden Personen auch an, sich konkrete Zielsetzungen zu überlegen, diese Zielsetzungen dann derart zu definieren, dass sie auf plausiblen Ursache-Wirkungs-Zusammenhänge aufbauen und Interventionslogiken erkennbar machen, die dann wiederum im Rahmen von Wirkungsevaluationen untersucht werden können (siehe Beitrag von Hense in diesem Band). Das wäre sehr wertvoll und ich glaube, das sollte man mehr machen.

Herausgeber:innen: Das klingt schon beinahe wie ein Schlusswort. Damit noch mal abschließend in die Runde gefragt: Wie lässt sich aus Ihrer Sicht der Wert von Wirkungsuntersuchungen in einem Satz zusammenfassen?

Hemma Mayrhofer: Ein Aspekt, den ich auf Basis meiner Erfahrungen als eine wichtige Wirkung von Wirkungsevaluation bezeichnen würde, ist ihr potenzieller Beitrag zu einem besseren Verständnis für die spezifischen Arbeitsweisen und Wirkzusammenhänge eines Handlungsfeldes aufseiten der politisch Verantwortlichen.

Jörg Faust: Der Wert von unabhängigen Wirkungsevaluationen ist ausgesprochen hoch, um die Aufmerksamkeit von Entscheidungstragenden zu erlangen und unabdingbar, damit die Kernfunktionen von Evaluierung – praxisrelevanter Erkenntnisgewinn, Lernen und Rechenschaftslegung – erfüllt werden können.

Helle Becker: Ich wiederhole einen Punkt, den ich eben schon mal gesagt habe: Für mich wäre der Wert von Wirkungsuntersuchungen eindeutig darin gegeben, wenn die Praxis Erkenntnisse an die Hand bekommt, um sich weiterzuentwickeln, und die weniger darin bestehen, Förderprogramme weiterzuentwickeln. Das würde ich mir wünschen und das wäre dann meines Erachtens wiederum auch fruchtbar im Hinblick auf die Konzipierung von Programmen.

Das Gespräch führten Frank Greuel, Franziska Heinze und Frank König.

Verzeichnis der Autorinnen und Autoren

Becker, Helle, Dr. phil., Kultur- und Erziehungswissenschaftlerin, ist Geschäftsführerin von Transfer für Bildung e. V. in Essen und leitet dort die Fachstelle politische Bildung. Sie arbeitet zu politischer und kultureller Jugend- und Erwachsenenbildung, internationaler/europäischer Jugend- und Bildungsarbeit sowie zur Zusammenarbeit unterschiedlicher Bildungssektoren und deren Qualitätsentwicklung.

Braun, Marina, M. A. Bildungswissenschaften, war von 2022 bis 2023 wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte liegen in der qualitativen Evaluationsforschung und der Demokratieförderung. In ihrer Promotion an der TU Dortmund beschäftigt sie sich mit Wissen im Kontext der politischen (Erwachsenen-)Bildung.

Euteneuer, Jonas, M. A. Empowerment Studies, war von 2022 bis 2023 wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind qualitative Forschung, (politisches) Engagement und Demokratietheorie(n).

Faust, Jörg, Dr. rer. pol., ist Direktor des Deutschen Evaluierungsinstituts der Entwicklungszusammenarbeit (DEval) und außerordentlicher Professor für Politikwissenschaft an der Universität Duisburg-Essen. Zudem ist er gegenwärtig Vorsitzender der Evaluierungsgruppe des Entwicklungsausschuss der OECD (OECD Evalnet). Neben seiner Arbeit im Feld der Evaluation forscht er unter anderem zu bi- und multilateraler Entwicklungszusammenarbeit sowie zu Prozessen politischen Wandels.

Figlestahler, Carmen, Dr. phil., M. A. Soziologie, ist Verwaltungsprofessorin an der Fakultät Soziale Arbeit und Gesundheit der HAWK Hochschule für angewandte Wissenschaft und Kunst Hildesheim/Holzminde/Göttingen. Zuvor war sie wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ am Deutschen Jugendinstitut e.V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind: Konstruktion und Bearbeitung sozialer Probleme, (sozial-)pädagogische Auseinandersetzungen mit politischem Extremismus und Jugendsoziologie.

Greuel, Frank, Dr. rer. pol., Dipl.-Pädagoge, wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind Evaluationsforschung zu Demokratieförderung und zur pädagogischen Prävention von Radikalisierung und die politische Sozialisation von jungen Menschen.

Haase, Katrin, Dr. phil., Dipl.-Sozialarbeiterin/Sozialpädagogin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind rekonstruktive Soziale Arbeit und Beratung im Kontext menschen- und demokratiefeindlicher Phänomene.

Heinze, Franziska, Dr. phil., Erziehungswissenschaftlerin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind unter anderem Forschung zu Organisationen und Strukturen der Kinder- und Jugendarbeit und politischen Bildung sowie Evaluation komplexer Programme. Sie ist Co-Sprecherin des AK Methoden in der Evaluation in der DeGEval.

Hemann, Max, M. A. Soziologie, wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind qualitative Sozialforschung, Wirkungsevaluationen und in der Demokratieförderung.

Hense, Jan Ulrich, Prof. Dr., ist selbstständiger Berater für Evaluation und Wirkungsorientierung in Gießen. Schwerpunkte seines Interesses sind unter anderem wirkungsorientierte Programmsteuerung, Wirkungsmodelle in der Evaluation sowie die Professionalisierung der Evaluation.

Johann, Tobias, M. A. Soziologie, wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind die wissenschaftliche Begleitung von Modellprojekten der Demokratieförderung sowie der Bearbeitung von Antisemitismus und Gruppenbezogener Menschenfeindlichkeit (GMF).

König, Frank, Dipl.-Politikwissenschaftler, Dipl.-Sozialpädagoge, wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind Evaluations- und Wirkungsforschung, Forschung

zum Verhältnis Staat und Zivilgesellschaft sowie zur Beratung im Kontext Rechtsextremismus.

Langer, Sarah, Dr. rer. pol., Organisationswissenschaftlerin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte liegen unter anderem in qualitativer Organisations- und kritischer Managementforschung, Fragen der Transformation von Management und Organisation sowie pädagogischer Organisationsberatung.

Loick Molina, Steffen, Dr. phil., Soziologe, wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte liegen in den Bereichen Soziologie des Körpers und der Kindheit, Geschlechtersoziologie, praxis- und diskurstheoretische Zugänge, Ethnografie sowie qualitativ-rekonstruktive Methoden.

Mayrhofer, Hemma, Dr. phil., ist Assistenzprofessorin und Leiterin des Instituts für angewandte Rechts- und Kriminalsoziologie der Universität Innsbruck. Ihre Arbeitsschwerpunkte sind unter anderem rechtssoziologische Forschung, sozialwissenschaftliche Sicherheitsforschung, Gewaltforschung, Jugendarbeitsforschung, Evaluations- und Wirkungsforschung.

Meijer, Laura, M. A. Affaires Internationales – Politikwissenschaft, wissenschaftliche Referentin im Projekt „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“ (2022–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind Bedarfslagen der Demokratieförderung und Extremismusprävention sowie Prozesse politischer Sozialisation im Jugendalter.

Milbradt, Björn, Dr. phil., Soziologe, war von 2016 bis 2024 Leiter der Fachgruppe „Politische Sozialisation und Demokratieförderung“ am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind Prozesse politischer Sozialisation, ihre politischen, gesellschaftlichen und institutionellen Kontexte sowie Gelingensbedingungen und methodologische Aspekte ihrer Erforschung.

Rehse, Aline, Dr. rer. pol., Soziologin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre aktuellen Arbeitsschwerpunkte betreffen die Evaluierung von Programmen sowie wirkungs- und evaluationsforschende Fragestellungen.

Reiter, Stefanie, Dipl.-Soziologin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind unter anderem die Evaluation komplexer Programme sowie Netzwerkforschung und Evaluationsforschung. Sie ist Co-Sprecherin des AK Soziale Dienstleistungen in der DeGEval.

Rüger, Stella, M.A. Kulturanalysen, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind qualitative Forschungsmethoden, insbesondere (Reflexive) Grounded Theory Methodologie und Situationsanalyse, soziale Ungleichheit und Intersektionalität (insbesondere mit Fokus auf die Schnittstelle Migration, Flucht und Behinderung) und (Anti-)Diskriminierung in Bildungskontexten.

Sammet, Kornelia, Dr. phil., Dipl.-Soziologin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte liegen in der Kultur- und Religionssoziologie, der Professionssoziologie, der Soziologie sozialer Ungleichheiten und der Armutsforschung sowie in der qualitativ-rekonstruktiven Sozialforschung.

Sauermann, Pia, Dr. rer. pol., M.A. Soziologie, wissenschaftliche Referentin im Projekt „Ermittlung von Bedarfslagen im Bereich Demokratieförderung und Extremismusprävention“ (2022–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind Bedarfslagen der Demokratieförderung und Extremismusprävention sowie Prozesse politischer Sozialisation im Jugendalter.

Schau, Katja, M.A. Soziologie, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind pädagogische Spannungsfelder in der politischen Bildung und Sozialen Arbeit, Problemkonjunkturen in der Auseinandersetzung mit politischer und religiös motivierter Radikalisierung, Bewertungen als Herausforderung in der Evaluationspraxis.

Schroeter, Ellen, M.A. Erziehungswissenschaften, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind Methoden der qualitativen Sozialforschung, Methoden der Evaluationsforschung sowie die wissenschaftliche Begleitung der Implementierung von Bundesprogrammen zur Extremismusprävention und Demokratieförderung.

Stärck, Alexander, Dr. phil., Dipl.-Erziehungswissenschaftler, war von 2020 bis 2024 wissenschaftlicher Referent im Projekt „Programmevaluation ‚Demokratie leben!‘“ am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Seine Arbeitsschwerpunkte sind Demokratieförderung, Vorurteils- und Diskriminierungsprävention, (frühe) Kindheit und Differenz, Methoden der rekonstruktiven Evaluations- und Kindheitsforschung.

Widmer, Thomas, Dr. phil., ist Professor für Politikwissenschaft an der Universität Zürich und leitet seit 2003 den dort angesiedelten Forschungsbereich „Policy-Analyse und Evaluation“. Er forscht und arbeitet unter anderem zu Evaluation, Policy-Analyse, Schweizer Politik, institutionellem Wandel und methodologischen Fragen.

Zierold, Diana, Dipl.-Soziologin, wissenschaftliche Referentin im Projekt „Programmevaluation ‚Demokratie leben!‘“ (2020–2024) am Deutschen Jugendinstitut e. V., Außenstelle Halle (Saale). Ihre Arbeitsschwerpunkte sind Partizipation, Extremismusprävention und Methoden der qualitativen und quantitativen Sozialforschung.