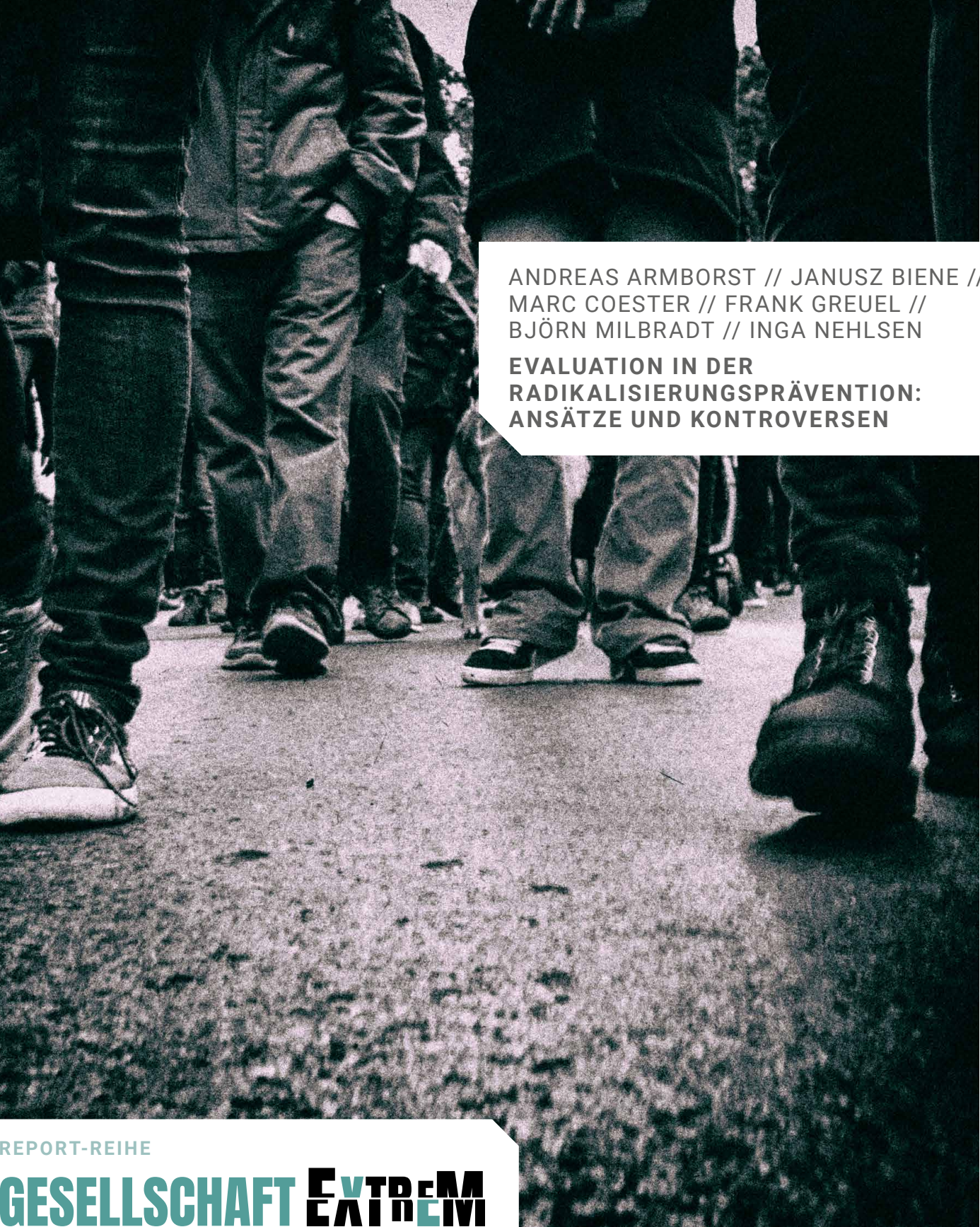


PRIF REPORT

PEACE RESEARCH INSTITUTE FRANKFURT / LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG



ANDREAS ARMBORST // JANUSZ BIENE //
MARC COESTER // FRANK GREUEL //
BJÖRN MILBRADT // INGA NEHLSSEN

**EVALUATION IN DER
RADIKALISIERUNGSPRÄVENTION:
ANSÄTZE UND KONTROVERSEN**

REPORT-REIHE

GESELLSCHAFT EXTREME

PRIF Report 11/2018

EVALUATION IN DER RADIKALISIERUNGSPRÄVENTION: ANSÄTZE UND KONTROVERSEN

ANDREAS ARMBORST // JANUSZ BIENE // MARC
COESTER // FRANK GREUEL // BJÖRN MILBRADT //
INGA NEHLSSEN

LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG (HSFK)
PEACE RESEARCH INSTITUTE FRANKFURT (PRIF)

Coverbild:

© Laloracio, iStock ID:512402784

iStock Standard-Lizenz: <https://www.istockphoto.com/de/legal/license-agreement>

Textlizenz:

Creative Commons CC-BY-ND 4.0 (Namensnennung/Keine Bearbeitungen/4.0 International).
Das Coverbild unterliegt eigenen Lizenzbedingungen.



Adresse:

Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK)
Baseler Straße 27–31
60329 Frankfurt am Main
Telefon: +49 69 95 91 04–0
E-Mail: gesellschaft.extrem@hsfk.de
<https://www.hsfk.de>

ISBN: 978-3-946459-39-2

DAS AUTORENTEAM

Dr. Andreas Armbrorst ist Leiter des Nationalen Zentrums für Kriminalprävention (NZK) in Bonn. Seine Forschungsschwerpunkte sind der sunnitische Islamismus und die Prävention islamistischer Gewalt.

Janusz Biene ist Friedens- und Konfliktforscher und Koordinator von PRO Prävention, einem im Integrationsbüro des Kreises Offenbach angesiedelten Radikalisierungspräventionsprojekt.

Prof. Dr. Marc Coester ist Professor für Kriminologie an der Hochschule für Wirtschaft und Recht in Berlin. Seine Forschungsschwerpunkte sind Kriminalprävention, Radikalisierungs- und Extremismusforschung, Evaluationsforschung, Jugendgewalt, Jugendstrafvollzug und Rückfallforschung.

Dr. Frank Greuel ist Erziehungswissenschaftler und wissenschaftlicher Referent am Deutschen Jugendinstitut in Halle. Sein Arbeitsschwerpunkt ist die wissenschaftliche Begleitforschung zur pädagogischen Prävention von Rechtsextremismus und gruppenbezogener Menschenfeindlichkeit.

Dr. Björn Milbradt ist Soziologe und Leiter der Fachgruppe „Politische Sozialisation und Demokratieförderung“ am Deutschen Jugendinstitut in Halle (Saale). Zuletzt erschienen ist von ihm im Verlag Barbara Budrich die Monographie „Über autoritäre Haltungen in ‚postfaktischen‘ Zeiten“. Sein Forschungsinteresse gilt insbesondere Bildungs- und Sozialisationsprozessen, ihren politischen, gesellschaftlichen und institutionellen Hintergründen und Bedingungen sowie methodischen und methodologischen Aspekten ihrer Erforschung.

Inga Nehlsen ist Friedens- und Konfliktforscherin und wissenschaftliche Mitarbeiterin am Nationalen Zentrum für Kriminalprävention (NZK) in Bonn. Ihre dortigen Forschungsschwerpunkte sind islamistische Radikalisierung und die Evaluation von Präventionsmaßnahmen gegen diese.

HSFK-REPORTREIHE „GESELLSCHAFT EXTREM“

Die sieben Beiträge dieser HSFK-Reportreihe bieten eine Bestandsaufnahme des Forschungsstands zu Radikalisierung und Deradikalisierung. Folgende zentrale Dimensionen werden dabei beleuchtet: Radikalisierungsprozesse von Individuen und von Gruppen, Radikalisierungstendenzen von Gesellschaften, Herausforderungen in der Deradikalisierungsarbeit, der Stellenwert von Online-Radikalisierung sowie Ansätze und Kontroversen bei der Evaluierung von Präventionsmaßnahmen.

Die Zusammenführung eines bislang eher fragmentarisch vorhandenen Wissensstandes ist auch deshalb wichtig, weil liberale Demokratien einmal mehr durch Extremismen – gleich ob politisch oder religiös begründet – herausgefordert sind. Dies ist insbesondere dann der Fall, wenn die plurale Verfasstheit einer Gesellschaft infrage gestellt wird und dabei von einigen auch antidemokratische Mittel zur Durchsetzung ihrer Ziele genutzt werden, bis hin zur Anwendung von Gewalt. Extreme politische Ansichten haben Konjunktur. In Deutschland radikalisierten sich Positionen auf der rechten und linken Seite des politischen Spektrums sowie im Kontext religiösen Sektierertums. Um der Entwicklung zu einer „Gesellschaft der Extreme“ vorzubeugen, liberale Werte und Institutionen zu stärken und die Ambivalenz von Radikalität zwischen gesellschaftlicher Herausforderung und Chance zu ergründen, müssen die Mechanismen individueller und kollektiver Radikalisierung verstanden werden – und zwar vergleichend über aktuelle Konjunkturen der Aufmerksamkeit für Islamismus oder salafistischen Dschihadismus hinaus.

Die Beiträge dieser Serie eint ein breites Verständnis von Radikalisierung, das den Ambivalenzen der Geschichte dieses umstrittenen Begriffs gerecht wird. Gleichwohl setzt jeder Report eigene, dem jeweiligen Thema angepasste Akzente in der Begriffsverwendung. Es ist genau dieser Pluralismus, den die Radikalisierungsforschung so dringend benötigt. Denn nur dann kann sie umfassend auf gesellschafts- wie sicherheitspolitisch virulente Fragen mögliche Erklärungen liefern und Handlungsoptionen generieren. Alle Reporte eint zudem die Empfehlung, noch stärker als bisher eine umfassende Präventionsagenda umzusetzen.

Die Autorentteams der einzelnen Reporte sind gegenstandsangemessen interdisziplinär und heterogen hinsichtlich ihrer eher wissenschaftlichen oder eher praxisbezogenen Expertise zusammengesetzt. Die Autorinnen und Autoren sind Teil eines vom Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK) koordinierten Forschungsnetzwerks „Gesellschaft Extrem: Radikalisierung und Deradikalisierung in Deutschland“, welches vom Bundesministerium für Bildung und Forschung gefördert wird. Weitere Informationen zu dem Projekt sowie Hinweise zu weiteren Publikationen und zu Informationsfilmen finden sich unter: www.gesellschaftextrem.hsfk.de.

Die Projektleitung

Prof. Dr. Christopher Daase

Prof. Dr. Nicole Deitelhoff

Dr. Julian Junk

IN DER HSFK-REPORTREIHE „GESELLSCHAFT EXTREM“ SIND ERSCHIENEN:

PRIF Report 5/2018

Was ist Radikalisierung? Präzisierungen eines umstrittenen Begriffs

Hande Abay Gaspar // Christopher Daase // Nicole Deitelhoff // Julian Junk // Manjana Sold

PRIF Report 6/2018

Radikalisierung von Individuen: Ein Überblick über mögliche Erklärungsansätze

Fabian Srowig // Viktoria Roth // Daniela Pisiu // Katharina Seewald // Andreas Zick

PRIF Report 7/2018

Brückennarrative: Verbindende Elemente für die Radikalisierung von Gruppen

David Meiering // Aziz Dziri // Naika Foroutan (mit Simon Teune // Esther Lehnert // Marwan Abou-Taam)

PRIF Report 8/2018

Radikalisierung der Gesellschaft? Forschungsperspektiven und Handlungsoptionen

*Eva Herschinger // Kemal Bozay // Oliver Decker // Magdalena von Drachenfels // Christian Joppke
(mit Klara Sinha)*

PRIF Report 9/2018

Herausforderung Deradikalisierung: Einsichten aus Wissenschaft und Praxis

Till Baaken // Reiner Becker // Tore Bjørgo // Michael Kiefer // Judy Korn // Thomas Mücke // Maximilian Ruf // Dennis Walkenhorst

PRIF Report 10/2018

Die Rolle des Internets und sozialer Medien für Radikalisierung und Deradikalisierung

Peter Neumann // Charlie Winter // Alexander Meleagrou-Hitchens // Magnus Ranstorp // Lorenzo Vidino

PRIF Report 11/2018

Evaluation in der Radikalisierungsprävention: Ansätze und Kontroversen

Andreas Armbrorst // Janusz Biene // Marc Coester // Frank Greuel // Björn Milbradt // Inga Nehlsen

Evaluation hilft uns, zu verstehen, wie die Prävention von Radikalisierung und Extremismus im gesellschaftlichen Kontext wirkt. Sie soll Antworten auf die häufig artikulierte Frage nach den sichtbaren Erfolgen von Prävention liefern. Allerdings werden an Evaluation teils überzogene Erwartungen hinsichtlich ihrer Leistungsfähigkeit und Durchführbarkeit gestellt. Das berechtigte Interesse an belastbaren Wirksamkeitsnachweisen stößt bei der Planung und Umsetzung von Evaluationsstudien im Bereich Deradikalisierung, Distanzierung und Prävention von Radikalisierung auf beträchtliche Schwierigkeiten. Dieser Bericht behandelt einige dieser Schwierigkeiten und zeigt, wie sich Evaluationen zur Erforschung der Wirksamkeit im Rahmen realistischer Möglichkeiten umsetzen lassen.

Die Idee einer „evidenzbasierten“ Prävention kann nur dann funktionieren, wenn die Evaluationsforschung sich der Eigenheiten, Widersprüche und Kontroversen in Wissenschaft und Praxis bewusst ist und diese kritisch reflektiert. Dazu beginnt der Bericht mit einer Erläuterung der Begriffe Radikalisierung, Prävention (von Radikalisierung) und Evidenzbasierung (der Prävention von Radikalisierung). Bereits dabei wird deutlich, dass Radikalisierung als ein Gegenstand der Prävention grundsätzlich umstritten ist, weil er Personen sehr pauschal als radikal oder extremistisch etikettiert.

Weil Prävention zeitlich und in Bezug auf die Zielgruppe endlos ausgeweitet werden kann, ergibt sich aus konzeptioneller und normativer Sicht die Frage, wo die Prävention (von Radikalisierung) anfangen und aufhören soll. Ohne eine endgültige Antwort auf diese Frage geben zu können, möchte der Bericht auf die Tendenz einer zunehmenden „Versicherheitlichung“ zivilgesellschaftlicher Akteure aufmerksam machen. Diese entsteht, wenn der Staat Erfordernisse der inneren Sicherheit in Bereiche überträgt, die originär nicht mit dieser Aufgabe betraut sind. Dabei soll allerdings nicht in Frage gestellt werden, dass die Prävention von gewaltbereiter Radikalisierung eine Aufgabe ist, die sich verschiedenen politischen Ressorts und Berufsgruppen stellt.

Unter Berücksichtigung dieser Vorzeichen kann eine evidenzbasierte Prävention von Radikalisierung allmählich Gestalt annehmen. Evidenzbasiert meint, dass Maßnahmen zur Verhinderung von sozialen Problemen auf wissenschaftlicher Grundlage konzipiert, implementiert und evaluiert werden. In Deutschland ist ein bestimmtes Verständnis von evidenzbasierter Präventionspraxis entstanden, das das Resultat einer längeren Entwicklung auf nationaler und internationaler Ebene ist. Der Report stellt anhand von konkreten Praxisbeispielen aus den Bereichen des Rechtsextremismus und religiös begründeten Extremismus Herausforderungen für die Evaluation dar.

Darauf aufbauend erörtert er mögliche Ansätze für die Evaluierung der *Effekte* von Präventionsmaßnahmen im Bereich der Radikalisierungsprävention, Deradikalisierung und Demobilisierung (sogenannte Wirkungsevaluationen). Durch diese bewusste thematische Ausrichtung soll aber nicht verschwiegen werden, dass andere Formen der Evaluation (z. B. die Prozessevaluation oder die formative Evaluation) eine ebenso große Bedeutung für die Entwicklung und Qualitätssicherung von Ansätzen zur Prävention von Radikalisierung haben.

Für die Untersuchung komplexer Programmeffekte gibt es keine Blaupause. In diesem Bericht werden jedoch erfolgversprechende Evaluationsansätze, wie die *Realist Evaluation*, *logische Modelle* und andere Evaluationsformen, vorgestellt. Sie erscheinen als pragmatischer und realistischer Mittel-

weg zwischen dogmatischen Positionen in der Evaluationsforschung in Bezug auf Evidenzbasierung. Eine entsprechend ausgerichtete Evaluationskultur könnte in der Projektlandschaft wiederkehrende Programmmechanismen identifizieren und deren kontextspezifische Effekte empirisch untersuchen.

Abschließend geht der Bericht auf Qualitätsmerkmale und Standards der Evaluationsforschung ein. Gütekriterien helfen den Auftraggebern und Adressatinnen von Evaluationsstudien, deren Verlässlichkeit und Aussagekraft abzuschätzen.

Die Handlungsempfehlungen benennen wichtige Voraussetzungen, unter denen Evaluationen ihren eigentlichen Zielen überhaupt erst gerecht werden können. Obwohl die Förderung von Projekten oft an eine verbindliche Evaluation gekoppelt ist, sind die dafür nötigen Voraussetzungen nicht immer gegeben. Wenn entsprechende Vorgaben und die existierenden Rahmenbedingungen besser aufeinander abgestimmt sind, werden auch die Ergebnisse aus Evaluationsstudien verwertbar und anschlussfähig.

1. Einleitung	1
2. Definition zentraler Begriffe	3
2.1 Radikalisierung	4
2.2 Prävention von Radikalisierung	5
2.3 Evaluation und evidenzorientierte Prävention von Radikalisierung	7
3. Stand der Evaluation von Radikalisierungsprävention	8
3.1 Evaluation im Bereich der Radikalisierungsprävention Rex	9
3.2 Evaluation im Bereich der Radikalisierungsprävention RelEx	11
4. Ansätze von Wirkungsevaluation	15
4.1 (Quasi-)Experimentelle Studien und Feldversuche	16
4.2 Qualitative Evaluationsstudien	17
4.3 Das Logische Modell	18
4.4 Realist Evaluation	19
5. „Studien belegen, dass ...“ – Verlässlichkeit und Aussagekraft von Evaluationsstudien	22
6. Empfehlungen	25
Literatur	27

1. EINLEITUNG

Terrorismus und politischer Extremismus waren 2017 laut einer Langzeitstudie die beiden größten Ängste der Deutschen.¹ Die Schere zwischen wahrgenommener und tatsächlicher Bedrohung geht dabei auseinander, aber gesellschaftliche Verunsicherung führt auch zu politischem Handlungsdruck. Ein Blick auf bestehende Förderprogramme zur Prävention von Radikalisierung und Extremismus zeigt: Der Staat und die Zivilgesellschaft unternehmen viel (Bundeszentrale für politische Bildung o.J.). Aber hilft es auch? Um Antworten auf diese Frage zu finden, kann man eine Evaluation in Auftrag geben. Doch so wie es für die Lösung gesellschaftlicher Herausforderungen kein Allheilmittel gibt, gibt es auch für Evaluationen kein Patentrezept. Unter die Bezeichnungen „Prävention von Radikalisierung“ oder „Extremismusprävention“ fallen so unterschiedliche Aktivitäten wie z. B. Medienschutz, Bildungsarbeit, Netzworkebildung und Beratungsarbeit. Hinzu kommen Ansätze zur Distanzierung von extremistischen Szenen und zur Deradikalisierung.² Vor allem die jüngst erheblich aufgestockte staatliche Förderung in diesem Bereich hat zu einer Vielzahl neuer Angebote geführt, die Kritiker und Kritikerinnen als „Wildwuchs“ (Steinke 2017) bezeichnen, während andere gerade in dieser Vielfalt ein erfolversprechendes Rezept sehen (Unger 2016; Nordbruch 2017).

Weil Evaluation keine geschützte Bezeichnung ist, kann sich hinter diesem Etikett von einem Erlebnisbericht bis hin zu einem großangelegten Forschungsprojekt alles verbergen. Evaluationen können die Machbarkeit, Prozesse, Ergebnisse, Zielerreichung und Wirkung von Präventionsansätzen untersuchen. Dabei kommen sehr unterschiedliche sozialwissenschaftliche Methoden zum Einsatz. Der richtige Methoden-Mix ermöglicht es Sozialforschern und Sozialforscherinnen, bei der Evaluation flexibel, gegenstandsangemessen und themengeleitet vorzugehen. Der Methodenpluralismus führt regelmäßig aber auch zu fachlichen Auseinandersetzungen über die Eignung von Evaluationsmethoden, insbesondere, aber nicht nur, bei Studien zur Erforschung der Wirksamkeit von Präventionsprogrammen. Hier können, selbst bei ansonsten gleichlautendem Untersuchungsziel, die Wahl der Methoden, deren Gütestandards und damit auch die erzielten Ergebnisse sehr unterschiedlich ausfallen.

Der vorliegende Report behandelt das Thema der Evaluation von Distanzierungs-, Deradikalisierungs-, und Präventionsansätzen im Bereich des politischen und religiös begründeten Extremismus und der Radikalisierung. An diesem eher abstrakten Inhalt lassen sich zwei zusammenhängende gesellschaftliche Trends beobachten. Der Trend zur „Versicherheitlichung“ und der Trend zur „Evidenzbasierung“. Den ersten Trend kennzeichnet das zunehmende Verschwimmen der Grenze zwischen zivilgesellschaftlicher Prävention und staatlicher Gefahrenabwehr, insbesondere im Bereich des religiös begründeten Extremismus.³ Zwar verüben auch Rechtsterroristen Anschläge, aber es ist vor allem das Schreckgespenst des dschihadistischen Selbstmordattentäters, das weite Teile der Gesellschaft beunruhigt und dadurch einer Entwicklung Vorschub leistet, die Radikalisierungsprävention zunehmend zu einer Frage der inneren Sicherheit macht. Im Zuge dieses Trends werden

1 <https://www.ruv.de/presse/aengste-der-deutschen/grafiken-die-aengste-der-deutschen>; 6.6.2018.

2 Zum Überblick s. Trautmann/Zick (2016) und Gruber/Lützing (2017).

3 Für einen Überblick zu jüngeren Entwicklungen s. Biene et al. (2016).

insbesondere Maßnahmen im Bereich der Deradikalisierungs- und Distanzierungsarbeit „versicherheitlich“, d. h. an den Anforderungen der Sicherheitspolitik ausgerichtet und nach deren Maßstäben gemessen. Dies verstärkt den zweiten Trend: Weil ihre Arbeit Relevanz für ein sicherheitspolitisches Thema hat, das zu den größten Sorgen der Menschen in Deutschland zählt, drängt sich den handelnden Akteuren und Akteurinnen immer stärker die Frage auf, „was wirkt?“. Ein erfolgreiches Präventionsprogramm verhindert nach dieser verkürzten Sichtweise Terroranschläge. Angesichts einer stetig größer werdenden Schutzerwartung in unserer Gesellschaft akzeptiert die Öffentlichkeit heute aber offensichtlich kaum mehr, wenn Fachleute darauf aufmerksam machen, dass es meistens keine einfachen Antworten gibt auf die Frage nach dem „was wirkt?“.

Forderungen nach Nachweisen für die Wirksamkeit von Prävention und Deradikalisierung stehen allerdings der professionellen Haltung vieler Praxisakteure entgegen, die in diesem Handlungsfeld tätig sind. Sie setzen aufgrund von langjähriger Praxiserfahrung und wissenschaftlichen Erkenntnissen über Radikalisierungsprozesse auf die Wirksamkeit ihres professionellen Handelns. Plump artikuliert Forderungen nach „Evidenzbasierung“ können außerdem von der Praxis als anmaßend und aufdringlich empfunden werden. Andere wiederum sehen in der Evidenzbasierung grundsätzlich ein legitimes und geeignetes Mittel für den gezielten Einsatz knapper Ressourcen, für die ergebnisorientierte Weiterentwicklung sowie für die Verbreitung erfolgreicher Präventionsmodelle. Schließlich lässt sich mit provokanten Forderungen nach Belegen für die Wirksamkeit staatlich finanzierter Programme vortrefflich öffentlichkeitswirksamer Druck auf die politisch verantwortlichen Akteure aufbauen. Evidenzbasierung entpuppt sich in dieser Perspektive als Multifunktionswerkzeug für die Durchsetzung unterschiedlicher Interessen. Wenn die Wissenschaft keine überzeugenden Antworten auf die Frage nach der Wirksamkeit von Prävention und Deradikalisierung liefert, gerät das ganze Praxisfeld unter Legitimationsdruck.

Vor dem Hintergrund dieser Interessenlagen erörtert der Report Möglichkeiten und Grenzen der Evaluation von Maßnahmen der Extremismusprävention und Deradikalisierung. Eine Evaluation ist ein regelgeleitetes Verfahren, in dem Daten erhoben, ausgewertet und dargestellt werden, um neue Erkenntnisse über ebensolche Maßnahmen zu gewinnen. Das Erkenntnisinteresse einer Evaluation kann sich im weitesten Sinne auf alle möglichen Eigenschaften, Voraussetzungen, Prozesse und Ergebnisse einer Maßnahme beziehen. Im Mittelpunkt dieses Reports stehen Evaluationen, die mögliche Ergebnisse von Maßnahmen der Extremismusprävention und Deradikalisierung zum Gegenstand haben. Maßnahmen, deren Wirkung auf diese Weise als abgesichert gilt, werden im Fachjargon als „evidenzbasiert“ bezeichnet.

Im wissenschaftlichen Diskurs und in Teilen der Praxis gibt es grundsätzliche Bedenken gegenüber diesem Konzept. Im Kern spiegeln diese Kontroversen unterschiedliche Denkschulen in den (Sozial-)Wissenschaften wider. Kritiker der evidenzbasierten Prävention mahnen an, dass die Wirksamkeit sozialer Programme kaum eindimensional erfassbar sei und äußern außerdem ethische, methodologische und theoretische Bedenken (Bellmann/Müller 2011). Verfechterinnen evidenzbasierter Praxis entgegnen diesen Einwänden: „[W]enn evidenzbasierte Entscheidungsfindung in der Praxis funktioniert, spielt es dann eine Rolle, dass sie in der Theorie nicht funktioniert?“ (Chalmers 2005; Übersetzung des Autorenteam). In diesem Spannungsfeld steht auch der vorliegende Re-

port. Er schlägt sich auf keine bestimmte Seite dieser kontroversen Debatte, obgleich im Mittelpunkt von Kapitel 4 Evaluationen stehen, die die erzielten Ergebnisse von Maßnahmen zum Gegenstand haben, die also auf Evidenzbasierung fokussieren. Der Report möchte dem Leser und der Leserin eine eigene aufgeklärte Meinung ermöglichen. Einige der theoretischen Gegensätze entschärfen sich tatsächlich, wenn man die Möglichkeiten und Grenzen evidenzbasierter Prävention pragmatisch und realistisch betrachtet; andere treten in der Praxis noch deutlicher zutage.

Die folgenden Kapitel setzen sich kritisch mit dem Begriff der Evidenzbasierung auseinander und beleuchten das Für und Wider ebensolcher Evaluationen. Kapitel 2 formuliert Arbeitsdefinitionen für zentrale Begriffe des Berichts: Radikalisierung, Prävention (von Radikalisierung) und Evidenzbasierung (der Prävention von Radikalisierung). Kapitel 3 nennt Praxisbeispiele für laufende und abgeschlossene Evaluationen im Bereich des Rechtsextremismus (Rex) und im Bereich des religiös begründeten Extremismus (RelEx). Die Frage nach den sichtbaren Erfolgen der evaluierten Programme stellt sich damals wie heute (Frindte/Preiser 2007). Daran anknüpfend stellt Kapitel 4 die unterschiedlichen Zugänge qualitativer und quantitativer Methoden der Datenerhebung- und -auswertung vor. Am Beispiel der „logischen Modelle“ und der *Realist Evaluation* zeigt das Kapitel Möglichkeiten auf, wie man aus der kombinierten Betrachtung qualitativer und quantitativer Daten mögliche Anhaltspunkte für die Wirkungsmechanismen präventiver Maßnahmen gewinnen kann. Kapitel 5 widmet sich der Frage nach einheitlichen Qualitätsstandards für Evaluationen. Am Ende des Textes erwarten den Leser und die Leserin praxisorientierte Empfehlungen, mit denen das Autorenteam dazu beitragen möchte, „eine begleitende Evaluationskultur zu etablieren, die Verbindlichkeit, gewisse wissenschaftliche Standards, Transparenz und auch eine Fehlerkultur vereint“ (El-Mafaalani et al. 2016: 27). Dies kann erreicht werden, indem die Auftragslagen für Evaluationen klarer artikuliert werden, als dies momentan häufig der Fall ist. Eine unklare Auftragslage kann im Ergebnis zu schlechten Evaluationen führen, die niemandem etwas nutzen, unnötige Kosten verursachen und die beteiligten Akteure frustrieren. In solchen Fällen kann der Verzicht auf eine Evaluation durchaus die bessere Wahl sein.

2. DEFINITION ZENTRALER BEGRIFFE

Dieser Text ist Teil der Reihe „Gesellschaft extrem. Radikalisierung und Deradikalisierung in Deutschland“. Die Reporte „Was ist Radikalisierung? Präzisierungen eines umstrittenen Begriffs“ (Abay Gaspar et al. 2018) und „Herausforderung Deradikalisierung: Einsichten aus Wissenschaft und Praxis“ (Baaken et al. 2018) gehen bereits ausführlich auf wesentliche Begriffe des vorliegenden Reports ein. Hier werden diese daher nur kurz aufgegriffen, um ihre gegenseitigen Bezüge zu beschreiben. Der Staat und die Zivilgesellschaft reagieren auf ein wahrgenommenes oder tatsächliches soziales Problem (Radikalisierung) mit Prävention. Darauf folgen die für Evaluation typischen Fragen wie z. B. die nach den erzielten Ergebnissen dieser Maßnahmen und Angebote.

2.1 RADIKALISIERUNG

Die weite Verbreitung des Begriffs Radikalisierung im öffentlichen Diskurs mag mitunter den Eindruck erwecken, es handle sich um einen funktionalen Begriff, der klar umrissen und nicht erklärungsbedürftig sei. Doch dies ist nicht der Fall. Radikalisierung ist ein grundsätzlich umstrittener Begriff in dem Sinne, dass darüber, was als „radikal“ gilt, was Anfangs- und Endpunkte von Radikalisierungsprozessen sind und inwieweit Gewaltbereitschaft oder Gewaltanwendung notwendigerweise Folge von Radikalisierung sind, keine Übereinstimmung herrscht (Abay Gaspar et al. 2018; Neumann 2017b: 43; Schmid 2013: 5). Für Radikalisierungsprävention und ihre Evaluation ist dieser Befund relevant, da sowohl in der Projektkonzeption als auch in der Evaluationsforschung diese Unterbestimmtheit reflektiert und gleichsam eine für den Zweck des Vorhabens nachvollziehbare, zweckmäßige und möglichst präzise Bestimmung des Präventionsgegenstandes vorgenommen werden muss. Erst diese konzeptionelle Arbeit ermöglicht es zu formulieren, was verhindert oder reduziert werden soll, um darauf aufbauend zu klären, ob und wie vorbeugende Maßnahmen Wirkungen zeitigen (können) oder nicht (Kiefer 2017; Kober 2017).

Für den Zweck des vorliegenden Reports gehen wir von den Ausführungen bei Srowig et al. (2018: 1) und Baaken et al. (2018: 7)⁴ aus und verstehen Radikalisierung als sozialen Prozess, in dessen Verlauf sich ein Mensch oder eine Gruppe extremistische Einstellungen aneignet. Als extremistisch können, mit Blick auf die Phänomenbereiche des Rechtsextremismus und des religiös begründeten Extremismus,⁵ beispielsweise Einstellungen gelten, die sich gegen die Prinzipien einer offenen und pluralistischen Gesellschaft richten und/oder Menschen die Gleichwertigkeit (aufgrund von Herkunft oder Glaube) absprechen.⁶ Die Legitimation von oder die Bereitschaft zur Gewalt kann mit Radikalisierung einhergehen, die Anwendung von Gewalt ist dabei nur eine mögliche Manifestation von Radikalisierung (Srowig et al. 2018: 1; Frindte et al. 2016: 3).

Wird Radikalisierung als *Prozess* verstanden, folgt daraus, dass es Anknüpfungspunkte auf verschiedenen Ebenen zu unterschiedlichen Zeitpunkten gibt, um Radikalisierung vorzubeugen, Menschen aus einer radikalisierten Szene zu lösen oder mit ihnen an ihrer Deradikalisierung zu arbeiten (Biene et al. 2015: 9). Die Entscheidung, welcher Zugang letztendlich gewählt wird, hängt vom jeweiligen Fall, Kontext und von den zur Verfügung stehenden Ressourcen ab und sollte idealerweise bereits in der Programmentwicklung und in der Implementationsphase mit Unterstützung einer wissenschaftlichen Begleitung reflektiert werden. Die grundlegende Erkenntnis der Radikalisierungsforschung, dass es „grundsätzlich keine Blaupause“ gebe, „die sich auf jedes Individuum und jede Gruppe übertragen lässt“ (Maurer 2018: 85) zeigt auf, dass Radikalisierung grundsätzlich „ein fragiler

4 Damit folgen wir für den Zweck dieses Reports einem engeren Verständnis von Radikalisierung, als es beispielsweise Abay Gaspar et al. (2018) in die Diskussion eingebracht haben.

5 Der Bereich der linken Militanz bleibt aufgrund der spärlichen Forschungslage, Präventionsangebote und damit einhergehenden mangelnden Evaluationen in diesem Bericht ausgespart.

6 Dieses Begriffsverständnis, nach dem Radikalisierung gerade keine Zunahme von Radikalität, sondern der oben benannten extremistischen Einstellungen und u.U. auch Handlungen beschreibt, ist angelehnt an Quent (2016) und Pfahl-Traughber (2015). Zum fragwürdigen Zusammenhang von Radikalität und Radikalisierung s. auch Schmid (2013: 3).

Prozess [ist], der von einer Vielzahl von Faktoren und Ereignissen abhängt, die von keinem Algorithmus und keiner Theorie systematisch erfasst werden können“ (Neumann 2016: 35). Diese Erkenntnis markiert dabei sowohl für Ansätze der Radikalisierungsprävention wie für deren Evaluation die Notwendigkeit einer stetigen Reflexion über die situations- und kontextabhängige Angemessenheit des eigenen Vorgehens.

2.2 PRÄVENTION VON RADIKALISIERUNG

Radikalisierungsprävention ist ein Oberbegriff für Strategien und Ansätze, die darauf abzielen, Radikalisierung zu verhindern oder abzumildern. Im deutschsprachigen Raum werden die Bezeichnungen Radikalisierungsprävention und Extremismusprävention häufig synonym verwendet, obwohl ihr jeweiliger Präventionsgegenstand (Extremismus bzw. Radikalisierung) nicht identisch ist.⁷

International variiert die Terminologie ebenfalls. Die Organisation für Sicherheit und Zusammenarbeit in Europa spricht von „Preventing Terrorism and Countering Violent Extremism and Radicalization that Lead to Terrorism“ (OSCE 2014), während in der US-amerikanischen Diskussion der Begriff Countering Violent Extremism (kurz: CVE) verwendet wird. Im Kontext der Vereinten Nationen spricht man von Preventing Violent Extremism (PVE) (vgl. Frazer/Nünlist 2015; Neumann 2017a). Auch wenn es im Detail konzeptionelle Unterschiede zwischen den damit überschriebenen Strategien geben mag, lässt sich festhalten, dass Präventionsarbeit gegen Radikalisierung auf der Handlungsebene internationaler Akteure etabliert ist.

Ansätze der Prävention unterscheiden sich hinsichtlich ihrer zeitlichen Vorverlagerung und ihrem Adressatenkreis. In diesem Zusammenhang sind die Bezeichnungen primäre, sekundäre und tertiäre (Caplan 1964) bzw. universelle, selektive und indizierte Prävention gebräuchlich (Gordon 1983).⁸ Ferner können sich alle drei Typen von Prävention personenorientiert an einzelne Individuen richten oder umweltorientiert an die Personen und ihr soziales Umfeld (Perrez 1998).

Primäre bzw. universelle Prävention richtet sich an einen breiten Adressatenkreis und versucht, unerwünschten Entwicklungen bereits zu einem sehr frühen Zeitpunkt grundlegend entgegenzuwirken. Sekundäre bzw. selektive Prävention richtet sich an Personen bzw. Personengruppen, bei denen erste Ausprägungen problematischer Entwicklungen erkennbar sind oder die als „gefährdet“ gelten, und versucht Radikalisierungsprozesse zu verhindern. Hinzu kommen Weiterbildungen für Fachkräfte, die potenziell mit gefährdeten Menschen arbeiten. Tertiäre bzw. indizierte Prävention adressiert Personen, bei denen das eigentlich zu verhindernde Problem schon ausgeprägt ist und zielt darauf

7 Zur deutschsprachigen Debatte um den Extremismusbegriff s. u.a. Bötticher (2017) und Teune (2018).

8 Obwohl die Konzepte nicht identisch sind, verwenden wir in diesem Bericht die Begriffe universelle/primäre, selektive/sekundäre und indizierte/tertiäre Prävention der Einfachheit halber synonym und schließen damit an ein weit verbreitetes Begriffsverständnis an (s. auch Trautmann/Zick 2016:7; Ceylan/Kiefer 2018). Der Unterschied zwischen beiden Konzepten liegt darin, dass Gordon mit den Bezeichnungen universell/selektiv/indiziert Bezug auf die Zielgruppe nimmt, während Caplan mit den Bezeichnungen primär/sekundär/tertiär Präventionsmaßnahmen danach unterscheidet, zu welchem Zeitpunkt sie stattfinden (s. auch Greuel 2018).

ab, ein erneutes Auftreten bzw. einen Rückfall zu verhindern. Genaugenommen handelt es sich bei tertiärer bzw. indizierter Prävention schon um eine reaktive Maßnahme, die an bereits manifesten Problemen ansetzt. Dieser terminologischen Unschärfe zum Trotz werden Demobilisierung und Deradikalisierung in der Regel dem Bereich tertiärer bzw. indizierter Prävention zugeordnet.

Ansätze der tertiären bzw. indizierten Prävention von Radikalisierung können idealtypisch in Distanzierung, Demobilisierung und Deradikalisierung unterteilt werden. Ziel von Demobilisierungsarbeit ist es, dass eine Person, die einer extremistischen Szene angehört, keine Straftaten (mehr) begeht, der Gewalt entsagt und sich aus der Szene zurückzieht (Ceylan/Kiefer 2018: 73), wobei letzterer Punkt keine notwendige Bedingung von Distanzierung bzw. Demobilisierung sein muss (Neumann 2013: 8; Köhler 2015). Entscheidend ist, dass ihr Fokus auf der Handlungsebene liegt.

Deradikalisierungsarbeit setzt auf der Einstellungsebene an. Ihr Ziel ist es – je nach Ansatz –, nicht nur die Herauslösung einer Person aus einer extremistischen Szene oder die Unterlassung von Gewalt zu erreichen, sondern auch die Ablehnung extremistischer Einstellungen (s. auch Baaken et al. 2018). Dies schließt die Auseinandersetzung mit individuellen Bedürfnissen, familiären Konfliktkonstellationen, rechtsextremistischen oder dschihadistischen Ideologien, den von ihnen geprägten individuellen Identitäten und die Möglichkeiten der sozialen (Re-)Integration in nicht-extremistische Milieus ein. Neben pädagogischen und sozialarbeiterischen Interventionen kann diese Einzelfallar-

VERSICHERHEITLICHUNG VON PRÄVENTION?

Prävention soll problematische Entwicklungen und unerwünschte Ereignisse verhindern. Angesichts der möglichen negativen Folgen präventiver Rahmungen (z. B. Stigmatisierungen von Bevölkerungsgruppen) sollte immer sorgsam abgewogen werden, ob das prinzipiell uferlose Konzept „Prävention“ angewendet werden sollte (Bröckling 2008; Lüders 2011).

Bei der Prävention gewaltbereiter Radikalisierung handelt es sich um ein Problem mit direkter sicherheitspolitischer Relevanz. Hier treten Fragen nach der Versicherheitlichung bzw. der Abgrenzung zwischen staatlichen und zivilgesellschaftlichen Akteuren besonders deutlich zutage, bspw. bei der Auseinandersetzung um das Zeugnisverweigerungsrecht von Sozialarbeitern, die mit radikalisierten Personen arbeiten. Im Bereich der primären Radikalisierungsprävention können unerwünschte Nebeneffekte auftreten, wenn dem adressierten Personenkreis ein Problem unterstellt wird, wo es evtl. noch gar keines gibt. Aus konzeptioneller und normativer Sicht stellt sich hier die Frage, wo die Prävention (von Radikalisierung) anfangen und aufhören soll (Greuel 2018). Frazer und Nünlist schlagen diesbezüglich vor:

„[es] soll konzeptionell unterschieden werden zwischen Programmen, die ‚CVE-spezifisch‘ sind, das heißt deren primäres Ziel es ist, gewalttätigen Extremismus zu verhindern, und solchen, die ‚CVE-relevant‘ sind, das heißt die nebenbei positive CVE-Effekte bewirken können. Nicht alles, was CVE-relevant ist, muss als CVE-Programm benannt werden“ (2015: 3).

beit auch psychologische Interventionen umfassen (Sischka 2018). Demobilisierungsarbeit kann dabei als ein Schritt auf dem Weg zu kognitiver Deradikalisierung verstanden werden (Baaken et al. 2018: 8).

2.3 EVALUATION UND EVIDENZORIENTIERTE PRÄVENTION VON RADIKALISIERUNG

Der Zusatz „evidenzbasiert“ verweist darauf, dass zwischen der Präventionsmaßnahme und dem anvisierten Präventionsziel (Vermeidung eines unerwünschten Zustandes bzw. Reduzierung von Risikofaktoren) ein (wissenschaftlich) nachweisbarer Wirkungszusammenhang bestehen sollte.

Theorie, Forschung, Evaluation und die praktische Umsetzung ihrer Erkenntnisse machen Radikalisierungsprävention zu evidenzbasierter Radikalisierungsprävention.⁹ In diesem Zusammenhang meint evidenzbasiert, dass Maßnahmen zur Verhinderung von sozialen Problemen auf wissenschaftlicher Grundlage konzipiert, implementiert und evaluiert werden (Beelmann 2017: 14–15). Neben der theoretischen und empirischen Grundlagenforschung über die Ursachen sozialer Probleme (hier: Radikalisierung) ist demnach die Evaluation fester Bestandteil evidenzorientierter Präventionsansätze.

Was Evidenzbasierung darüber hinaus heißt und wie sie bewertet wird, ist strittig.¹⁰ In der kontroversen Debatte können idealtypisch drei Positionen ausgemacht werden: Die erste gibt ausschließlich den Ergebnissen aus Evaluationsstudien mit randomisierten Kontrollgruppen den Status „Evidenz“ und schließt andere Methoden zur Generierung von Wissen aus. Konträr dazu verläuft jene Position, die Evidenzbasierung anhand von Kontrollgruppenstudien grundsätzlich ablehnt. Dazwischen finden sich vielfach nuancierte Positionen, die den dogmatischen Ansatz der Evidenzbasierung ablehnen und stattdessen nach pragmatischen Alternativen suchen. Ein solches Modell schlagen de la Chaux et al. (2018) in Ergänzung zu Beelmann (2011) vor. Sie empfehlen die „Triangulation von Evidenzen“, d. h. die Berücksichtigung von Befunden aus qualitativen und quantitativen Studien, sofern sich daraus Erkenntnisse über die Funktionsweise von Radikalisierungsprävention gewinnen lassen.

Der vorliegende Report kann keinen erschöpfenden Überblick über die verschiedenen Formen der Evaluation geben, sondern konzentriert sich stattdessen auf eine ihrer Spielarten, die Ergebnisevaluation oder Wirkungsevaluation. Diese stehen im engen Zusammenhang mit der häufig artikulierten Forderung nach Evidenzbasierung in der Prävention. Mit dieser Themensetzung soll aber nicht verschwiegen werden, dass andere Formen der Evaluation, z. B. vor der Implementierung einer Maßnahme zur Unterstützung der Programmentwicklung (proaktive Evaluation), während der Umsetzung als Beratungsfunktion (Prozessevaluation) oder auch als vergleichende Überblicksarbeit mehrerer Evaluationsstudien (*Systematic Review*) eine ebenso große Bedeutung für die Entwicklung und Qualitätssicherung von Ansätzen zur Prävention von Radikalisierung haben und zudem eine Wirkungsevaluation sinnvoll ergänzen können (z. B. Widmer et al. 2009).

9 Zur Kontroverse um den Evidenzbegriff s. etwa Kap. 2, S. 3.

10 S. Positionspapier der Gesellschaft für Evaluation e. V. (DeGEval) Evidenz und Evaluation: https://www.degeval.org/images/stories/Publikationen/Positionspapier_Evaluation_-_Evidenz_-_Effekte.pdf; 14.6.2018.

3. STAND DER EVALUATION VON RADIKALISIERUNGSPRÄVENTION

Nach der Einführung in die drei Themenbereiche Radikalisierung, Prävention (von Radikalisierung) und Evaluation (von Radikalisierungsprävention) widmet sich dieses Kapitel nun konkreten Praxisbeispielen. Es benennt laufende und abgeschlossene Evaluationsaktivitäten in den beiden Bereichen Rex (Rechtsextremismus) und RelEx (Religiös begründeter Extremismus), bei denen sich damals wie heute (umstrittene) Fragen nach der Wirksamkeit von Maßnahmen stellen.

In Deutschland hat sich ein bestimmtes Verständnis von evidenzbasierter Präventionspraxis entwickelt, das das Resultat einer längeren Entwicklung auf nationaler und internationaler Ebene ist. Entscheidende Wegmarken sind z. B. die Aufforderung des US-amerikanischen Kongresses, einen Bericht über die Effektivität von Programmen zur Prävention von Kriminalität zu erstellen. Der sogenannte *Sherman-Report* (Sherman et al. 1998) hat auch die Prävention in Deutschland maßgeblich beeinflusst und zur Verbreitung des evidenzbasierten Paradigmas beigetragen. Heute stehen im Mittelpunkt dieser Strategie die Entwicklung von Qualitätsstandards für Evaluationen und Programmplanungen, Metaanalysen und Überblicksarbeiten (z. B. die *Systematic Reviews* der *Campbell Collaboration*)¹¹ sowie die Bestimmung von *Best-Practice*-Modellen und deren Verbreitung durch entsprechende Datenbanken (z. B. die *Blueprints for Healthy Youth Development* des *Center for the Study and Prevention of Violence* der Universität Boulder in Colorado).¹² Ein Pionier evidenzbasierter Ansätze in Deutschland war der Landespräventionsrat Niedersachsen, der durch die Verbreitung der *Beccaria Standards*¹³ und anhand der Grünen Liste Prävention¹⁴ das Konzept in Politik und Praxis getragen hat.

Auch im Feld der Radikalisierungsprävention wird der evidenzbasierte Ansatz heute zunehmend adaptiert (Lindekilde 2012). Das 2012 von der Europäischen Kommission gegründete Expertennetzwerk *Radicalisation Awareness Network* (RAN)¹⁵ sucht und erarbeitet europaweit *Best-Practice*-Ansätze in diesem Feld. Madriaza und Ponsot erstellten 2015 eine der wenigen Überblicksarbeiten zu internationalen Präventionsansätzen gegen Radikalisierungsprozesse (Madriaza/Ponsot 2015); Christmann (2012) vergleicht entsprechend evaluierte Ansätze in Großbritannien und Litmanovitz et al. legten 2017 den ersten *Systematic Review* für die *Campbell Collaboration* zu den Risikofaktoren einer Radikalisierung mit Ableitungen für die Prävention vor (Litmanovitz et al. 2017). Handreichungen wie das *Program Evaluation Toolkit for Countering Violent Extremism* der *RAND Corporation* (Helmus et al. 2017) oder das *Online Evaluation Toolkit* des EU-Forschungskonsortiums IMPACT Europe¹⁶ zielen ebenfalls auf die Messung von Präventionseffekten ab.

11 <https://www.campbellcollaboration.org>; 5.2.2018.

12 <http://www.blueprintsprograms.com>; 5.2.2018.

13 https://www.beccaria.de/nano.cms/de/Beccaria_Standards1; 7.6.2018.

14 <http://www.gruene-liste-praevention.de/nano.cms/datenbank/information>; 7.6.2018.

15 https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network_en; 16.4.2018.

16 <http://impacteurope.eu>; 4.7.2018.

Gerade zwischen europäischen Ländern und Kulturen sind Möglichkeiten des Vergleichens und des Übertragens sowie die Bestimmung des Erfolges entsprechender Ansätze für die beteiligten Länder von großem Interesse. Auch für diesen Zweck stellen Evaluationen eine wichtige Grundlage dar. Der folgende Abschnitt geht auf die verschiedenen Evaluationsaktivitäten in Deutschland ein.

3.1 EVALUATION IM BEREICH DER RADIKALISIERUNGSPRÄVENTION REX

Im Bereich der Extremismus- bzw. Radikalisierungsprävention wird in Deutschland seit einigen Jahren über Strategien der Evaluation diskutiert (vgl. z. B. Lynen von Berg/Roth 2003; Strobl et al. 2012). Ebenso vielfältig wie die Programme selbst sind die Ansätze für ihre Evaluation. Der folgende Abschnitt richtet den Blick vorrangig auf die Bundesprogramme des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) und des Bundesministeriums des Innern, für Bau und Heimat (BMI). Aber auch auf Landesebene und auf regionaler Ebene gibt es zahlreiche Programme und Projekte, zu denen mittlerweile Ergebnisse aus Evaluationen und wissenschaftlichen Begleitungen vorliegen.

Schon das erste große Bundesprogramm nach der Wiedervereinigung „Aktionsprogramm gegen Aggression und Gewalt“ (AgAG), welches zwischen 1992 und 1996 in ostdeutschen Schwerpunktregionen 124 Projekte förderte, stellte Mittel für eine wissenschaftliche Begleitung bereit, die durch das Institut für Sozialpädagogik und Sozialarbeit an der TU Dresden durchgeführt wurde (Bohn/Münchmeier 1997). Dabei ist jedoch zu berücksichtigen, dass das Programm kein reines Präventionsprogramm war, sondern zugleich auch infrastrukturfördernde Aufgaben in den damaligen neuen Bundesländern zu erfüllen hatte. So verweist Möller (2003) darauf, dass die Evaluation eher auf infrastrukturelle Effekte als auf Aussagen zu den konkreten Wirkungen (vor allem Reduzierung von Gewalt und Aggression) ausgelegt war. Für ihn sprechen damals steigende Zahlen von rechtsextremer Gewalt und von Mitgliedschaften in einschlägigen Gruppierungen in Deutschland „deutlich gegen durchschlagende und nachhaltige extremismusreduzierende Wirkungen von AgAG“ (Möller 2003: 28).

Auch das zwischen 2001 und 2006 durchgeführte Bundesprogramm „Jugend für Toleranz und Demokratie – gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus“, welches deutschlandweit in drei Programmteilen (ENTIMON, CIVITAS und XENOS (letzteres verlängert bis 2013)) 4.500 Projekte, Initiativen und Maßnahmen zur Stärkung der Zivilgesellschaft in Höhe von rund 192 Millionen Euro förderte, beinhaltete Mittel für breit angelegte Evaluations- und Monitoringstrategien. So setzte das Deutsche Jugendinstitut (DJI) für die Evaluation des Programmteils ENTIMON auf eine Strukturdatenanalyse aller geförderten Teilprojekte und auf eine qualitative Analyse ausgewählter Projekte zur exemplarischen Bestimmung innovativer Strategien (Klingelhöfer et al. 2007). Dabei wurde als Evaluationsinstrument auf das Konzept der logischen Modelle zurückgegriffen (s. Kapitel 4.3).

Auch die beiden Bundesprogramme „kompetent. für Demokratie – Beratungsnetzwerke gegen Rechtsextremismus“ (Auf- und Ausbau einer Beratungsinfrastruktur) und „VIELFALT TUT GUT. Jugend für Vielfalt, Toleranz und Demokratie“ (Lokale Aktionspläne und Modellprojekte), die vom BMFSFJ zwischen 2007 und 2010 umgesetzt wurden, sowie deren Nachfolgeprogramme „TOLERANZ

FÖRDERN – KOMPETENZ STÄRKEN“ (Beratungsnetzwerke, Lokale Aktionspläne und Modellprojekte) (Bischoff et al. 2015) und die „Initiative Demokratie Stärken“ zwischen 2011 und 2014 (BMFSFJ 2014) setzten entsprechend der Vielfalt der Programmbereiche auf unterschiedliche evaluative Zugänge und Evaluatoren.

Die Evaluation des Gesamtprogramms wurde vom DJI durchgeführt (Bischoff et al. 2011). Neben Dokumentenanalysen, der Erhebung von Monitoringdaten, Befragungen externer Expertinnen und Experten (Außenperspektive) und an den Projekten beteiligter Fachkräfte (Innenperspektive) wurden u.a. Vorher-Nachher-Befragungen von Adressatinnen und Adressaten eines Modellprojekts in einem quasiexperimentellen Längsschnittdesign als eine Form der Wirkungsanalyse durchgeführt. Aufgrund der Ergebnisse der Programmevaluation konnten schon während der Programmdurchführung sowohl bestimmte Anpassungsimpulse gegeben (Prozessevaluation) als auch Empfehlungen zur Weiterentwicklung beider Programme erarbeitet werden (Bischoff et al. 2011). Das aktuelle Programm des BMFSFJ, „Demokratie leben!“¹⁷ wie auch das Programm „Zusammenhalt durch Teilhabe“ des BMI werden ebenfalls wissenschaftlich begleitet.

Gerade im Bereich der Bundesprogramme zeigt sich insgesamt betrachtet die ganze Vielfalt und Komplexität des Feldes: So sind Bundesprogramme zur Demokratieförderung, zur Radikalisierungs-, Gewalt- und Vorurteilsprävention dazu konzipiert, Strukturen erst auszubilden und didaktische Ansätze sowie Zielgruppenzugänge und Kooperationsmöglichkeiten zu allererst zu entwickeln und zu erproben. Insofern stellen sie einen Evaluationsgegenstand dar, dessen Erfolg nicht vorrangig beispielsweise in der breitenwirksamen Reduktion von Vorurteilen (oder gar Terroranschlägen) liegt, sondern in der Entwicklung von Ansätzen der Demokratieförderung und Radikalisierungsprävention, die in die Regelpraxis der Kinder- und Jugendhilfe übertragbar sind, sowie von entsprechenden Strukturen auf Bundes-, Länder- und kommunaler Ebene. An diese Ausgangslage schließen sich Fragen über die Aufgaben und die Funktion von Evaluation an: Was genau soll untersucht werden, an welchen Kriterien soll Erfolg bzw. Misserfolg gemessen werden, welche Kausalitäten bzw. Ursache-Wirkungszusammenhänge werden angenommen und (wie) können die Folgen pädagogischer Arbeit in der Biografie von Kindern und Jugendlichen untersucht werden?

3.2 EVALUATION IM BEREICH DER RADIKALISIERUNGSPRÄVENTION RELEX

Radikalisierungsprävention im Bereich RelEx ist ein relativ junges Arbeitsfeld, in dem es bisher entsprechend wenige Evaluationsstudien gibt. Ohne Anspruch auf Vollständigkeit werden in diesem Abschnitt einige der aktuell vorliegenden Evaluationen vorgestellt.

Im Jahr 2017 wurde der Evaluationsbericht über die Arbeit der im Bundesamt für Migration und Flüchtlinge (BAMF) ansässigen Beratungsstelle „Radikalisierung“ und ihrer vier zivilgesellschaftli-

17 Für mehr Informationen zu den Modellprojekten von „Demokratie leben!“, die sowohl im Bereich Rex als auch RelEx angesiedelt sind, s. Birtsch et al. (2016).

chen „Partner vor Ort“ auf Bundesebene veröffentlicht.¹⁸ Dabei handelt es sich um keine Wirkungs-, sondern um eine Prozess- und Netzwerkanalyse sowie eine Analyse des logischen Rahmens der Arbeit der Beratungsstellen (Uhlmann 2017: 11). Eines der zentralen Erkenntnisse ist, dass das Feld der Beratungsarbeit sich stetig professionalisiert – u.a. durch die Entwicklung einer Standardsystematik im Rahmen eben dieser Evaluation, es jedoch einer weiteren Systematisierung von „Standards, Deradikalisierungsindikatoren und Metriken zur Erfolgsmessung“ (Uhlmann 2017: 52) bedarf. Darauf aufbauend sei eine umfassende Wirkungsanalyse der Beratungsarbeit erst möglich.

Anfang des Jahres 2018 wurde der Bericht der Evaluation der Beratungsstelle Hessen des *Violence Prevention Network (VPN)* veröffentlicht. Ziel der formativen Evaluation war, „den Kenntnisstand über die Adressatengruppierungen der Radikalisierungsgefährdeten sowie der bereits Radikalisierten zu erweitern“ und „Erfolg versprechende Möglichkeiten der Auslösung und stabilisierenden Begleitung von Distanzierungsprozessen zu identifizieren“ (Möller/Neuscheler 2018: 4). Auch im Rahmen dieser Evaluation wurde die Professionalisierung vorangetrieben, indem Evaluierende und Evaluierte gemeinsam eine Zielsystematik entwarfen, die in der Folge ermöglichen kann, Wirkungen zu identifizieren und zu überprüfen (Möller/Neuscheler 2018: 12). Des Weiteren wurde eine Prozessanalyse sowie eine Analyse des strukturellen Rahmens der Beratungsstelle durchgeführt (Möller/Neuscheler 2018: 14–23), die in der Gesamtsicht eine positive Einschätzung der Beratungsarbeit zeichnen. Evaluationen weiterer Beratungsstellen sowie Präventionsprojekte laufen zum Zeitpunkt der Erstellung dieses Textes noch oder befinden sich in der Abstimmung zwischen den Akteuren. Es ist daher davon auszugehen, dass der Kenntnisstand bezüglich Struktur- und Prozesslogiken sowie der Wirksamkeit von Präventions-, Demobilisierungs- und Deradikalisierungsarbeit auch im Phänomenbereich RelEx in den kommenden Jahren sukzessive ausgebaut wird.

ZWISCHENFAZIT

Trotz der insgesamt regen Evaluationsaktivitäten herrscht nach wie vor ein Mangel an wissenschaftlichen Strategien im Bereich der Extremismus- bzw. Radikalisierungsprävention. Grund hierfür ist, trotz einiger Fortschritte und positiver Beispiele (s. Kapitel 3.1 und 3.2), nach wie vor ein Mangel an entsprechenden Evaluationen. An anderen Stellen mangelt es an Transparenz, denn mitunter bleiben die Ergebnisse aus Evaluationen unveröffentlicht.¹⁹

Die Diskussion um die wissenschaftlichen Begleitungen und Evaluationen der Bundesprogramme sowie der Programme und Projekte auf Landes- und regionaler Ebene im Bereich der Extremismusprävention und Demokratieförderung spiegeln nicht zuletzt innerakademische Paradigmen und Kontroversen wider. Dabei wird insbesondere vonseiten der Protagonisten statistisch angelegter,

18 Die vier „Partner vor Ort“ sind die Beratungsstelle Hayat des Trägers ZDK Gesellschaft Demokratische Kultur gGmbH, das Beratungsnetzwerk Grenzgänger des Trägers IFAK e.V., die Beratungsstelle kitab des Trägers Verein zur Förderung akzeptierender Jugendarbeit e.V. (Vaja) und das *Violence Prevention Network* e.V..

19 Zu berücksichtigen bleibt, dass uneingeschränkte Transparenz einem Vertrauensverhältnis zwischen Evaluation und Evaluierten im Wege steht. Letztlich kann uneingeschränkte Transparenz sogar wiederum zu Intransparenz führen, wenn etwa die Evaluierten aus Angst vor der Veröffentlichung von Fehlern und Schwierigkeiten ihren Evaluatoren gegenüber ausschließlich über eigene Erfolge berichten.

FALLBEISPIEL: DIE EVALUATION DES PROJEKTS TEACH2REACH

Teach2Reach ist ein Projekt des *Violence Prevention Network* e.V. (VPN) im Bereich der Primärprävention von Radikalisierung in religiös begründeten Extremismus. Es handelt sich um eine Reihe von Workshops zu den Themen „Islam“, „Interreligiösität“ und „Nahost-Konflikt“, die im Schulkontext von VPN-Trainerinnen und -Trainern in Kooperation mit Lehrkräften durchgeführt werden. Finanziert wird das Projekt seit 2016 von der Landeskommision Berlin gegen Gewalt der Senatsverwaltung für Inneres und Sport, die die Arbeitsstelle Jugendgewaltprävention der Camino-Werkstatt mit der Evaluation des Projekts beauftragte.

Evaluationsziele

Die Evaluation von *Teach2Reach* war explizit auf die Analyse der Zielerreichung und Wirksamkeit des Projekts ausgerichtet. Dazu wurde erfasst, in welchem Umfang der Workshop von den Teilnehmenden angenommen wurde. Außerdem wurde untersucht, ob der Workshop bei den Teilnehmenden Veränderungen persönlicher Eigenschaften in Gang gesetzt hat, von denen angenommen wird, dass sie gegen Radikalisierung schützen (z. B. Wissenszuwachs, Steigerung von Dialogfähigkeit, Einstellungsveränderungen). Die Evaluation untersucht auch, welche Programmmerkmale förderlich und hinderlich für die Zielerreichung waren, um dadurch eine bessere Erkenntnis über die jeweiligen Wirkungsmechanismen zu erlangen.

Kontextuelle Herausforderungen und Methoden

Da das Projekt einen wichtigen Teil des neu aufgelegten „Landesprogramms Radikalisierungsprävention“ ausmacht, wollten die Verantwortlichen möglichst früh Anhaltspunkte über dessen Wirksamkeit gewinnen. Mit relativ begrenzten finanziellen und zeitlichen Ressourcen für die Evaluation konnte mittels eines standardisierten Fragebogens eine Vorher-Nachher-Befragung der Teilnehmenden realisiert werden. Zusätzlich wurde eine Gruppe befragt, die den gleichen Workshop vor einigen Monaten absolviert hatte (Nachhaltigkeit).

Erhoben wurden sowohl die subjektive Zufriedenheit mit den Workshops und die Einschätzung der eigenen Kompetenzen (nur nachher) als auch generelle Einstellungen z. B. zur Religion und zum Zusammenleben sowie zu Vorurteilen (vorher und nachher). Die Evaluierenden verwendeten für den Fragebogen bereits geprüfte Items aus renommierten Studien, die zum Teil sprachlich oder aufgrund der Workshopziele angepasst wurden. Die Evaluation integrierte auch partizipative Verfahren und bezog die Mitarbeiterinnen und Mitarbeiter von VPN mit ein. So wurde zu Beginn eine Zielexplication mit ihnen durchgeführt, um zentrale Wirkannahmen des Projekts zu identifizieren.

Kritisch reflektiert wird von den Evaluierenden, inwieweit Effekte zweifelsfrei der Maßnahme zugeschrieben werden können bzw. Einflüsse externer Faktoren angenommen werden müssen. Der Umstand, dass die Anzahl der Teilnehmenden in der zur Verfügung stehenden Zeit begrenzt war und keine Kontrollgruppen gebildet werden konnten, ist mit Blick auf die Wirkungsmessung ebenfalls zu berücksichtigen. Als alternative Methode, um eine bessere Kontextualisierung und eine „Absicherung“ der quantitativen Ergebnisse aus den Fragebögen zu ermöglichen, wurden diese mittels qualitativer Daten aus leitfadengestützten Gruppendiskussionen und Einzelinterviews mit Lehrkräften und Trainerinnen kreuzvalidiert.

Eine weitere Herausforderung stellte die tiefergehende Prüfung der im Fragebogen verwendeten Messinstrumente und die Entwicklung von niedrigschwelligen und methodologisch tragfähigen Instrumenten bzw. Ansätzen dar, um die Einstellungen und Meinungen der Schülerinnen und Schüler, und davon insbesondere jene in sozial-schwachen Kontexten, zu erfassen. Hier zeigt sich der nach wie vor bestehende Bedarf an der Entwicklung von Instrumenten zur Messung von „Wirkung“ oder „Erfolg“ - hier zum Beispiel von Einstellungsänderungen - im Kontext von Radikalisierungsprävention.

Anschlussfähigkeit – So what?

Die Evaluation zeigt im Ergebnis, das von Teach2Reach wünschenswerte Effekte ausgehen. Diese Ergebnisse wurden in erweiterten Gremien und Fachressorts vorgestellt und die Evaluierenden wurden in die strategische Weiterentwicklung des übergeordneten Programms, unter dem das Projekt gefördert wird, involviert. Dieser vertrauensvolle Austausch ist wohl u.a. einem langjährigen guten Verhältnis zwischen Förderer und evaluierender Institution zu verdanken und daher nicht unbedingt selbstverständlich.

Der durch die Studie gewonnene Einblick in die Prozesse, Strukturen und Inhalte der Workshops trug zu einem vertieften Verständnis der inhärenten Projektlogik bei den Fördergebern bei und unterstützte damit etwa die Weiterentwicklung der „fachlichen Identität“ von (Einzel-)Projekten und Landesprogrammen. So wurden die aus dieser Evaluation gewonnenen Erkenntnisse über *Best Practices* z. B. an ähnliche Projekte im Landesprogramm mitgeteilt. Weiterhin wurden einzelne Empfehlungen aus der Evaluation durch die Fördergeber umgesetzt.

Inwieweit die erfolgte Diskussion der Evaluationsergebnisse und -empfehlungen im Evaluationsbericht sowie durch die Fördergeberin unter den Projektmitarbeiterinnen und -mitarbeitern im Sinne einer reflexiven Fehlerkultur auch die weitere Umsetzung des Projekts beeinflusste, muss an dieser Stelle offen bleiben. Das partizipative Vorgehen während der Evaluation (z. B. durch die Zielexplication) dürfte jedoch Reflexionsprozesse in Gang gesetzt haben, sowohl auf organisatorischer als auch auf praktischer/pädagogischer Ebene.

Darstellung auf Grundlage von Glock/Lüter 2017 sowie aus persönlichen Gespräch mit Dr. Lüter und Frau Vialet (Landeskommission Berlin gegen Gewalt) (Juni 2018).

kontrollgruppenbasierter und darauf aufbauend evidenzorientierter Forschung der Mangel an einschlägigen Studien – gemessen an den damit einhergehenden Kriterien – beklagt. So kommen Frindte und Preiser (2007: 35) zu dem Ergebnis, zahlreiche Projekte stünden „bislang auf instabilem Fundament. Ihre Wirksamkeit war häufig weder theoretisch begründet noch empirisch überprüft“.

Eine Bestandsaufnahme des Bundeskriminalamtes (BKA) von insgesamt 721 in 2014/2015 realisierten Projekten der Extremismusprävention aus den drei Phänomenbereichen des Rechts-, Links-extremismus und Islamismus kommt bezüglich der Frage nach Evaluation zu dem Ergebnis: „Die vorliegenden Informationen zu Evaluationsmaßnahmen müssen sowohl hinsichtlich Quantität als auch Qualität als äußerst dürftig bezeichnet werden“ (Lützinger/Gruber 2017: 20). Diese Erhebung liefert außerdem wichtige Anhaltspunkte über das Selbst- und Fremdverständnis von Evaluation. So wird deutlich, dass die Evaluationsforschung in der Trägerlandschaft immer noch auf teilweise berechtigte Skepsis gegenüber der Objektivität, dem Nutzen, der Aussagekraft und der Durchführbarkeit von Evaluationsstudien stößt. Bedenken bestehen dabei insbesondere gegenüber einer Legitimitäts- und Kontrollfunktion externer Evaluation. Dagegen steht die Trägerlandschaft Evaluationsvorhaben aufgeschlossen gegenüber, wenn sie die Möglichkeit eröffnen, das eigene professionelle Handeln zu reflektieren und zu entwickeln (Lützinger/Gruber 2017: 19–20). Der mögliche Nutzen von Begleitforschung kann zusätzlich erhöht werden, wenn „das umfassende Repertoire an Erhebungsmethoden eine Option [eröffnet], Erfolge auch jenseits von Fallzahlen darzustellen und in ihrer Ganzheitlichkeit zu würdigen“ (Lützinger/Gruber 2017: 20).

Zu ähnlichen Ergebnissen kommt auch eine Bestandsaufnahme des Instituts für Konflikt- und Gewaltforschung der Universität Bielefeld. Trautmann et al. (2017: 3) resümieren für die in Deutschland angebotenen und durchgeführten Präventionsprogramme gegen islamistisch motivierte Radikalisierung: „Die positiven Effekte lassen sich häufig nur schwer messen und der Evaluationsansatz trifft zuweilen auf Befindlichkeiten der Akteure. Mancherorts fehlt es an ausreichenden Angeboten oder aber die Qualität ist fraglich“. Auch Kober (2017: 233) kommt in einer Übersichtsarbeit zur Evaluation in der Prävention von islamistischem Extremismus zu dem Schluss: „Es können daher in Deutschland keine Studien oder Evaluationsberichte identifiziert werden, die unter Anwendung quantitativer und/oder qualitativer Methoden, statistisch belastbare Aussagen zur Wirksamkeit von Maßnahmen oder Projekten der Prävention religiöser Radikalisierung zulassen“.

Trotz dieses teilweise defizitären Forschungsstandes lässt sich ein klarer Trend hin zu einer besseren Wissensbasierung erkennen. So ist die Idee der Wissens- und Evidenzbasierung z. B. im aktuellen „Nationalen Präventionsprogramm gegen islamistischen Extremismus“²⁰ fest verankert mit dem Ziel, die „Wirksamkeit der Extremismusprävention zu erhöhen“ (5). Deutschland verfügt über leistungsstarke Forschungsstrukturen und damit über eine wesentliche Voraussetzung, um soziale Probleme wissens- und evidenzbasiert zu adressieren und gegenstandsangemessene Evaluationen durchzuführen.

20 https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/praeventionsprogramm-islamismus.pdf?__blob=publicationFile&v=2; 4.7.2018.

4. ANSÄTZE VON WIRKUNGSEVALUATION

Selbst für Experten und Expertinnen ist die Bandbreite unterschiedlicher Evaluationsansätze bisweilen schwer überschaubar. Evaluation ist keine geschützte Bezeichnung. Sie wird mitunter „inflationär verwendet“ (Döring/Bortz 2016: 978) und reflexartig gefordert, ohne dass dabei immer von vornherein klar ist, wer wen oder was, nach welchen Kriterien, mit welchem Erkenntnisinteresse und anhand welcher Untersuchungsmethoden bewertet.

Für die Untersuchung der Effekte von Ansätzen der Deradikalisierung, Demobilisierung und Radikalisierungsprävention gibt es keine Blaupause. Der Prototyp der Wirksamkeitsforschung, d. h. das Experiment mit Versuchs- und Kontrollgruppe, lässt sich nicht immer sinnvoll auf die Untersuchung sozialer Programme übertragen. Experten und Expertinnen verschiedener Fachrichtungen suchen daher nach alternativen Ansätzen, um zu untersuchen, welche Wirkungen von „komplexen Interventionsprogrammen in der Prävention“ ausgehen (Robert Koch-Institut 2012). Das Problem ist nicht neu. Eine Studie der Bertelsmann Stiftung von 2004 mit dem Titel „Evaluation politischer Bildung – ist Wirkung messbar?“ widmet sich der Frage „[...] ob man solche komplexen Veränderungsprozesse überhaupt erfassen, ja gar ‚messen‘ kann“ (Uhl et al. 2004: 9–10).

Um eine Wirkungsevaluation durchzuführen, benötigt man ein Bündel an beobachtbaren Merkmalen, auf deren Beeinflussung das Programm abzielt (sog. Indikatoren). Köhler erläutert: „[D]ie Möglichkeit, gesetzte Ziele mit der Programmorganisation und den letztendlich erreichten Ergebnissen abzugleichen, macht eine Evaluation erst möglich“ (2016: 25). Woran erkennt man, ob ein Programm oder eine Maßnahme Radikalisierung erfolgreich verhindert bzw. entsprechende Entwicklungen rückgängig gemacht hat? Klar ist: Einen Standardindikator, der uneingeschränkt für das gesamte Spektrum der Radikalisierungsprävention gültig wäre, kann es nicht geben.

Konkret für den Bereich der Deradikalisierung sind folgende grobe Zielkategorien in der Diskussion (Neumann 2013, Köhler 2015): „Demobilisierung“, „Ausstieg“, „Integration“ und „Stabilisation“. Auf der ersten Stufe steht der Verzicht auf Gewalt zu politischen Zwecken, während die letzte Stufe eine vollständige Deradikalisierung auf der kognitiven und affektiven Ebene umfasst. Präventionsarbeit (hier präziser: Deradikalisierungsarbeit) ist demnach in dem Maße erfolgreich, in dem es ihr gelingt, Personen durch eine intensive Fallberatung dazu zu verhelfen, von einer höheren auf eine niedrigere Radikalisierungsstufe zu wechseln.

Diese Zielkategorien können eine wertvolle Orientierungshilfe im Bereich der tertiären Prävention und deren Evaluation sein. Sie sind aber kein sinnvoller Maßstab für Präventionsmaßnahmen, die zu einem sehr viel früheren Zeitpunkt im Prozess der Radikalisierung ansetzen. Für diesen primären Bereich benennt z. B. Beelmann (2017) in den „Grundlagen einer entwicklungsorientierten Prävention des Rechtsextremismus“ als Ziele insbesondere den Abbau von Vorurteilsstrukturen. Dazu empfiehlt er pädagogische Ansätze, wie z. B. Antirassismusprogramme, Diversity Trainings sowie ein Bündel an Maßnahmen aus dem Bereich der Vorurteilsprävention. Ein Teil dieser Ansätze ist bereits sehr gut erforscht und es liegen entsprechend validierte Indikatoren für eine gegenstandsangemessene Evaluation vor.

Die Auswahl und die Gestaltung (die sog. Operationalisierung) geeigneter Erfolgskriterien ist ein kritischer Schritt jeder Wirkungsevaluation. Sie sollten insbesondere durch konkrete Zielbestimmungen durch die jeweiligen Akteure und Akteurinnen vor Maßnahmenstart festgelegt und sorgfältig abgestimmt sein, um zu gewinnbringenden Ergebnissen zu kommen, die insbesondere auch der Erkenntnisfunktion einer Evaluation gerecht werden. Wenn fest steht, welche Änderungsprozesse im Rahmen einer Evaluation näher untersucht werden sollen, dann schließt sich daran die Frage nach dem forschungspraktischen Vorgehen der Studie an.

Die folgenden Abschnitte beschreiben sozialwissenschaftliche Forschungsmethoden, anhand derer Erkenntnisse über die Wirksamkeit und Wirkungsmechanismen von sozialen Interventionen und Programmen gewonnen werden können. Anhand dieser Beschreibung soll der Leser bzw. die Leserin nachvollziehen, wie die Resultate in einem Evaluationsprozess entstehen, welche Rückschlüsse sie (nicht) zulassen, welche Reichweite sie haben und wie belastbar sie sind.

4.1 (QUASI-)EXPERIMENTELLE STUDIEN UND FELDVERSUCHE

Das Ziel einer experimentellen Studie ist es, einen kausalen Zusammenhang zwischen Ursache und Wirkung nachzuweisen. Sie gilt in der Wissenschaft häufig als der „Königsweg der Erkenntnis“ (Döring/Bortz 2016: 194), weil sie unter idealen Voraussetzungen alle möglichen Störfaktoren ausschließt, die das Ergebnis beeinflussen könnten. An dem Ergebnis einer experimentellen Studie (den sogenannten Effektstärken) kann man ablesen, ob eine Maßnahme einen bestimmten Einfluss auf ihre Teilnehmerinnen und Teilnehmer hat, und wie groß der Effekt (im Vergleich zu einer Kontrollgruppe) ausfällt. In seiner rudimentärsten Form (einer sogenannten *Black-Box-Evaluation*) bringt das Experiment, neben der Erkenntnis, dass das untersuchte Programm wirkt, keine weitere Erkenntnis bezüglich der Fragen, warum, auf welche Weise und für wen es wirkt. Weil dies in aller Regel kein befriedigendes Ergebnis ist, gibt es eine Reihe von Variationen experimenteller Studien (z. B. faktorielle Designs, Döring/Bortz 2016: 194), die zu differenzierten Ergebnissen kommen, indem sie Variationen der Maßnahme zwischen mehreren Teilnehmergruppen miteinander vergleichen.

Bei ausreichend großer Stichprobe (Teilnehmerzahl) können außerdem mittels statistischer Auswertungsverfahren, wie z. B. der Mehrebenenanalyse, im Nachhinein differenzierte Effekte für spezifische Teilgruppen bestimmt werden. Man kann dann etwa ablesen, ob ältere Teilnehmende stärker von dem Programm profitieren als jüngere oder umgekehrt. Wenn die Studie darüber hinaus an mehreren Standorten (d. h. multizentrisch) durchgeführt wurde, erlaubt dies weitere wertvolle Rückschlüsse darüber, wie Präventionsprogramme in den jeweiligen, z. B. sozialräumlichen Kontexten, in die sie eingebettet sind, wirken.

Ein entscheidender Nachteil experimenteller Studien ist die mangelnde Verallgemeinerbarkeit ihrer Ergebnisse. Sofern bei der Auswahl der Teilnehmenden nicht entsprechende Strategien der Stichprobenziehung berücksichtigt werden, so wie sie etwa bei repräsentativen Bevölkerungsumfragen eingesetzt werden, sind die Ergebnisse ausschließlich für die kleine Gruppe der Versuchsteilnehmer und -teilnehmerinnen gültig. Großangelegte Feldversuche in der Präventionsforschung (Eisner

et al. 2012), Längsschnitt-/Panelstudien und „natürliche Experimente“, bei denen sich Unterschiede zwischen Experimental- und Kontrollgruppe durch natürliche Begebenheiten ergeben, sind Möglichkeiten, diesen Mangel zu kompensieren. Solche Studien sind allerdings aufwändig und ihr Einsatz muss aus forschungsökonomischer Sicht im richtigen Verhältnis zu der Größe des zu evaluierenden Programms stehen.

4.2 QUALITATIVE EVALUATIONSSTUDIEN

Qualitative Studien nähern sich ihrem Gegenstand in der Regel mit einem anderen Verständnis von Wirkung. Im Kern zielen diese darauf ab, soziale Wirklichkeit, also auch „Wirkungen“, als Ergebnis kommunikativer und interaktiver Prozesse zu verstehen. In der empirischen Erhebung und Rekonstruktion von Deutungs-, Handlungs- und Bewertungsmustern sowie Diskursen werden die Sichtweisen der Akteure, ihr Handeln einschließlich der eingebetteten evaluativen Momente und den ihnen zugrunde liegenden Strukturen und Regeln zum Gegenstand der Analyse gemacht (vgl. von Kardorff 2017: 244; Flick 2006). Auch wenn Begriffe wie „Evidenz“ und „Wirkung“ nicht abgelehnt werden, herrscht im qualitativen Paradigma eine kritische Begriffsreflexion vor. Kelle (2006: 129–130) weist darauf hin, dass standardisierte Evaluationsdesigns nicht in der Lage seien, die für soziale Prozesse typischen intendierten Nebeneffekte und überraschenden Wirkungen überhaupt methodisch in den Blick zu bekommen und plädiert für eine qualitativ ausgerichtete und für verschiedene Kausalitäten offene Wirkungsevaluation.

In der Radikalisierungsprävention könnten beispielsweise Wirkungen eines Anti-Aggressionstrainings in einer standardisierten Erhebung kausal auf die *Inhalte* dieses Trainings zurückführbar sein – in einem flankierenden teilstandardisierten Interview mit Teilnehmenden des Angebots stellt sich dann jedoch heraus, dass die vor und nach dem eigentlichen Training ablaufende *Beziehungsarbeit* der gefährdeten Jugendlichen mit dem Sozialarbeiter bzw. der Sozialarbeiterin der eigentliche Wirkfaktor ist. Eine einfache *Black-Box*-Evaluation wäre nicht in der Lage, dies zu erkennen und könnte daher zu fehlerhaften Schlussfolgerungen führen.

Außerdem gibt es feldspezifische Gründe, die für den Einsatz qualitativer Methoden sprechen können. Die Bedingungen für quasi-experimentelle Wirkungsuntersuchungen sind in (sozial-)pädagogischen Settings, den Praxisfeldern der Sozialen Arbeit und der Kinder- und Jugendhilfe häufig nicht vorzufinden und auch nicht ohne Weiteres herzustellen. Jene Settings zeichnen sich vielmehr durch eine hohe Kontextabhängigkeit, individuelle Problemlagen und Lebensweltbezogenheit der Ansätze und Interventionen aus. In der komplexen sozialen Wirklichkeit erweisen sich einfache Kausalitätsannahmen oftmals als unterkomplex, denn „jede kausale Analyse sozialen Handelns muss berücksichtigen, dass Akteure sich in lokalen Kontexten bewegen, dort nach Regeln handeln und auf Wissensvorräte zurückgreifen, die Forschern *a prima vista* nicht bekannt sind, und dass sie oft ihre Handlungsziele auf Wegen verfolgen, die die Forscher nicht antizipieren können“ (Kelle 2006: 133).

Teilweise unabhängig von der Art der erhobenen Daten (statistisch oder qualitativ) ist das Design der Untersuchung. In diesem Zusammenhang empfiehlt Lösel (2008), die jeweiligen Vorteile qualita-

tiver und quantitativer Methoden der Evaluation zu kombinieren und verweist auf die Tatsache, dass viele Studien dies auf eine ganz selbstverständliche Art und Weise tun: „Ein experimentelles Design impliziert nicht notwendigerweise statistische Analyse (Shadish et al. 2002). [...] [S]elbst der Pionier von Evaluation empfahl die Kombination qualitativ-ethnographischer Methoden mit den rigorosen Kriterien von Experimenten (Campbell 1979)“ (Lösel 2008: 153; Übersetzung des Autorenteam). Das logische Modell und die *Realist Evaluation* sind zwei Evaluationsverfahren, die den kombinierten Einsatz qualitativer und quantitativer Methoden der Datenerhebung und Datenauswertung nahelegen, um Erkenntnisse über die Wirkungsmechanismen der untersuchten Programme zu erzielen.

4.3 DAS LOGISCHE MODELL

Evaluationen anhand logischer Modelle untersuchen bzw. klären die Programmmerkmale und Abläufe sowie deren strukturelle Rahmenbedingungen, durch die Präventionsprogramme ihre Wirkung erzielen wollen:

„Einigkeit besteht darin, dass der Fokus nicht mehr alleine auf Prozess- oder Strukturqualität gerichtet sein soll, sondern dass sich die Evaluation pädagogischer Arbeit auch auf die Ergebnisqualität, auf summative Aspekte beziehen soll und der Blick außerdem auf Wirkungen, Effekte und Zielerreichungen gerichtet sein muss. [...] Diese Forderung kann aber nicht darüber hinwegtäuschen, dass die Frage, wie und wozu, d. h. vor allem auch, auf welche Weise die Bewertung pädagogischer Arbeit angemessen erfolgen kann, nicht zufriedenstellend geklärt ist [...]. Auch die mancherorts fast schon inflationäre Verwendung des Begriffs der Evidenzbasierung vermag dies nicht zu verbergen.“ (Yngborn/Hoops 2018: 349)

Grundannahme logischer Modelle ist, „dass Programme bzw. Projekte auf der Grundlage bestimmter beschriebener Rahmenbedingungen und Ressourcen sowie auf der Grundlage der auf dieser Basis geplanten Aktivitäten bestimmte Zielsetzungen verfolgen“ (Klingelhöfer 2007: 38). Entsprechend werden in der Modellierung die Aktivitäten einer Maßnahme, vorliegende Rahmenbedingungen, Ressourcen und Zielstellungen identifiziert und zueinander in Beziehung gesetzt.

„Für eine Wirkungsevaluation ist diese Programmtheorie insofern interessant, als die Wirkungen – also die im Programm erreichten Ziele und ihre Verbindung zu den im Programm durchgeführten Aktivitäten – nicht mehr nur als Effekte beobachtbar sind, sondern in der Programmtheorie sichtbar werden und in einer Evaluation untersucht werden können.“ (Yngborn/Hoops 2018: 352).

Gerade weil das Angebot an Maßnahmen im Bereich der Radikalisierungsprävention so vielfältig ist und sich über verschiedene Fachbereiche und Kontexte erstreckt, muss der Fokus einer Wirksamkeitsevaluation auf den *Wirkungsmechanismen* liegen, die im Konzept einer Maßnahme wurzeln. Wodurch wurde die Veränderung (vermutlich) herbeigeführt, welche Mechanismen setzen bestimm-

te Aktivitäten in Gang? Dies setzt voraus, dass die theoretischen Annahmen über solche Mechanismen, die einer Maßnahme zugrunde liegen, explizit und den Evaluierenden bekannt sind und dass möglichst viele kontextuale Faktoren erhoben und mit einbezogen werden. Die Praktikerinnen und Praktiker werden dabei auch zu ihren eigenen Annahmen über Maßnahme, Konzeption, Abläufe, Mechanismen und Wirkungen befragt. In diesem Sinne ist das logische Modell ein Mittel, mit dem die Perspektiven der Praxis einbezogen werden können.

Logische Modelle können die (angenommenen) Wirkungsmechanismen einer Maßnahme sichtbar machen und verhelfen dadurch den beteiligten Fachkräften zu einem „mikroskopischen Blick“ und zu einem „Überblick über Veränderungsprozesse“ in der Zielgruppe (Yngborn/Hoops 2018: 363). Dieser Ansatz erscheint für die Evaluation im Bereich der Radikalisierungsprävention daher vielversprechend.²¹

4.4 REALIST EVALUATION

Ein weiterer Ansatz, die Wirkungsmechanismen von sozialen Programmen zu untersuchen, ist die *Realist Evaluation* (gleichbedeutend mit *Realistic Evaluation*) (Pawson/Tilley 1997; 2004). Ihre Gründer, die beiden Soziologen Ray Pawson und Nick Tilley, sehen die Sozialwissenschaften in der Pflicht, sich an der Lösung sozialer Probleme zu beteiligen, indem sie Wissen über die Wirkung sozialer Programme generieren. Im Einklang mit vielen anderen Kritikern und Kritikerinnen störten sie sich allerdings daran, dass zur Evidenzbasierung sozialer Programme und Präventionsmaßnahmen fast ausschließlich Ergebnisse aus Studien mit randomisierten Kontrollgruppen berücksichtigt wurden. Am Beispiel einer in den USA vielfach (durch randomisierte Kontrollgruppen) evaluierten polizeilichen Maßnahme im Bereich der häuslichen Gewalt beschreibt Tilley, wie die gleiche Maßnahme in einigen Städten das Risiko für einen erneuten Fall häuslicher Gewalt verhafteter Personen verringerte und in anderen erhöhte (Tilley 2010; Sherman et al. 1992). Aus der Beobachtung, dass die mehrmalige Evaluation der gleichen Maßnahme an unterschiedlichen Orten zu scheinbar widersprüchlichen Ergebnissen führen kann, leitet sich die Idee der *Realist Evaluation* ab. Anstatt einfach nur danach zu fragen, „ob etwas funktioniert“, steht im Mittelpunkt ihrer Untersuchung die Frage: „Was funktioniert für welche Zielgruppen und unter welchen Bedingungen?“.

Die *Realist Evaluation* untersucht die Effekte sozialer Programme und eignet sich durch diese Ausrichtung grundsätzlich auch zur Evaluation von Radikalisierungsprävention, ohne dabei den engen Voraussetzungen von randomisierten Kontrollgruppenstudien und den Einschränkungen ihrer Ergebnisse unterworfen zu sein.

Das Ziel jeder *Realist Evaluation* ist es, (angenommene) Kausalmechanismen in einer sozialen Intervention, wie beispielsweise einem Aussteigerprogramm, zu identifizieren und empirisch zu testen, unter welchen Bedingungen diese Mechanismen (nicht) greifen. Daraus sollen generalisierbare Aussagen über die richtige Konstellation von Programminhalten, Merkmalen der Zielgruppe und

21 Zum logischen Modell s. auch WKKF (1998), Wyatt Knowlton/Phillips (2009) und Beywl et al. (2007).

dem sozialstrukturellen Kontext abgeleitet werden. Dabei ist nicht eine bestimmte Maßnahme (z. B. Fallberatung durch einen bestimmten Träger) der Analysegegenstand der Untersuchung, sondern es sind die konkreten Methoden, Ressourcen und andere charakteristische Merkmale der Maßnahme, die den Analysegegenstand bilden. Entsprechend sagt das Ergebnis der Evaluation nicht pauschal etwas darüber aus, ob die Fallberatung des Trägers X erfolgreich ist oder nicht, sondern welche spezifischen Elemente des Programms in Kombination mit anderen Einflussfaktoren (k)einen Effekt für unterschiedliche Teilgruppen von Klienten und Klientinnen haben. „Mechanismen sind der Motor“ (Pawson/Tilley 2004: 18; Übersetzung des Autorenteam) für die Beeinflussung sozialer Prozesse. *Realist Evaluation* hat den Anspruch zu untersuchen, welche Bedingungen zusammentreffen müssen, damit ein Programm seine gewünschte Wirkung entfalten kann.

Überträgt man dieses Verständnis von Programmwirkung auf den Bereich der Prävention von Radikalisierung, bedeutet dies, systematisch zu erforschen, welche Mechanismen intendierte Veränderungen herbeiführen. Obwohl hier eine nahezu unüberschaubare Vielfalt von Einzelmaßnahmen auf eine noch größere Vielfalt von Klienten und Fällen trifft, liegt die Vermutung nahe, dass es wiederkehrende Programmmechanismen gibt. Diese Muster zu erkennen, wäre Element einer evidenzbasierten Prävention von Radikalisierung.

An dieser Stelle kann keine umfassende Anleitung zur Durchführung einer *Realist Evaluation* gegeben werden.²² Es soll jedoch nicht verschwiegen werden, dass diese Form der Evaluation hohe fachliche Anforderungen an die Evaluierenden stellt. Im Mittelpunkt der Untersuchung steht die empirische, d. h. auf qualitative und quantitative Daten gestützte Überprüfung von Hypothesen.

Abbildung 1 veranschaulicht die vier typischen Abschnitte: Da jede Maßnahme auf eine Veränderung abzielt, steckt dahinter immer auch eine mehr oder weniger klar artikulierte Programmtheorie und eine Annahme, warum und mit welchen Mitteln die Maßnahme die gewünschte Änderung herbeiführen soll (*Theory of Change*). Die Aufgabe der Evaluierenden ist es, idealerweise zusammen mit den Architekten der Maßnahme und den ausführenden Fachkräften, die Wirkungsannahmen in überprüfbare Hypothesen zu überführen.

22 Für eine kurze Übersicht s. Hewitt et al. (2012) und Westhorp (2014); für den Originalbeitrag s. Pawson/Tilley (1997).

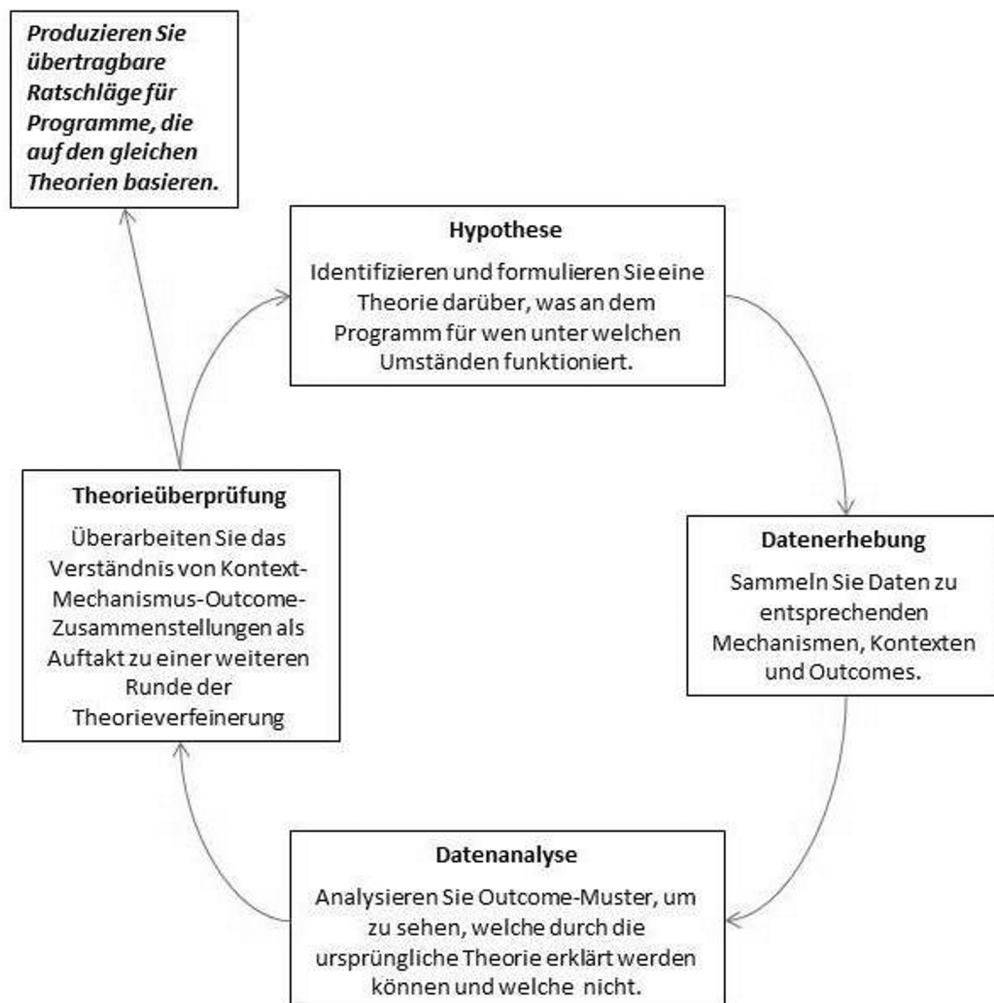


Abb. 1: Die 4 Phasen der Realist Evaluation (Übersetzung des Autorenteam aus Pawson/Tilley (2004) und Hewitt et al. (2012)).

Die Formulierung und Überprüfung von Hypothesen ist der Dreh- und Angelpunkt der *Realist Evaluation*. Dazu müssen entsprechende qualitative und/oder quantitative Daten erhoben und ausgewertet werden (Hewitt et al. 2012: 253). Die Hypothesen müssen demnach so formuliert und die Daten so ausgewählt werden, dass die Hypothesen mit den Daten widerlegt werden können. Die *Realist Evaluation* löst ihren Anspruch, Ursache-Wirkungs-Beziehungen abzubilden, ein, indem sie in der Datenanalyse Vergleichsgruppen bildet: „[...] die grundlegende Annahme ist, dass es differenzierte Auswirkungen auf den Erfolg oder Misserfolg innerhalb und zwischen Interventionen gibt. Die grundlegende

Vorgehensweise ist daher die Hypothesen mittels des Vergleichs von Untergruppen abzufragen“ (Pawson/Tilley 2004: 11; Übersetzung des Autorenteam).

Die Ergebnisse der Evaluation sind nicht der Weisheit letzter Schluss, sondern sie sind provisorisch. Sie bedeuten einen momentanen Erkenntnisgewinn, der mit Durchlaufen weiterer Evaluationszyklen von Zeit zu Zeit einer erneuten Prüfung unterzogen werden sollte. Dadurch verdichten sich Erkenntnisse in Bezug auf Fragen, die typischerweise an eine (Wirkungs-)Evaluation gerichtet sind (s. Tabelle 1):

Administrative/Politische Fragen	Mögliche Antwort aus einer <i>Realist Evaluation</i>
Hat die Maßnahme gewirkt?	Das hängt vom Kontext ab.
Sollen wir Maßnahme X anstatt Maßnahme Y fördern?	Erst prüfen, ob sie vergleichbar kommensurabel [evaluativ vergleichbar] sind.
Das Pilotprojekt war ein Erfolg, sollen wir es in die Fläche verbreiten?	Nein, überträgt nur seine Stärken.

Tabelle 1: Auswahl möglicher Fragen/Antworten einer *Realist Evaluation* (Pawson/Tilley 2004: 21; Übersetzung des Autorenteam).

Evaluationen nach dem Vorbild der *Realist Evaluation* eröffnen neue Wege in der manchmal festgefahrenen Debatte rund um Evidenzbasierung und Wirkung von Prävention von Radikalisierung (Gielen 2017).²³ Auch in anderen Politik- und Praxisfeldern hat sie den Diskurs bereichert und neue Handlungsmöglichkeiten eröffnet.

5. „STUDIEN BELEGEN, DASS ...“ – VERLÄSSLICHKEIT UND AUSSAGEKRAFT VON EVALUATIONSTUDIEN

Damit eine Evaluation verlässliche Ergebnisse liefern kann, müssen bei der Datenerhebung und -auswertung eine ganze Reihe bestimmter Voraussetzungen erfüllt sein. Üblicherweise können bei einer Evaluation nicht alle diese Voraussetzungen gleichermaßen erfüllt werden, sodass es notgedrungen zu Einschränkungen kommt.

In diesem Kapitel geht es um Anhaltspunkte für die Qualität von Evaluation. Eine solche fachliche Einschätzung gibt den Adressatinnen und Adressaten von Evaluationsergebnissen (bspw. aus Praxis, Verwaltung, Politik, Wissenschaft) eine Orientierung dafür, wie verlässlich und belastbar die Ergebnisse der Studie sind. Die Beurteilung einer Evaluation ist im wörtlichen Sinne also die Evaluation einer Evaluation.²⁴

²³ S. auch „Wirkungsorientierung und Evaluation“ 21. Jahrestagung der DeGEval: <https://www.degeval.org/veranstaltungen/jahrestagungen/dresden-2018>; 14.6.2018.

²⁴ Widmer (1996) bezeichnet sie daher als Meta-Evaluation.

Diese Einschätzung ist aus verschiedenen Gründen enorm wichtig. Erstens: Evaluation ist keine geschützte Bezeichnung. Die Einhaltung von Standards soll sicherstellen, dass in einer Evaluation auch das steckt, was erwartet wird: valide, relevante und sachgerechte Bewertung einer sozialen Intervention. Zweitens: Keine Evaluation ist perfekt, aber hochwertige Studien schließen eine Reihe von möglichen „technischen Fehlern“ (Parkhurst 2017: 7) von vornherein aus. Technische Fehler können unbemerkt zu *systematischen Verzerrungen* der Ergebnisse und damit letzten Endes zu fehlerhaften Empfehlungen und Entscheidungen führen. Drittens: Gütekriterien geben in qualitativen wie quantitativen Studien einen Anhaltspunkt für die Qualität der Ergebnisse. Robuste Studienergebnisse lassen sich schwieriger entkräften, wohingegen Studien mit vermeidbaren Schwachstellen nur allzu leicht von Kritikern und Kritikerinnen „zerrissen“ werden können. Gütekriterien helfen den Adressatinnen und Adressaten von Evaluationen, verlässliche von unverlässlichen Befunden zu trennen.

Angesichts des enormen Einflusses, den Forschungs- und Evaluationsdaten in politischen Prozessen haben können, erkennt man schnell, wie wichtig es ist, deren Robustheit verlässlich einzuschätzen. „Eine Bewertung der forschungsmethodischen Qualität von Evaluationsstudien ist somit unerlässlich, will man zu einer rationalen und reflektierten Entscheidung über die Einsatzmöglichkeiten von Fördermaßnahmen und Präventionsprogrammen kommen“ (Beelmann/Hercher 2016: 98).

Dabei gilt für die (wirkungsorientierte) Evaluation einer Präventionsmaßnahme der gleiche Grundsatz wie für die Maßnahme selbst: Ohne eine klare und präzise Benennung ihrer Ziele und darauf bezogene Methoden gibt es keinen Anhaltspunkt für deren Bewertung. Ein Deradikalisierungsprojekt, das nicht bestimmt, welche konkreten Aspekte von (De-)Radikalisierung es auf welche Weise beeinflussen möchte, kann nicht sinnvoll im Hinblick auf seine Wirkungen evaluiert werden. Genauso kann man die Qualität von Evaluationsstudien immer nur im Hinblick auf das jeweilige Erkenntnisinteresse in Kombination mit der Forschungsmethode bewerten.

Evaluation muss immer gegenstandsbezogen sein, d. h. Erkenntnisinteresse und Evaluationsgegenstand leiten die Wahl der Methode und nicht umgekehrt. In der Evaluationsforschung können daher nahezu alle Methoden der empirischen Sozialforschung zum Einsatz kommen. Dieser durchaus wünschenswerte Methodenpluralismus verkompliziert allerdings manchmal die Einschätzung der wissenschaftlichen Belastbarkeit von Evaluationsstudien. Wissenschaftler und Wissenschaftlerinnen aller Disziplinen streiten darüber, wie angemessen solche Bewertungssysteme zur Beurteilung von Forschungsergebnissen sind und zu welchen Zwecken man sie überhaupt einsetzen sollte. In philosophischen und erkenntnistheoretischen, methodologischen und methodischen Debatten haben sich einige forschungsmethodische und fachliche Gütestandards durchgesetzt.

Die Gesellschaft für Evaluation e.V. (DeGEval) hat beispielsweise in einem aufwändigen Verständigungsprozess ihrer Mitglieder einen Katalog fachlicher Standards erarbeitet.²⁵ Demnach sind die vier grundlegenden Merkmale einer guten Evaluation Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit. Die Standards dienen einerseits als Leitbild für die Planung und Durchführung von Evaluationsstudien und andererseits als Bewertungsmaßstab für die Ergebnisse bereits abgeschlossener

25 <https://www.degeval.org/degeval-standards-alt/archiv/revisionsprozess>; 5.2.2018.

Evaluationen. Sie beanspruchen, „politikfeldübergreifend“ (DeGEval 2017: 26) und „unabhängig vom konkret angewendeten Evaluationsansatz“ (DeGEval 2017: 28) gültig und anwendbar zu sein.

Neben diesen allgemeinen fachlichen Standards für Evaluation existieren forschungsmethodische Standards. Wie genau und verlässlich die Ergebnisse und Schlussfolgerungen von Evaluationsstudien sind, hängt letzten Endes davon ab, in welchem Umfang die Forscher und Forscherinnen Regeln der Datenerhebung und -auswertung einhalten. Für diese Bewertung existieren wiederum eigenständige Bewertungssysteme, wie beispielsweise die *Maryland Scientific Methods Scale* (SMS) (Sherman et al. 1998; Farrington et al. 2002). Beelmann und Hercher (2016) ergänzen die SMS um weitere Merkmale, die im Kontext einer Evaluation relevant sind, wie z. B. die Verallgemeinerbarkeit der Befunde und die Operationalisierung von Variablen zur Erfolgsmessung.

Im Bereich der quantitativen Forschung existiert ein relativ hohes Maß an fachlichem Konsens darüber, welche Faktoren die Qualität der Daten und Ergebnisse beeinflussen. Kriterien wie interne, externe und Konstruktvalidität, Reliabilität der Messung, Irrtumswahrscheinlichkeit, Stichprobenfehler, Scheinkorrelationen, ökologischer Fehlschluss und methodische Artefakte können von entsprechend geschultem Personal ohne großen interpretativen Spielraum erkannt und entsprechend bewertet werden.

Bei qualitativen Studien gibt es naturgemäß einen größeren Spielraum bei der Bewertung durch Gütekriterien, wie beispielsweise die intersubjektive Nachvollziehbarkeit. Einige qualitative Forscher und Forscherinnen haben Vorschläge für „Standards nicht standardisierter Forschung“ (Flick 2014: 416; Steinke 1999) gemacht, während andere dies als grundsätzlich problematisch betrachten (Hannes/Pearson 2011: 27). Ein verbreitetes Argument lautet, dass genormte Ansprüche an die Qualität qualitativer Forschung deren Diversität und Denkweisen beeinträchtigten und so letzten Endes das Gegenteil davon bewirken, was sie eigentlich bezwecken wollen (Sandelowski/Barroso 2006: 134–137).

Weitgehende Übereinstimmung zwischen den verschiedenen qualitativen Forschungssträngen herrscht dahingehend, dass auch für die qualitative Forschung Qualitätsmaßstäbe wichtig sind (Tracy 2010). Das „Handbuch Forschungsmethoden und Evaluation“ berichtet mit Verweis auf die *Cochrane Qualitative & Implementation Method Group* von „mehr als einhundert verschiedenen Kriterienkatalogen“ und stellt fest: „Ein konsensfähiger einheitlicher Kanon von Kernkriterien ist indessen bislang nicht entstanden“ (Döring/Bortz 2016: 107).

Da der Auswertung qualitativer Daten in aller Regel interpretative Paradigmen zugrunde liegen und verschiedene Methoden angewendet werden, spiegelt die Diversität von Gütekriterien zu einem guten Teil die Forderung nach kontextangemessener Forschung wider. Gleichzeitig steht qualitative Evaluationsforschung damit vor erheblichen Herausforderungen, was die jeweilige Begründung der Ergebnisse angeht.

6. EMPFEHLUNGEN

Aufträge zur Evaluation klar artikulieren

Eine unklare Auftragslage kann im Ergebnis zu schlechten Evaluationen führen. Der Auftraggeber bzw. die Auftraggeberin einer Evaluation sollte daher in der Ausschreibung klar artikulieren, welche Art der Evaluation er bzw. sie sich wünscht. Einfach nur eine „Evaluation“ oder eine „wissenschaftliche Begleitung“ in Auftrag zu geben, käme in einem Restaurant der Bestellung „etwas zu essen“ gleich.

Kein Automatismus

Eine Pflicht zur Evaluation ist wünschenswert – allerdings nur unter bestimmten Voraussetzungen. Im Bereich der Radikalisierungsprävention brauchen wir nicht einfach nur „mehr Evaluation“, sondern mehr gute Evaluation. Wirkungsorientierte Evaluationen sind nur dann sinnvoll, wenn, erstens, das Konzept der Maßnahme eine präzise Bestimmung des Präventionsgegenstands, der Präventionsebenen und der Präventionsziele sowie der Projektlogik (Wirkmechanismen) aufweist; zweitens, das Erkenntnisinteresse einer Evaluationsstudie in Abstimmung mit den Beteiligten formuliert ist und, drittens, ausreichend finanzielle Mittel für eine gegenstandsangemessene Evaluation vorgesehen sind (Richtwert: zehn Prozent der beantragten Fördersumme).

Geduld

Schnelle Ergebnisse sind weder in der Prävention noch bei deren Evaluation zu erwarten. Der Erkenntnisgewinn einer Evaluationsstudie steigt in der Regel mit ihrer Laufzeit. Vor allem Längsschnittstudien erlauben Rückschlüsse über die Nachhaltigkeit von Präventionseffekten. Bei Präventionsangeboten, Distanzierungs- und Deradikalisierungsmaßnahmen handelt es sich in aller Regel um mittel- und langfristig angelegte sozialpädagogische Prozesse und mitunter um mehrjährige Fallarbeit. Förderer sollten daher sowohl die Präventionsarbeit als auch deren Evaluation langfristig ausrichten und mit entsprechenden Mitteln ausstatten. Eine übereilte Evaluation (z. B. gegen Ende einer Projektlaufzeit) nützt niemandem etwas und frustriert zudem die beteiligten Akteure und Akteurinnen.

Berücksichtigung aller relevanten Daten

Die Qualität einer Evaluationsstudie hängt maßgeblich von den ausgewerteten Daten ab. Relevante Projektdaten sollten unter Berücksichtigung aller – z. B. datenschutzrechtlichen – Erfordernisse für die Evaluation nutzbar sein. Oft liegen die wesentlichen Daten zur Evaluierung erst nach Projektabschluss vollständig vor, wenn die Finanzierung ausgelaufen ist. Wünschenswert ist, die programmbegleitende Evaluation über das Ende der Maßnahme hinaus zu fördern, um alle relevanten Projektdaten berücksichtigen und längerfristige Effekte beobachten zu können.

Methodenpluralismus

Welche empirischen Methoden der Sozialforschung für eine Evaluation geeignet sind, hängt vom konkreten Erkenntnisinteresse und dem Gegenstand der Evaluation ab. Für die Evaluation von Radikalisierungsprävention gibt es kein Patenrezept. Die in diesem Report vorgestellten Ansätze, wie *Realist Evaluation*, *logische Modelle* und anderen Formen der Evaluation, erlauben die Anwendung unterschiedlicher Forschungsmethoden, um der Frage nachzugehen, wie Präventionseffekte in Abhängigkeit von dem Kontext, in den sie eingebettet sind, funktionieren und wirken.

Hol- und Bringschuld

Der Transfer von Wissen zwischen Forschung, Politik und Praxis muss in geeigneter Weise gestaltet werden. Dabei gibt es eine Hol- und eine Bringschuld zwischen Wissenschaft, Politik, Verwaltung und Praxis. Auf Seiten der Wissenschaft gibt es eine Bringschuld in der Form, dass die Ergebnisse von Evaluationsstudien adressatengerecht kommuniziert werden müssen und dass dabei deren Mehrwert konkret benannt wird. Andererseits besteht eine Holschuld seitens der Programmverantwortlichen und Förderer, die den Wissenstransfer durch ihre aktive Nachfrage unterstützen können.

Reflexivität

Politik, Gesellschaft und Evaluationsforschung sollten sich darüber im Klaren sein, dass Evaluationsvorhaben nicht einfach „neutral“ und „extern“ auf die Praxis blicken. In jedem Fall formen sie ein Stück weit auch ihren Evaluationsgegenstand mit, beispielsweise durch die Wahl des Untersuchungsdesigns, das wiederum bestimmte Anforderungen an die evaluierten Projekte und Maßnahmen stellt. Dies gilt für formative Evaluationen ganz dezidiert und gewollt, trifft jedoch für alle Evaluationsdesigns in unterschiedlichen Maßen zu. Daraus resultiert für alle Akteure die Notwendigkeit, sich mit dieser Komplexität reflexiv auseinanderzusetzen.

- Abay Gaspar, Hande/Daase, Christopher/Deitelhoff, Nicole/Junk, Julian/Sold, Manjana* 2018: Was ist Radikalisierung? Präzisierungen eines umstrittenen Begriffs, PRIF Report 5/2018, Frankfurt a.M.
- Baaken, Till/Becker, Reiner/Bjorgo, Tore/Kiefer, Michael/Korn, Judy/Mücke, Thomas/Ruf, Maximilian/Walkenhorst, Dennis* 2018: Herausforderung Deradikalisierung: Einsichten aus Wissenschaft und Praxis, PRIF Report 9/2018, Frankfurt a.M.
- Beelmann, Andreas* 2011: The Scientific Foundation of Prevention. The Status Quo and Future Challenges of Developmental Crime Prevention, in: Bliesener, Thomas/Beelmann, Andreas/Stemmler, Mark (Hrsg.): Antisocial Behavior and Crime: Contributions of Developmental and Evaluation Research to Prevention and Intervention, Cambridge, MA, 137–164.
- Beelmann, Andreas* 2017: Grundlagen einer entwicklungsorientierten Prävention des Rechtsextremismus (Gutachten im Rahmen des Wissenschafts-Praxis-Dialogs zwischen dem Landespräventionsrat Niedersachsen und der Friedrich-Schiller-Universität Jena), <http://lpr.niedersachsen.de/html/download.cms?id=2596&datei=Gutachten-LPR+Niedersachsen-Version3.pdf>; 15.7.2018.
- Beelmann, Andreas/Hercher, Judith* 2016: Methodische Beurteilung von Evaluationsstudien im Bereich der Gewalt- & Kriminalitätsprävention: Beschreibung und Begründung eines Methodenprofils, in: Stiftung Deutsches Forum für Kriminalprävention (Hrsg.): Entwicklungsförderung & Gewaltprävention – Aktuelle Beiträge aus Wissenschaft und Praxis, Bonn, 97–116.
- Bellmann, Johannes/Müller, Thomas* 2011: Evidenzbasierte Pädagogik – ein Déjà-vu? Einleitende Bemerkungen zur Kritik eines Paradigmas, in: Bellmann, Johannes/Müller, Thomas (Hrsg.): Wissen, was wirkt: Kritik evidenzbasierter Pädagogik, Wiesbaden, 9–32.
- Beywl, Wolfgang/Kehr, Jochen/Mäder, Susanne/Niestroj, Melanie* 2007: Planung von Evaluationen, Evaluation Schritt für Schritt, Münster, 25–26.
- Biene, Janusz/Daase, Christopher/Junk, Julian/Müller, Harald* (Hrsg.) 2016: Salafismus und Dschihadismus in Deutschland: Ursachen, Dynamiken, Handlungsempfehlungen, Frankfurt a.M.
- Biene, Janusz/Daphi, Priska/Fielitz, Maik/Müller, Harald/Weipert-Fenner, Irene* 2015: Nicht nur eine Frage der Sicherheit. Salafismus in Deutschland als gesamtgesellschaftliche Herausforderung (HSFK Standpunkte 2015/1), https://www.hsfk.de/fileadmin/HSFK/hsfk_downloads/standpunkt0115.pdf; 15.7.2018.
- Birtsch, Vera/Kurz-Adam, Maria/Lippmann, Claus/Merten, Roland/Speck, Karsten* (Hrsg.) 2016: unsere jugend: Die Zeitschrift für Studium und Praxis der Sozialpädagogik, 68: 11–12, München.
- Bischoff, Ursula/Gehne, Carsten/Greuel, Frank/Johansson, Susanne/König, Frank/Schlimbach, Tabea/Zierold, Diana/Zimmermann, Eva* 2011: Abschlussbericht der Programmevaluation der Bundesprogramme „VIELFALT TUT GUT. Jugend für Vielfalt, Toleranz und Demokratie“ und „kompetent. für Demokratie – Beratungsnetzwerke gegen Rechtsextremismus“ Berichtszeitraum 01.07.2007 bis 31.12.2010, https://www.dji.de/fileadmin/user_upload/bibs/686_14317_PEabschlussbericht2010.pdf; 6.6.2018.
- Bischoff, Ursula/Greuel, Frank/Johansson, Susanne/König, Frank/Reiter, Stefanie/Zimmermann, Eva* 2015: Nachhaltige Impulse für eine wirksame Interventions- und Präventionsarbeit. Ergebnisse der Programmevaluation des Bundesprogramms „TOLERANZ FÖRDERN – KOMPETENZ STÄRKEN“, Abschlussbericht 2015, unter Mitarbeit von Claudia Menge, <https://www.dji.de/filead->

- min/user_upload/toko/2015_Nachhaltige_Impulse_f%c3%bcr_eine_wirksame_Interventions_Pr%c3%a4ventionsarbeit.pdf; 6.6.2018.
- BMFSFJ (Bundesministerium für Familie, Senioren, Frauen und Jugend) 2014: Abschlussbericht des Bundesprogramms „Initiative Demokratie stärken“, <https://www.demokratie-leben.de/fileadmin/content/PDF-DOC-XLS/Abschlussberichte/Abschlussbericht-IDS.pdf>; 6.6.2018.
- Bohn, Irina/Münchmeier, Richard 1997: Dokumentation des Modellprojekts. Das Aktionsprogramm gegen Aggression und Gewalt AgAG, Bd. 1, Münster.
- Böttcher, Astrid 2017: Radikalismus und Extremismus: Konzeptualisierung und Differenzierung zweier umstrittener Begriffe in der deutschen Diskussion (Dissertation an der Universität Leiden).
- bpb – (Bundeszentrale für Politische Bildung) o.J.: Bundesweite Übersicht der Anlaufstellen, <http://www.bpb.de/politik/extremismus/radikalisierungspraevention/208847/uebersicht-anlaufstellen>; 6.6.2018.
- Bröckling, Ulrich 2008: Vorbeugen ist besser... Zur Soziologie der Prävention, in: Behemoth. A Journal on Civilisation 1, 38–48.
- Caplan, Gerald 1964: Principles of Preventive Psychiatry, 5. Aufl., New York, NY.
- Ceylan, Rauf/Kiefer, Michael 2018: Radikalisierungsprävention in der Praxis: Antworten der Zivilgesellschaft auf den gewaltbereiten Neosalafismus, Wiesbaden.
- Chalmers, Iain 2005: If Evidence-Informed Policy Works in Practice, Does It Matter if It Doesn't Work in Theory?, in: Evidence & Policy: A Journal of Research, Debate and Practice 1: 2, 227–242.
- Christmann, Kris 2012: Preventing Religious Radicalisation and Violent Extremism: A Systematic Review of the Research Evidence, http://www.safecampuscommunities.ac.uk/uploads/files/2016/08/yjb_preventing_violent_extremism_systematic_review_requires_uploading.pdf; 7.6.2018.
- DeGEval (Gesellschaft für Evaluation e.V.) 2017: Standards für Evaluation, Erste Revision 2016, Mainz.
- de la Chaux, Marlen/Kober, Marcus/Nabo, Mitra Moussa 2018: Überlegungen zur Schaffung einer Evidenzgrundlage für die Präventionsarbeit im Bereich islamistischer Extremismus, in: Walsh, Maria, et al. (Hrsg.): Evidenzorientierte Kriminalprävention in Deutschland : Ein Leitfaden für Politik und Praxis, Wiesbaden, 489–509.
- Döring, Nicola/Bortz, Jürgen 2016: Evaluationsforschung, in: Döring, Nicola/Bortz, Jürgen (Hrsg.): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, Heidelberg, 975–1036.
- Eisner, Manuel/Malti, Tina/Ribeaud, Denis/Müller, Barbara 2012: Groß angelegte Feldversuche in der kriminologischen Präventionsforschung. Das Zürcher Projekt zur sozialen Entwicklung von Kindern, in: Strobl, Rainer/Lobermeier, Olaf/Heitmeyer, Wilhelm (Hrsg.): Evaluation von Programmen und Projekten für eine demokratische Kultur, Wiesbaden, 69–97.
- El-Mafaalani, Aladin/Fathi, Alma/Mansour, Ahmad/Müller, Jochen/Nordbruch, Götz/Waleciak, Julian 2016: Ansätze und Erfahrungen der Präventions- und Deradikalisierungsarbeit (HSFK-Reportreihe „Salafismus in Deutschland“, Report Nr. 6/2016), http://www.hsfg.de/fileadmin/HSFK/hsfk_publicationen/report_062016.pdf; 15.7.2018.

- Farrington, David P./Gottfredson, Denise C./Sherman, Lawrence W./Welsh, Brandon C.* 2002: The Maryland Scientific Methods Scale, in: Sherman, Lawrence W./Farrington, David P./ Welsh, Brandon C./MacKenzie, Doris Layton (Hrsg.): Evidence-Based Crime Prevention, 1. Aufl., London, 13–21.
- Flick, Uwe* (Hrsg.) 2006: Qualitative Evaluationsforschung: Konzepte, Methoden, Umsetzung, Reinbek.
- Flick, Uwe* 2014: Gütekriterien qualitativer Sozialforschung, in: Baur, Nina/Blasius, Jörg (Hrsg.): Handbuch Methoden der empirischen Sozialforschung, Wiesbaden, 411–423.
- Frazer, Owen/Nünlist, Christian* 2015: Countering Violent Extremism in der Extremismusbekämpfung, in: CSS Analysen zur Sicherheitspolitik 183, Zürich.
- Frindte, Wolfgang/Ben Slama, Brahim/Dietrich, Nico/Pisoiu, Daniela/Uhlmann, Milena/Kausch, Melanie* 2016: Wege in die Gewalt. Motivationen und Karrieren salafistischer Jihadisten (HSFK-Reportreihe „Salafismus in Deutschland“, Report Nr. 3/2016), https://www.hsfk.de/fileadmin/HSFK/hsfk_publicationen/report_032016.pdf; 15.7.2018.
- Frindte, Wolfgang/Preiser, Siegfried* 2007: Präventionsansätze gegen Rechtsextremismus, in: Aus Politik und Zeitgeschichte 11, 32–38.
- Gielen, Amy-Jane* 2017: Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How?, in: Terrorism and Political Violence, 1–19 (Online First), DOI: 10.1080/09546553.2017.1313736.
- Glock, Birgit/Lüter, Albrecht* 2017: Teach2Reach: Aufklärungs-Workshops zu religiös begründetem Fundamentalismus. Ein Projekt des Violence Prevention Network e.V. – Abschlussbericht, <https://www.berlin.de/lb/lkbgg/landesprogramm/teach2reach-abschlussbericht-22-november-2017.pdf>; 7.6.2018.
- Gordon, Robert S.* 1983: An Operational Classification of Disease Prevention, in: Public Health Reports 98: 3, 107–109.
- Greuel, Frank* 2018: Das (zu) weite Feld der Prävention oder: Wo Prävention beginnen und enden sollte, PRIF Blog, <https://blog.prif.org/2018/04/16/das-zu-weite-feld-der-praevention/>; 3.5.2018.
- Gruber, Florian/Lützing, Saskia* 2017: Extremismusprävention in Deutschland – Erhebung und Darstellung der Präventionslandschaft. Modulabschlussbericht, unter Mitarbeit von Iris Klima, Georg Sielaff und Stefan Wick (Eine Auswertung im Rahmen des BKA-Forschungsprojekts „Entwicklungsmöglichkeiten einer phänomenübergreifend ausgerichteten Prävention politisch motivierter Gewaltkriminalität (PüG)“), Wiesbaden, https://www.bka.de/SharedDocs/Downloads/DE/Publicationen/Publicationsreihen/Forschungsergebnisse/2017PueG_Literaturanalyse_Entwicklungsmoeglichkeiten.html?nn=27638; 15.7.2018.
- Hannes, Karin/Pearson, Alan* 2011: Obstacles to the Implementation of Evidence-Based Practice in Belgium: A Worked Example of Meta-Aggregation, in: Hannes, Karin/Lockwood, Craig (Hrsg.): Synthesizing Qualitative Research, Chichester, 21–39.
- Helmus, Todd/Matthews, Miriam/Ramchand, Rajeev/Beaghley, Sina/Stebbins, David/Kadlec, Amanda/Brown, Michael A./Kofner, Aaron/Acosta, Joie* 2017: RAND Program Evaluation Toolkit for Countering Violent Extremism, Santa Monica, CA.

- Hewitt, Gillian/Sims, Sarah/Harris, Ruth 2012: The Realist Approach to Evaluation Research, in: International Journal of Therapy and Rehabilitation 19: 5, 250–259.
- Kelle, Udo 2006: Qualitative Evaluationsforschung und das Kausalitätsparadigma, in: Flick, Uwe (Hrsg.): Qualitative Evaluationsforschung: Konzepte, Methoden, Umsetzung, Reinbek, 117–135.
- Kiefer, Michael 2017: Radikalisierungsprävention in Deutschland – ein Problemaufriss, in: Toprak, Ahmet/Weitzel, Gerrit (Hrsg.): Salafismus in Deutschland: Jugendkulturelle Aspekte, pädagogische Perspektiven, Wiesbaden, 121–134.
- Klingelhöfer, Susanne 2007: Das Programm „Entimon“: Spezifika, Potenziale und Herausforderungen einer induktiv-rekonstruierenden Evaluation anhand Logischer Modelle, in: Glaser, Michaela/Schuster, Silke (Hrsg.): Evaluation präventiver Praxis gegen Rechtsextremismus – Positionen, Konzepte und Erfahrungen, Halle a.d.S., 32–52.
- Klingelhöfer, Susanne/Schmidt, Mareike/Schuster, Silke/Brüggemann, Ulrich 2007: Abschlussbericht der wissenschaftlichen Begleitung des Programms „Entimon – gemeinsam gegen Gewalt und Rechtsextremismus“, Jahre 2002–2006, https://www.dji.de/fileadmin/user_upload/bibs/188_8436_AB-Entimon.pdf; 15.7.2018.
- Kober, Marcus 2017: Zur Evaluation von Maßnahmen der Prävention von religiöser Radikalisierung in Deutschland, in: Journal for Deradicalization 11, 219–256.
- Köhler, Daniel 2015: De-Radicalization and Disengagement Programs as Counter-Terrorism and Prevention Tools. Insights from Field Experiences Regarding German Right-Wing Extremism and Jihadism, in: Lombardi, Marco/Ragab, Eman/Chin, Vivienne/Dandurand, Yvon/de Divitiis, Valerio/Burato, Alessandro (Hrsg.): Countering Radicalisation and Violent Extremism Among Youth to Prevent Terrorism, Amsterdam.
- Köhler, Daniel 2016: Strukturelle Qualitätsstandards in der Interventions- und Präventionsarbeit gegen gewaltbereiten Extremismus, Stuttgart.
- Lindekilde, Lasse 2012: Introduction: Assessing the Effectiveness of Counter-Radicalisation Policies in Northwestern Europe, in: Critical Studies on Terrorism 5: 3, 335–344.
- Litmanovitz, Yael/Weisburd, David/Hasisi, Badi/Wolfowicz, Michael 2017: What Are the Social, Economic, Psychological and Environmental Risk Factors that Lead to Radicalization and Recruitment to Terrorism?, 27.9.2017, <https://campbellcollaboration.org/library/social-economic-psychological-environmental-risk-factors-radicalization-terrorism.html>; 7.6.2018.
- Lösel, Friedrich 2008: Doing Evaluation Research in Criminology: Balancing Scientific and Practical Demands, in: King, Roy D./Wincup, Emma (Hrsg.): Doing Research on Crime and Justice, Oxford.
- Lüders, Christian 2011: Von der scheinbaren Selbstverständlichkeit präventiven Denkens, in: DJI Impulse: Das Bulletin des Deutschen Jugendinstituts 2: 94, 4–6.
- Lützing, Saskia/Gruber, Florian 2017: PÜG – Extremismusprävention in Deutschland – Herausforderungen und Optimierungspotential, Modulabschlussbericht (Eine Auswertung im Rahmen des BKA-Forschungsprojekts „Entwicklungsmöglichkeiten einer phänomenübergreifend ausgerichteten Prävention politisch motivierter Gewaltkriminalität (PÜG)“), Wiesbaden, https://www.bka.de/SharedDocs/Downloads/DE/Publikationen/Publikationsreihen/Forschungsergebnisse/2017PueG_ExtremismuspraeventionInDeutschland_Herausforderung.html; 15.7.2018.

- Lynen von Berg, Heinz/Roth, Roland* (Hrsg.) 2003: Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen, Wiesbaden.
- Madriaza, Pablo/Ponsot, Anne-Sophie* 2015: Preventing Radicalization: A Systematic Review, Montreal.
- Maurer, Thomas* 2017: Die Pluralität der Radikalisierung – Eine systematische Analyse der Theorieansätze zur Radikalisierungsforschung, in: *Journal for Deradicalization* 13, 49–100.
- Möller, Kurt* 2003: Aktuelle politische Programme und Aktivitäten der pädagogischen und sozialarbeiterischen Bekämpfung von Rechtsextremismus und Fremdenfeindlichkeit – eine kritische Zwischenbilanz, in: *Lynen von Berg, Heinz/Roth, Roland* (Hrsg.): Maßnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen, Wiesbaden, 27–50.
- Möller, Kurt/Neuscheler, Florian* 2018: Evaluationsbericht Beratungsstelle Hessen 2018: Bericht über zentrale Ergebnisse der Evaluation der Beratungsstelle Hessen – Religiöse Toleranz statt Extremismus (Managementfassung), Esslingen, <http://www.violence-prevention-network.de/de/publikationen/evaluationsberichte/>; 6.6.2018.
- Neumann, Peter* 2013: Radikalisierung, Deradikalisierung und Extremismus, <http://www.bpb.de/apuz/164918/radikalisierung-deradikalisierung-und-extremismus?p=0>; 15.7.2018.
- Neumann, Peter* 2016: Der Terror ist unter uns: Dschihadismus, Radikalisierung und Terrorismus in Europa, Berlin.
- Neumann, Peter* 2017a: Countering Violent Extremism and Radicalisation that Lead to Terrorism: Ideas, Recommendations, and Good Practices from the OSCE Region, <http://www.osce.org/chairmanship/346841?download=true>; 15.7.2018.
- Neumann, Peter* 2017b: Was wir über Radikalisierung wissen – und was nicht, in: Kärger, Jana (Hrsg.): „Sie haben keinen Plan B“. Radikalisierung, Ausreise, Rückkehr – zwischen Prävention und Intervention, Bonn.
- Nordbruch, Götz* 2017: Staatliche Kontrolle statt zivilgesellschaftlichem „Wildwuchs“? Präventions- und Ausstiegsarbeit in Deutschland, in: Bundesarbeitsgemeinschaft Religiös Begründeter Extremismus, <https://bag-relex.de/2017/10/23/staatliche-kontrolle-statt-zivilgesellschaftlichem-wildwuchs-praeventions-und-ausstiegsarbeit-in-deutschland/>; 6.6.2018.
- OSCE (Organization for Security and Co-operation in Europe)* 2014: Preventing Terrorism and Countering Violent Extremism and Radicalization that Lead to Terrorism: A Community-Policing Approach, Wien.
- Parkhurst, Justin* 2017: The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence, New York, NY.
- Pawson, Ray/Tilley, Nicholas* 1997: Realistic Evaluation, London.
- Pawson, Ray/Tilley, Nicholas* 2004: Realist Evaluation, British Cabinet Office, http://www.community-matters.com.au/RE_chapter.pdf; 15.7.2018.
- Perrez, Meinrad* 1998: Prävention und Gesundheitsförderung, in: Baumann, Urs/Perrez, Meinrad (Hrsg.): Lehrbuch Klinische Psychologie – Psychotherapie (Psychologie-Lehrbuch), 2., vollst. überarb. Aufl., Bern, 366–391.

- Pfahl-Traughber, Armin* 2015: Der Islamismus ist kein grüner Faschismus, sondern ein religiöser Extremismus. Eine kritische Prüfung einschlägiger Kriterien anlässlich einer öffentlichen Debatte, in: Schneiders, Thorsten Gerald (Hrsg.): Salafismus in Deutschland. Ursprünge und Gefahren einer islamisch-fundamentalistischen Bewegung, Bielefeld, 149–167.
- Quent, Matthias* 2016: Rassismus, Radikalisierung, Rechtsterrorismus. Wie der NSU entstand und was er über die Gesellschaft verrät, Weinheim.
- Robert Koch-Institut* 2012: Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? (Beiträge zur Gesundheitsberichterstattung des Bundes), http://edoc.rki.de/documents/rki_fv/req9cSaGLfIA/PDF/29PjAJd8GC8w.pdf; 15.7.2018.
- Sandelowski, Margarete/Barroso, Julie* 2006: Handbook for Synthesizing Qualitative Research, New York, NY.
- Schmid, Alex P.* 2013: Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review (ICCT Research Paper), <https://www.icct.nl/download/file/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf>; 15.7.2018.
- Sherman, Lawrence W./Gottfredson, Denise C./MacKenzie, Doris L./Eck, John/Reuter, Peter/Bushway, Shawn D.* 1998: Preventing Crime: What Works, What Doesn't, What's Promising (Research in Brief. National Institute of Justice), <https://files.eric.ed.gov/fulltext/ED423321.pdf>; 15.7.2018.
- Sherman, Lawrence W./Schmidt, Janell D./Rogan, Dennis P.* 1992: Policing Domestic Violence: Experiments and Dilemmas, New York, NY.
- Sischka, Kerstin* 2018: Was können Psychotherapeuten zur Radikalisierungsprävention und Deradikalisierung beitragen?, 25.1.2018, <http://www.ufuq.de/was-koennen-psychotherapeuten-zur-radikalisierungspraevention-und-deradikalisierung-beitragen/>; 15.7.2018.
- Srowig, Fabian/Roth, Viktoria/Pisoiu, Daniela/Seewald, Katharina/Zick, Andreas* 2018: Radikalisierung von Individuen. Ein Überblick über mögliche Erklärungsansätze, PRIF Report 6/2018, Frankfurt a.M.
- Steinke, Ines* 1999: Kriterien qualitativer Forschung: Ansätze zur Bewertung qualitativ-empirischer Sozialforschung, Weinheim.
- Steinke, Ronen* 2017: Aussteigerprogramme für Extremisten: Hauptsache, sie legen keine Bombe, in: Süddeutsche Zeitung, 16.10.2017, <http://www.sueddeutsche.de/politik/extremismus-in-die-son-nemit-abu-adam-1.3709409>; 6.6.2018.
- Strobl, Rainer/Lobermeier, Olaf/Heitmeyer, Wilhelm* (Hrsg.) 2012: Evaluation von Programmen und Projekten für eine demokratische Kultur, Wiesbaden.
- Teune, Simon* 2018: Warum wir nicht vom „Extremismus“ reden sollten, PRIF Blog, <https://blog.prif.org/2018/04/30/warum-wir-nicht-vom-extremismus-reden-sollten/>; 15.7.2018.
- Tilley, Nicholas* 2010: Realistic Evaluation and Disciplinary Knowledge: Applications from the Field of Criminology, in: Vaessen, Jos (Hrsg.): Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences, New Brunswick, NJ, 203–236.
- Tracy, Sarah J.* 2010: Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research, in: Qualitative Inquiry 16: 10, 837–851.

- Trautmann, Catrin/Kahl, Wolfgang/Zick, Andreas* 2017: Prävention von islamistischer Radikalisierung und Gewalt. Eine Systematisierung von Präventionsansätzen in Deutschland, in: *forum kriminalprävention* 2017: 1, 3–9.
- Trautmann, Catrin/Zick, Andreas* 2016: Systematisierung von in Deutschland angebotenen und durchgeführten (Präventions-)Programmen gegen islamistisch motivierte Radikalisierung außerhalb des Strafvollzugs, https://www.kriminalpraevention.de/files/DFK/dfk-publicationen/2016_systematisierung_islamismuspr%c3%a4vention_kurzfassung_2.pdf; 15.7.2018.
- Uhl, Katrin/Ulrich, Susanne/Wenzel, Florian M.* (Hrsg.) 2004: *Evaluation politischer Bildung. Ist Wirkung messbar?*, Gütersloh.
- Uhlmann, Milena* 2017: Evaluation der Beratungsstelle „Radikalisierung“. Abschlussbericht (BAMF Forschungsbericht 31), https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Forschungsberichte/fb31-evaluation-beratungsstelle-radikalisierung.pdf?__blob=publicationFile; 6.6.2018.
- Unger, Christian* 2016: Die mühsame Suche nach den Wegen aus der Radikalität, in: *Der Westen*, 29.10.2016, <https://www.derwesten.de/politik/islamismus-die-muehsame-suche-nach-wegen-aus-der-radikalitaet-id12314633.html>; 6.6.2018.
- von *Kardorff, Ernst* 2017: Qualitative Evaluationsforschung, in: *Flick, Uwe/von Kardorff, Ernst/Steinke, Ines* (Hrsg.): *Qualitative Forschung*, Reinbek, 238–250.
- Westhorp, Gill* 2014: *Realist Impact Evaluation: An Introduction* (Methods Lab Publication), <https://www.odi.org/publications/8716-realist-impact-evaluation-introduction>; 15.7.2018.
- Widmer, Thomas* 1996: *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen*, Bern.
- Widmer, Thomas/Beywl, Wolfgang/Fabian, Carol* (Hrsg.) 2009: *Evaluation. Ein systematisches Handbuch*, Wiesbaden.
- WKKF (W.K. Kellogg Foundation)* 1998: *Logic Model Development Guide*, <https://www.bttop.org/sites/default/files/public/W.K.%20Kellogg%20LogicModel.pdf>; 13.6.2018.
- Wyatt Knowlton, Lisa/Phillips, Cynthia C.* 2009: *The Logic Model Guidebook: Better Strategies for Great Results*, Los Angeles, CA.
- Yngborn, Annalena/Hoops, Sabrina* 2018: Das Logische Modell als Instrument der Evaluation in der Kriminalitätsprävention im Kindes- und Jugendalter, in: *Walsh, Maria/Pniewski, Benjamin/Kober, Marcus/Armborst, Andreas* (Hrsg.): *Evidenzorientierte Kriminalprävention in Deutschland: Ein Leitfaden für Politik und Praxis*, Wiesbaden, 349–368.

WEITERE TEXTE DER REPORTREIHE „GESELLSCHAFT EXTREM“

Was ist Radikalisierung? Präzisierungen eines umstrittenen Begriffs (PRIF Report 5/2018)

Hande Abay Gaspar // Christopher Daase // Nicole Deitelhoff // Julian Junk // Manjana Sold

Radikalität und Radikalisierung werden heutzutage als zentrale Kennzeichen der globalen politischen Krise angesehen. Das täuscht darüber hinweg, wie ambivalent der Begriff ist. Dieser Report plädiert für ein weites Verständnis von Radikalisierung, um die ganze Bandbreite von Radikalisierungsphänomenen in den Blick nehmen zu können: von der Radikalisierung ohne Gewalt über die Radikalisierung in die Gewalt bis hin zur Radikalisierung in der Gewalt. Damit trägt er den verschiedenen Facetten des Radikalisierungsbegriffs stärker Rechnung, denn Radikalität kann politisch durchaus produktiv sein. Ein breiter Radikalisierungsbegriff verschließt sich weder der Kritik an Beschränkungen von Freiheitsrechten noch der Beförderung von Stigmatisierung und löst sich aus der scheinbar untrennbaren Verknüpfung mit unmittelbaren Gefährdungslagen. Er öffnet den diskursiven und regulativen Raum im Bereich der primären, sekundären und tertiären Prävention.

Radikalisierung von Individuen: Ein Überblick über mögliche Erklärungsansätze (PRIF Report 6/2018)

Fabian Srowig // Viktoria Roth // Daniela Pisoiu // Katharina Seewald // Andreas Zick

Warum radikalisiert sich Individuen? Dieser Report gibt einen systematischen Überblick über den Forschungsstand zu den Ursachen und Folgen der Radikalisierung von Individuen und beschreibt wie diese in Wechselwirkung und Interaktion mit anderen Personen, sozialen Gruppen sowie Organisationen oder Institutionen stattfinden. Die Aneignung extremistischer Denkmuster sowie die Zugehörigkeit zu einer extremistischen Gleichaltrigengruppe im Jugendalter helfen bei der Befriedigung allgemeiner Bedürfnisse wie Anerkennung und Gruppenzugehörigkeit, aber auch bei der Reduktion von Unsicherheiten und Identitätskonflikten. Ideologien bieten Individuen nachvollziehbare Deutungsmuster und individuelle Handlungsalternativen für spezifische Problemlagen an. Der Report leitet aus seinen Erkenntnissen Vorschläge für zukünftige präventive und therapeutische Maßnahmen ab.

Brückennarrative: Verbindende Elemente für die Radikalisierung von Gruppen (PRIF Report 7/2018)

David Meiering // Aziz Dziri // Naika Foroutan (mit Simon Teune // Esther Lehnert // Marwan Abou-Taam)

Radikale Gruppen stellen nicht nur die Sicherheitsbehörden, sondern die gesamte Gesellschaft vor enorme Herausforderungen. Dieser Report arbeitet aus der bestehenden Forschung heraus, wie Radikalisierungsprozesse innerhalb und zwischen Gruppen ablaufen und welche Rolle derartige Gruppenprozesse im gesamtgesellschaftlichen Kontext spielen. Er fokussiert insbesondere auf die Schnittmengen bestimmter ideologischer Elemente unterschiedlicher radikaler Gruppen. Diese Gemeinsamkeiten werden im Report als Brückennarrative bezeichnet. Das erste dieser Narrative umfasst Anti-Imperialismus, Anti-Modernismus und Anti-Universalismus und hat als gemeinsamen Fluchtpunkt den Antisemitismus. Im zweiten Brückennarrativ, dem Antifeminismus, treffen sich völkische Nationalisten, christliche und islamische Fundamentalisten und islamistische Dschihadisten. Das dritte Brückennarrativ bildet die Vorstellung, im (legitimen) Widerstand zu handeln und dadurch Gewalt zu rechtfertigen. Der Report legt dar, wie wichtig es ist, diese Narrative in der Präventionsarbeit zu berücksichtigen, das heißt, Maßnahmen zu entwickeln, die das gemeinsame ideologische Muster verschiedener radikaler Gruppen ansprechen.

Radikalisierung der Gesellschaft? Forschungsperspektiven und Handlungsoptionen (PRIF Report 8/2018)

*Eva Herschinger // Kemal Bozay // Oliver Decker // Magdalena von Drachenfels // Christian Joppke
(mit Klara Sinha)*

Welche Faktoren begünstigen eine gesamtgesellschaftliche Radikalisierung? Es gibt wenige Arbeiten in der internationalen und nationalen Radikalisierungsforschung, deren Interesse direkt auf die gesellschaftliche Ebene gerichtet ist, und die diskutieren, welche Wirkung radikalisierte Gruppen, Milieus und Schichten auf die Gesamtgesellschaft und ihre potenzielle Radikalisierung haben. Dieser Report arbeitet die aktuelle Forschung auf und diskutiert begünstigende Faktoren einer gesamtgesellschaftlichen Radikalisierung. Gesellschaftliche Radikalisierung entsteht in dem Maße, in dem die Legitimität des politischen Systems in Frage gestellt wird und eine Abkehr von herrschenden sozialen Normen im politischen Umgang, insbesondere eine Abkehr von der Ablehnung politischer Gewalt, stattfindet. Die Radikalisierung Einzelner, wie auch von Gruppen, Milieus oder Schichten kann gesamtgesellschaftliches Radikalisierungspotenzial bergen. Dabei können gesellschaftspolitische Veränderungen in Summe zu nachlassender gesellschaftlicher Kohäsion führen. Angesichts dieser Möglichkeit fordern die Autorinnen und Autoren gesellschaftliche Resilienz zu stärken sowie die öffentliche Debatte zu zivilisieren.

Herausforderung Deradikalisierung: Einsichten aus Wissenschaft und Praxis (PRIF Report 9/2018)

Till Baaken // Reiner Becker // Tore Bjørgo // Michael Kiefer // Judy Korn // Thomas Mücke // Maximilian Ruf // Dennis Walkenhorst

Verglichen mit dem Themenkomplex „Radikalisierung“ wurde „Deradikalisierung“ in der Wissenschaft bisher eher zweitrangig behandelt. Dieser Report arbeitet systematisch die zentralen Erkenntnisse aus der theoretischen Literatur und aus der Deradikalisierungspraxis auf. Es zeigt sich, dass zentrale Akteure aus Praxis, Wissenschaft, (Sicherheits-)Behörden und Politik nicht nur unterschiedliche Definitionen verwenden, es herrscht auch keine Einigkeit darüber, was Deradikalisierung (praktisch) zu bedeuten hat. Hinzu kommt, dass die Trägerlandschaft der Extremismusprävention in Deutschland so divers ist wie das föderale System der Bundesrepublik. Das in Deutschland bestehende Hybridmodell aus staatlichen und zivilgesellschaftlichen Zuständigkeiten sowie die Vielfalt an Ansätzen und Profilen der Beratenden können, bei richtiger Akzentuierung, als Chance für die Arbeit gewertet werden. Der Report schließt mit entsprechenden Handlungsempfehlungen für Entscheidungsträgerinnen und -träger.

Die Rolle des Internets und sozialer Medien für Radikalisierung und Deradikalisierung (PRIF Report 10/2018)

Peter Neumann // Charlie Winter // Alexander Meleagrou-Hitchens // Magnus Ranstorp // Lorenzo Vidino

Welche Rolle spielen die Möglichkeiten des Internets bei der Radikalisierung von Individuen und Gruppen? Dieser Report liefert eine Übersicht über die bestehende Forschung. Er geht der Frage nach, wie und warum extremistische Organisationen und Individuen das Internet verwenden. Darüber hinaus diskutiert er Möglichkeiten (und Grenzen), wie Online-Extremismus wirksam entgegengetreten werden kann. Es zeigt sich unter anderem, dass die Nutzung von Internetangeboten durch extremistische Gruppen oftmals eher laienhaft und herkömmlich ist. Dies sollte auch bei der Entwicklung von Gegenmaßnahmen beachtet werden. Reine Online-Gegenmaßnahmen stoßen an Grenzen, da Offline- und Online-Radikalisierung auf das engste verschränkt sind und nicht getrennt betrachtet werden können. Um der Struktur und den Nutzungsgewohnheiten des Internets gerecht zu werden, bedarf es einer engen Interaktion zwischen öffentlichen und privaten Akteuren in der Strategieentwicklung. Der öffentliche Sektor sollte hier Anreize setzen und muss die Konsequenzen von kritischen Maßnahmen im Bereich der Zensur gründlicher als bisher abwägen.

Alle Reporte der Reihe sind hier abrufbar: <https://gesellschaftstextrem.hsfk.de/ergebnisse/prif-reports/>

PRIF REPORT

Die PRIF Reports analysieren Hintergründe politischer Ereignisse und Entwicklungen und präsentieren wissenschaftliche Forschungsergebnisse in Deutsch oder Englisch.

Kreuzer, Peter (2018): Dealing with China in the South China Sea: Duterte Changing Course, PRIF Report 3/2018, Frankfurt/M.

Peace Research Institute Frankfurt (2018): Coercion and Peace. PRIF's New Research Program, PRIF Report 2/2018, Frankfurt/M.



www.hsfk.de/PRIF-Reports
www.hsfk.de/HSFK-Reports

PRIF SPOTLIGHT

Die PRIF Spotlights diskutieren aktuelle politische und gesellschaftliche Themen.

Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (2018): Im Auftrag der Gerechtigkeit. Die Verleihung des Hessischen Friedenspreises 2017 an Carla del Ponte, PRIF Spotlight 7/2018, Frankfurt/M.

Deitelhoff, Nicole/Dembinski, Matthias/Peters, Dirk (2018): Nach vorn, um nicht zurückzufallen. Deutsch-französische Initiativen zur Zukunft der EU-Außen- und Sicherheitspolitik, PRIF Spotlight 6/2018, Frankfurt/M.



www.hsfk.de/PRIF-Spotlights

PRIF BLOG

Auf dem PRIF Blog erscheinen Beiträge zu aktuellen politischen Fragen und Debatten der Friedens- und Konfliktforschung. Die Blogbeiträge erscheinen in loser Folge in Deutsch oder Englisch.



<https://blog.prif.org/>

PRIF Reports und PRIF Spotlights sind Open-Access-Publikationen und können kostenlos auf www.hsfk.de heruntergeladen werden. Sie möchten die digitalen Ausgaben abonnieren? Bitte wenden Sie sich an: publikationen@hsfk.de.

 www.facebook.com/HSFK.PRIF

 www.twitter.com/HSFK_PRIF

 <https://blog.prif.org/>

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

GESELLSCHAFT **EXTREM**

ANDREAS ARMBORST // JANUSZ BIENE //
MARC COESTER // FRANK GREUEL //
BJÖRN MILBRADT // INGA NEHLEN

EVALUATION IN DER RADIKALISIERUNGSPRÄVENTION: ANSÄTZE UND KONTROVERSEN

Dieser Report nimmt das gesteigerte öffentliche Interesse an verschiedenen Maßnahmen und Ansätzen der Radikalisierungsprävention zum Ausgangspunkt einer Diskussion über Evaluation. Evaluationen helfen zu verstehen, wie die Prävention von Radikalisierung und Extremismus im gesellschaftlichen Kontext wirkt. Sie können damit wesentliche Anhaltspunkte für die häufig artikulierte Frage nach den sichtbaren Erfolgen von Prävention liefern. Gleichzeitig existieren in der Debatte um die sogenannte „evidenzbasierte“ Prävention teilweise überzogene Erwartungen hinsichtlich der Leistungsfähigkeit und Durchführbarkeit von Wirkungsevaluationen. Das berechtigte Interesse an belastbaren Wirksamkeitsnachweisen stößt bei der Planung und Umsetzung von Evaluationsstudien im Bereich der Deradikalisierung, Distanzierung und Prävention von Radikalisierung auf beträchtliche Herausforderungen. Dieser Report geht auf einige dieser Schwierigkeiten ein und zeigt beispielhaft verschiedene Ansätze dafür, wie sich Evaluationen im Rahmen realistischer Möglichkeiten umsetzen lassen. Die Idee einer „evidenzbasierten“ Prävention kann nur dann funktionieren, wenn der Evaluationsforschung die Eigenheiten, Widersprüche und Kontroversen in Wissenschaft und Praxis bewusst sind und sie diese kritisch reflektiert.