
Evaluationsansätze in der Radikalisierungsprävention

Dr. Michail Logvinov

Vorbemerkung

Der Stellenwert von Evaluationen¹ steht in einem direkten Zusammenhang mit dem Ausbau von Infrastrukturen für soziale Innovationen. Auch das öffentliche Interesse an bestimmten sicherheitsrelevanten Phänomenen und zu verändernden Zuständen trägt zum Legitimationsdruck im Hinblick auf sozialpolitische Initiativen bei. Des Weiteren spielt das Verhältnis zwischen den zu prävenierenden sozialen Missständen, dem finanziellen Aufwand und den beobachtbaren Ergebnissen dahingehend eine wichtige Rolle, als sich die Frage nach (möglichen) Zweck-Mittel-Konflikten und nicht intendierten Wirkungen von Interventionen in sozialen Kontexten aufdrängt.

Es nimmt vor diesem Hintergrund nicht wunder, dass Bewertungen von extremismusrelevanten Präventionsprogrammen und Deradikalisierungsinitiativen – zu Recht – auf eine große Aufmerksamkeit stoßen, die zugleich aus verschiedenen pluralen Interessen sowie politischen Blickwinkeln resultiert. Grundlegend ist indes der Zusammenhang zwischen gesamtgesellschaftlichen Sicherheitsinvestitionen und Sicherheitszielen in einem demokratischen System. Insofern stehen, darin eingebettet, stets die Interdependenzen der relevanten gesellschaftlichen Akteure auf dem Prüfstand, wie sie Evaluationen konzipieren sowie systemisch zweckvoll und zielorientiert umsetzen. Zu beachten ist, dass in Deutschland derzeit zwei relativ voneinander abgekoppelte „Evaluationsinseln“ auszumachen sind, die jeweils an verschiedene Paradigmen der Politik und Forschung zu Extremismusphänomenen gebunden sind: zum einen die kriminalpolitisch intendierten Ansätze der evidenzbasierten Wirkungsevaluation und zum anderen jene der Sozialpädagogik oder pädagogisch orientierten Jugendhilfe.

¹ Siehe die nachfolgenden Definitionen: „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making” (Mertens 1998: 5). „[...] evaluation means a study designed and conducted to assist some audience to assess an object’s merit and worth” (Stufflebeam 2000: 35). Rossi et al. (1988: 3) bestimmen „Evaluation” als „systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluationsforschung bezeichnet den gezielten Einsatz sozialwissenschaftlicher Forschungsmethoden zur Verbesserung der Planung und laufenden Überwachung sowie zur Bestimmung der Effektivität und Effizienz von [...] sozialen Interventionsmaßnahmen“. Es ist überdies eine Art „Metaforschung“ in Deutschland entstanden, die die Methoden und Verfahren der Praxisforschung bzw. wissenschaftlichen Begleitung von Programmen/Projekten und der Evaluationsforschung auf den Prüfstand stellt. Im Gegensatz zur Meta-Evaluation wird hier mit Daten unterschiedlicher Qualität gearbeitet.

1. Entwicklungsstränge und Paradigmen der Evaluationsforschung

Die Evaluationsforschung stellt ohne Übertreibung einen Forschungszweig mit Tradition dar. Die ersten Evaluationsstudien als Instrumente zur Explikation sowie Generierung von Erfahrungswissen und dessen Bewertung anhand methodisch festgelegter Kriterien kamen bereits im 19. Jahrhundert in den USA und Großbritannien zustande. Im „Zeitalter der Reformen“ (1792-1900) wurden zahlreiche Programme zur Einführung von sozialen Innovationen umgesetzt, deren Bewertung – nach der ersten, impressionistischen Phase – mit standardisierten Datenerhebungsverfahren und experimentellen Methoden einherging (Madaus/Stufflebeam 2000: 4). Es seien an dieser Stelle die Evaluationsforschungen von Joseph Rice zwischen 1887 und 1898 genannt. Darauf folgten etwa im exponierten Bildungs- und Erziehungsbereich, in dem eine der wichtigsten Wurzeln von Evaluationen liegt, mindestens fünf weitere Phasen mit ihren jeweiligen wissenschaftlichen und evaluationstheoretischen Schwerpunktsetzungen (Madaus et al. 1983: 6 ff.).

1. Im „Zeitalter der Effizienz und Testung“ (1900-1930) spielte der Gedanke des wissenschaftlichen Managements (vgl. Taylorismus) eine herausragende Rolle, weshalb Systematisierung, Standardisierung und Effizienzsteigerung von Bildungsprogrammen befördert und ihre Performanz mit begleitenden Tests untersucht wurden.
2. Der Zeitraum zwischen 1930 und 1945 stand unter dem Einfluss des „Vaters der Bildungsevaluation“ und Begründers der Programmevaluation, Ralph W. Tyler, dessen Beitrag zur Outcome-Forschung kaum zu überschätzen ist. Im Gegensatz zum experimentellen Forschungsdesign von etwa Rice fokussierte dieses Evaluationsverfahren Programmziele sowie Effekte und legte den Schwerpunkt auf den Vergleich von intendierten und tatsächlich erreichten Zielen.
3. Die nächste Phase (etwa 1946 bis 1957) zeichnete sich eher durch eine „Ignoranz“ im Hinblick auf die Evaluation sozialpolitischer Maßnahmen und durch eine vermehrte Standardisierung sowie „Technologisierung“ der Testverfahren im psychologischen und pädagogischen Bereich, aber auch durch die Manualisierung des Ansatzes von Tyler aus.
4. Ab Ende der 1950er bis Anfang der 1970er Jahre erfuhr die angelsächsische Evaluationsforschung einen bis dahin nicht gekannten Entwicklungsschub, der mit der Ausweitung und Umsetzung von Bildungs- und Wohlfahrtsprogrammen einherging. Infolge dieser Entwicklung entstand ein millionenschwerer Markt für Evaluationen von sozialen Dienstleistungsprogrammen. „Das ‚Evaluationsbusiness‘ wurde zu einer Wachstumsindustrie“ (Stockmann 2004: 24 f.). Bereits Anfang der 1960er Jahre kritisierte etwa Lee J. Cronbach die Oberhand von Tests und experimentellen Methoden in summativen Programmevaluationen, die dazu führten, dass die Evaluationsergebnisse keine Relevanz für die Gestaltung sowie Optimierung von Programmen und Curricula-Entwicklung aufwiesen, und formulierte sein anwendungsorientiertes Evaluationsverständnis (Cronbach 1982). Nach der teils massiven Kritik seitens des „Phi Delta Kappa National Study Committee on Evaluation“ entstand Ende der 1960er Jahre eine Reihe neuer Bewertungsverfahren und -modelle.
5. Mit steigenden Erwartungen der Beteiligten und Betroffenen an Evaluationsstudien setzte ab Anfang der 1970er bis Anfang der 1980er Jahre das „Zeitalter der Professionalisierung“ ein. Infolgedessen wurden Fachperiodika und -gremien etabliert; Universitäten boten Kurse an; es entstanden dezidierte Evaluationstheorien und -standards. Zugleich wurden die bis dahin erarbeiteten Verfahren und Ansätze weiterentwickelt bzw. einer kritischen Prüfung unterzogen, so dass lebendige

Diskussionen und eine Methodenvielfalt die Evaluationsforschung der 1980er Jahre prägten. Als Begleiterscheinung dieses Ausdifferenzierungsprozesses wird in der Literatur ein „Kalter Krieg der Paradigmen“ beschrieben, der aus verschiedenen erkenntnistheoretischen und methodologischen Standpunkten resultierte und eine Zeit lang die Auseinandersetzung mit der Programmevaluation prägte (Lee 2004: 150 f., Madaus/Stufflebeam 2000: 4).

Ein Blick auf die einzelnen Entwicklungsstränge moderner Evaluationsverfahren ermöglicht ein besseres Verständnis der jeweiligen Prämissen und Implikationen der Evaluationsansätze, da die Evaluationsforschung im Hinblick auf weltanschauliche und erkenntnistheoretische Wurzeln von mindestens vier Trends geprägt wurde: dem mehr oder minder radikalen Rationalismus der 1960er Jahre und dem Experimentalgedanken der evidenzbasierten Forschung der 1980/90er Jahre, dem pluralistisch-dialogischen Ansatz der 1970/80er Jahre und dem neo-liberalen Konzept (Vedung 2010). Diese Trends lassen sich anhand einer paradigmensorientierten Systematisierung der Evaluationsverfahren veranschaulichen.

Evaluationsparadigmen entstanden einerseits in bestimmten sozialen und wissenschafts-historischen Kontexten, andererseits waren die jeweiligen Evaluationskonzepte in den methodischen Traditionen einzelner Disziplinen verwurzelt. So verwundert es kaum, dass Tests sowie Messungen der Zielerreichung und (Quasi-)Experimente in pädagogischen und psychologischen Verfahren ihren Ursprung hatten (Mertens 2004: 46).

Evaluationskonzepte waren zugleich durch bestimmte erkenntnistheoretische Annahmen geprägt. Der *Empirismus* und *Positivismus* postulierten etwa das Vorhandensein einer objektiven Realität, die Objektivität und Neutralität der Wissenschaft und verfolgten daraus resultierend einen methodischen Rigorismus (Lee 2004: 151, Stockmann 2007: 42). Im Gegensatz zu realistischen Forschungen streb(t)en die positivistisch orientierten Evaluationen ein von intervenierenden (Dritt-)Variablen befreites Setting an, in dem ein Ursache-Wirkung-Zusammenhang möglichst präzise eruiert werden soll.² In der Evaluationsforschung ließen sich vor allem die Arbeiten zu (quasi-)experimentellen Designs von Donald T. Campbell dieser Tradition zuordnen. Als Angehöriger der Sozialindikatorenbewegung hegte Campbell das Ideal einer experimentierenden Gesellschaft und deutete in diesem Kontext soziale Innovationen als Feldversuche, bei denen die Testung von Interventionen unter methodisch kontrollierten Bedingungen einen Kausalzusammenhang zwischen ergriffenen Maßnahmen und ihren Wirkungen offenlegen sollte (Giel 2013: 63 f.).

Im *interpretativen/konstruktivistischen Paradigma* wird die vermeintlich objektive Realität als Konstruktion verschiedener konkurrierender Perspektiven aufgefasst (bspw. Egon Guba, Yvonna Lincoln u. a.). Bei der Evaluation werden vor allem die Standpunkte, Werte, Bedürfnisse und Interessen der Betroffenen und Beteiligten betont, die mit nur quantitativen Methoden kaum zu erfassen wären (Stockmann 2004: 43). So hat Robert Stake in seiner responsiven Evaluation³ den Methodenpluralismus und die Aushandlung von Methoden mit Betroffenen und Beteiligten favorisiert, um die Auswirkungen der Programmaktivitäten möglichst vielseitig zu untersuchen (Lee 2004: 151).

² Die randomisierten Experimente gelten in diesem Paradigma als Königsweg.

³ „A responsive evaluation is a search and documentation of program quality. The essential feature of the approach is a responsiveness to key issues or problems, especially those recognized by people at the site. It is not particularly responsive to program theory or stated goals but more to stakeholder concerns. Its design usually develops slowly, with continuing adaptation of evaluation goal-setting and data-gathering in pace with the evaluators becoming well acquainted with the program and its contexts” (Stake 2003: 63).

Das *emanzipatorische/transformative Evaluationsparadigma* rückt bei der Bewertung sozialer Innovationen vor allem jene Faktoren in den Vordergrund, die auf verschiedene soziale Realitäten der Betroffenen Einfluss nehmen. Es handelt sich unter anderem um soziale, wirtschaftliche, politische oder etwa genderbezogene Variablen, die die jeweiligen Gruppen charakterisieren. In diesem multifaktoriellen Setting bestünde die Rolle der Evaluation darin, mit unterschiedlichen sozialen Gruppen zu interagieren und mittels Methodenpluralismus die Diversität sowie die Stimmen der „Erniedrigten und Beleidigten“ einzufangen, um Machtungleichgewichte und Ungleichheiten bei der Gestaltung und Umsetzung von sozialen Innovationen aufzuzeigen (Lee 2004: 151).

Das *anwendungsorientierte (auch nutzen- oder nutzungsorientierte)* Paradigma reagierte demgegenüber auf die Beobachtung der Evaluationspraxis, der zufolge eine unübersehbare Lücke zwischen den Evaluationsergebnissen und ihrem Mehrwert für die jeweiligen Zielgruppen klappte. Es zeigte sich, dass etwa Entscheidungen über die (Nicht-)Fortführung von Programmen anhand abweichender, von den Evaluationen nicht zwingend hervorgehobenen Kriterien gefällt wurden (Giel 2013: 73, Astbury 2017: 226). Daraus schlussfolgerte etwa Michael Q. Patton, dass anscheinend Diskrepanzen zwischen den Nutzenerwartungen sowie Nutzungsabsichten der Stakeholder und dem Erkenntnisinteresse sowie der exklusiven Methodenauswahl der Evaluierenden vorlagen (Beywl/Giel 2012: 101). Daraus resultierte ein pragmatischer Evaluationsansatz, dessen Schwerpunkt auf dem beabsichtigten Gebrauch durch vorgesehene Nutzer liegt.⁴ Dabei spielt es keine Rolle, welcher Zweck (formativ, summativ), welche Datenerhebungsverfahren (quantitativ, qualitativ, gemischt) und welcher Fokus einer Evaluation zugrunde liegt – alle Entscheidungsprozesse über diese Aspekte erfolgen in Kooperation mit klar definierten Nutzergruppen.⁵

Im Gegensatz zu rein ergebnisorientierten Ansätzen konzentriert sich die *Evaluation der Programmintegrität* auf die Ebenen der Interventionskonzepte, der Programmdauer und -intensität, der Ressourcen und des Verhältnisses zwischen Interventionen, Intervenierenden und Klienten (Quay 1977). So lässt etwa das 2005 aus „Correctional Program Assessment Inventory – 2000“ hervorgegangene Instrument „Evidence-Based Correctional Program Checklist“ fünf Programmbereiche anhand von 73 evidenzbasierten Indikatoren auf ihre Entsprechung den Qualitätsstandards hin untersuchen.⁶

⁴ „Utilization-focused program evaluation (as opposed to program evaluation in general) is evaluation done for and with specific intended primary users for specific, intended uses. [...] Utilization-focused evaluation is concerned with how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on intended use by intended users“ (Patton 2003: 224, 223).

⁵ Voraussetzungen und die Logik der nutzenorientierten Evaluationsforschung wird wie folgt umrissen: „In any evaluation there are many potential stakeholders and an array of possible uses. Utilization-focused evaluation requires moving from the general and abstract, i.e., possible audiences and potential uses, to the real and specific: actual primary intended users and their explicit commitments to concrete, specific uses. The evaluator facilitates judgment and decision-making by intended users rather than acting solely as a distant, independent judge. Since no evaluation can be value-free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations. In essence, utilization-focused evaluation is premised on the understanding that evaluation use is too important to be merely hoped for or assumed. Use must be planned for and facilitated“ (Patton 2003: 223).

⁶ „The CPC comprises five domains (compared to the CPAI’s six domains) and splits the domains into two basic areas. The first area, capacity, measures the degree to which a program has the ability to offer evidence-based interventions. The domains in this area are program leadership and development, staff characteristics, and quality assurance. The second area of the CPC, content, assesses the extent to which a program adheres to the RNR principles, and consists of an offender assessment and a treatment characteristics domain. [...] Since 2005, when the CPC was developed, the general CPC tool has been adapted to assess specific types of programs that have their own subset of research within the broader context of correctional treatment programs. These adaptations include assessments for Community Supervision Agencies (CPC-CSA), general correctional treatment groups (CPC-GA), and Drug Court programs (CPC-DC)“ (Duriez et al. 2018: 8 ff.).

Es sei darauf hingewiesen, dass der paradigmensorientierte Systematisierungsversuch von mindestens zwei Dutzend Evaluationsansätzen⁷ – unter anderem zielorientierte, managementorientierte, konsumentenorientierte, expertenorientierte und partizipative Ansätze – nur eine idealtypische Annäherung an die jeweiligen Evaluationsmethodologien ermöglicht. Zugleich wird aber bereits auf der höchsten Abstraktionsebene ersichtlich, dass die angesprochenen Forschungstraditionen jeweils unterschiedliche Schwerpunkte bzw. Ebenen fokussieren und teils abweichende Hauptziele – bspw. Erkenntnisse, Kontrolle, (Weiter-)Entwicklung, Legitimation – verfolgen, wobei sie unterschiedlichen Evaluationsschulen anhängen und verschiedene Rollenmodelle – Datensammlung, Beobachtung, Urteilsbildung, Moderation – voraussetzen (Scriven 1983, Stockmann 2007: 44 f.). Zielorientierte oder diskrepanzorientierte Ansätze folgen einer anderen Logik als dezidierte Wirkungsevaluationen; methoden- oder mechanismenorientierte Forschungen dienen einem anderen Zweck als Evaluationen, die sich einer Sozialagenda verschreiben, und entscheidungs- sowie rechenschaftsorientierte Evaluationen nehmen andere Zielgruppen in den Blick als die konsumentenorientierte Praxisforschung. Die meisten Konzeptionen und Verfahren beanspruchen zugleich für sich, mehr oder minder evidenzorientiert zu sein und Effekte von Programmen – (nicht-)intendiert, (nicht-)antizipiert, positiv/negativ, (nicht-)expliziert, positiv/negativ – offenzulegen.

Evaluation in der Radikalisierungsprävention

Ein unvoreingenommener Blick auf die Begleit- und Evaluationsforschung in der Kriminal- und Extremismusprävention offenbart ein durchaus facettenreiches Forschungsfeld mit einem ausgeprägten Methodenpluralismus. So hat etwa der Landespräventionsrat Niedersachsen mit seiner „Grünen Liste Prävention“ Pionierarbeit in Deutschland geleistet und verschiedene Programme nach festgelegten Gütekriterien auf Grade ihrer Wirksamkeit hin analysiert, um Interventionen mit nachgewiesenen Effekten zu identifizieren (Groeger-Roth 2018: 159 ff.). Bei der Einordnung in drei verschiedene Stufen – Effektivität *theoretisch gut begründet*, Effektivität *wahrscheinlich* und Effektivität *nachgewiesen* – werden die theoretische Begründung von Programmen selbst, ihre didaktische Umsetzung und die Qualität von Evaluationen systematisch geprüft (Groeger-Roth/Hasenpusch 2011: 3 f.). Als Königsweg gelten hier Programme, deren Evaluationen Experimentaldesigns mit einer großen Anzahl an Interventions- und Vergleichsgruppen zugrunde liegen und die eine „follow-up“-Messung mindestens sechs Monate nach der Maßnahme durchgeführt haben.

Einen vergleichbaren Ansatz verfolgt das Nationale Zentrum für Kriminalprävention (NZK) mit seiner Datenbank WESPE, in die auch dezidierte Extremismuspräventionsprogramme aufgenommen wurden. Darüber hinaus veröffentlicht das Zentrum Forschungssynthesen, die die projektbezogenen Informationen reflektieren.⁸ Auch die Forschungs- und Beratungsstelle Extremismus/Terrorismus des Bundeskriminalamtes (BKA) hat sich der Erforschung der

⁷ Vor 20 Jahren identifizierte Stufflebeam (2000) in seiner lesenswerten Abhandlung etwa 22 Evaluationsansätze: 1) Public Relations-Inspired Studies, 2) Politically Controlled Studies, 3) Objectives-Based Studies, 4) Accountability, Particularly Payment By Results Studies, 5) Objective Testing Programs, 6) Outcomes Evaluation As Value-Added Assessment, 7) Performance Testing, 8) Experimental Studies, 9) Management Information Systems, 10) Benefit-Cost Analysis Approach, 11) Clarification Hearing, 12) Case Study Evaluations, 13) Criticism and Connoisseurship, 14) Program Theory-Based Evaluation, 15) Mixed-Methods Studies, 16) Decision/Accountability-Oriented Studies, 17) Consumer-Oriented Studies, 18) Accreditation/Certification Approach, 19) Client-Centered Studies (or Responsive Evaluation), 20) Constructivist Evaluation, 21) Deliberative Democratic Evaluation, 22) Utilization-Focused Evaluation.

⁸ Das NZK legte zudem ein Excel-basiertes Evaluationsinstrument für die Islamismusprävention (EvIs) vor, in dem verschiedene Kriterien aus den Bereichen „Individuum“, „kritische Lebensereignisse“, „soziale Strukturen“ und „Religion und Ideologie“ zu unterschiedlichen Zeitpunkten bewertet und im Zeitverlauf abgebildet werden können (Ullrich et al. 2019).

Prävention phänomenübergreifender Gewaltkriminalität verschrieben und die Präventionslandschaften sowie Maßnahmen und Evaluationskontexte untersucht (Lützinger/Gruber 2017).

Auf der europäischen Ebene liegen mit dem „IMPACT Europe Toolkit“ eine Interventionsdatenbank und Werkzeuge zur Planung und Umsetzung von (Selbst-)Evaluationen vor. In einem Synthese-Bericht wurden im Rahmen eines von der EU geförderten Projekts die neuesten Erkenntnisse über die europäischen Maßnahmen der Extremismusprävention und Programmevaluation zusammengefasst (Hemert et al. 2014). RAN Centre of Excellence veröffentlichte seinerseits ein Ex Post-Paper mit dem Titel „Guideline Evaluation of PCVE Programmes and Interventions“ (Molenkamp et al. 2018). Auch die RAND-Corporation stellt einen „Werkzeugkasten“ zur (Selbst-)Evaluation von Projekten mit detaillierten Anweisungen und Auswertungstipps mittels Excel zur Verfügung (Helmus et al. 2017).

Während die erwähnten Institutionen und Forschungsverbände auf den Ebenen der Forschungssynthesen, Metaforschungen und Prozessstandardisierungen wirken, betreiben einige (Praxis-)Forschungsinstitute die Begleitforschung und Evaluation von zahlreichen Präventionsprogrammen und -maßnahmen in verschiedenen Settings. Auf der Bundesebene seien exemplarisch die Präventionsprogramme des BMI/BAMF (Uhlmann 2017) und des BMFSFJ⁹ genannt.

Wie anhand dieser, zugegebenermaßen unvollständigen¹⁰, Übersicht ersichtlich wird, herrscht auf Seiten der Praxis- und Metaforschung ein reger Betrieb. Zugleich springt ins Auge, dass wir es hier mit paradigmatisch differenzierten Forschungsgemeinschaften zu tun haben, deren Schwerpunkte auf verschiedenen Ebenen liegen. Mitunter liegen Forschungsbemühungen verschiedene Logiken, die die Belange der Praxis auf unterschiedliche Art und Weise reflektieren, zugrunde. Daher bedarf es einer „Frameangleichung“ sowie Anpassung von theoretischen und praktischen Blickwinkeln, ohne die es nur schwer möglich sein wird, zu einem Konsens zu gelangen.

Denn des Öfteren werden in der Forschung Befunde generiert und/oder aufgegriffen, die die Evaluationspraxis als defizitär herausstellen. So schlussfolgerten etwa Lützinger und Gruber (2017: 18 f.): „Aus vorangegangenen Untersuchungen wurde deutlich, dass Informationen hinsichtlich durchgeführter Evaluationsmaßnahmen in der Regel äußerst dürftig und wenig transparent sind. Ein voneinander Lernen wird hierdurch deutlich erschwert [...]. Möglicherweise ist die defizitäre Evaluationspraxis bzw. öffentlich zugängliche Dokumentation von Evaluationsberichten auch der Grund, warum sich in unserer Befragung eine deutlich ambivalente Haltung gegenüber dem Thema Evaluation offenbarte“.

„Dürftige“ und/oder „wenig transparente“ Informationen über Evaluationsmaßnahmen sind freilich nicht mit der „defizitären“ Evaluationspraxis gleichzusetzen. Ein Blick auf die Berichte der Begleitforschungen und Programmevaluationen¹¹ ergibt ein facettenreicheres Bild und führt zu dem Schluss, dass hier die management- und steuerungsorientierte Evaluation im Vordergrund steht. Die in die wissenschaftliche Begleitung involvierten Praxisforschungseinrichtungen folgen dem Organisationsmodell der Evaluation und vordergründig dem ziel- sowie anwendungsorientierten Ansatz. Primär genutzt werden die Evaluationsergebnisse durch Programm-

⁹ Siehe vor allem die einschlägigen Programmberichte von DJI, ISS und Camino: „Programmevaluation und wissenschaftliche Begleitung“, unter: <https://www.demokratie-leben.de/foerderperiode-2015-2019/programmevaluation-und-wissenschaftliche-begleitung.html> (22. Mai 2020).

¹⁰ Es fehlen bspw. Evaluationen von Landesprogrammen und kommunalen Maßnahmen der Extremismusprävention (siehe dazu die Präventionsdatenbank der Bundeszentrale für Politische Bildung und die einschlägigen BKA-Erhebungen).

¹¹ Siehe „Programmevaluation und wissenschaftliche Begleitung“, unter: <https://www.demokratie-leben.de/foerderperiode-2015-2019/programmevaluation-und-wissenschaftliche-begleitung.html> (22. Mai 2020).

verantwortliche der zuständigen Ministerien („Program Owner“). Das Deutsche Jugendinstitut begleitet unter anderem Modellprojekte und verantwortet die Gesamtevaluation des Bundesprogramms. So berichtet das für die wissenschaftliche Begleitung der Partnerschaften für Demokratie im Rahmen des Bundesprogramms „Demokratie Leben!“ verantwortliche Institut für Sozialarbeit und Sozialpädagogik aus Frankfurt a.M. kontinuierlich über zahlreiche Programmschwerpunkte wie bspw. Zielerreichung, Jugendpartizipation, Evaluation der Weiterbildungsangebote und Gesamtkoordination.

Die in der BKA-Analyse zitierte „Wahrnehmung“ – „Mitunter bestand das Gefühl, das wissenschaftliche Begleitinstitut forsche *im Sinne* des Bundesprogramms, also wenig kritisch und – in Gestalt einer Art Legitimationsforschung – mit dem Ziel einer positiven Bilanz.“ (ebd.: 19) –, bedürfte daher einer fachlichen Einordnung. Denn die mit der Begleitforschung betrauten Praxisforschungseinrichtungen arbeiten in der Tat dem „Program Owner“ zu, indem sie (Struktur-)Daten sammeln und konsolidieren, die jeweiligen Erkenntnisse zwecks Überprüfung der Programmziele ziehen und die Umsetzung der Programmschwerpunkte kontinuierlich dokumentieren. Zugleich wird dem konsumentenorientierten sowie partizipativen Evaluationsansatz durch die Bewertung einzelner Maßnahmen, problemorientierte Handreichungen und Evaluationsmethodik Rechnung getragen. Das Deutsche Jugendinstitut begleitet unter anderem Modellprojekte und verantwortet die Gesamtevaluation des Bundesprogramms. Somit benötigen Meta-Forschungen tiefere Einblicke in die Praxis- und Evaluationsforschung.

Zugleich sind Fragen nach Transparenz und methodischem Vorgehen der Evaluation bzw. wissenschaftlichen Begleitung im Hinblick auf die Programmtheorie sowie -formulierung, Implementation und Ergebniskontrolle mehr als berechtigt. Denn eine das Qualitätsmanagement flankierende Evaluation müsste bekanntlich über reine Tätigkeitsbilanzen hinausgehen und mindestens fünf Qualitätsdimensionen enthalten (Heil u.a. 2001: 32):

1. Angemessenheit der Zielsetzung (Fachlichkeit I),
2. Umsetzung der Zielplanung (Effektivität),
3. optimale Mittelverwendung (Effizienz),
4. Einhaltung professioneller Standards (Fachlichkeit II) und
5. Passung in vorhandene Kontexte ohne unerwünschte Nebenwirkungen (Verträglichkeit).

Dazu tritt die überwölbende Grundfrage, ob durch das jeweilige Programm Einflussgrößen auf das Extremismusgeschehen geschaffen werden, die eine reduktive Wirkung entfalten.

Ein weiteres Beispiel zeigt eindringlich, dass man – dem Diktum von Niklas Luhmann folgend – den Beobachter/die Beobachterin beobachten muss, um nachvollziehen zu können, wie Evidenzen geschaffen werden. In der NZK-Datenbank WESPE findet die Beratungsstelle HAYAT mit sieben von 100 möglichen Punkten im „Evidenz-Index“ Erwähnung. Die Punktevergabe bei HAYAT erfolgte lediglich anhand eines Aufsatzes, in dem das Selbstverständnis, die Verortung und die Handlungslogiken der Beratungsstelle beschrieben werden. Im Gegensatz dazu wurde für die Beurteilung von zwei weiteren Initiativen bzw. Projekten auf jeweils zwei Evaluationsstudien bzw. Abschlussberichte der wissenschaftlichen Begleitung zurückgegriffen. Durch die Subsummierung unter der entsprechenden Überschrift in der Datenbank – „Anzahl der *Evaluationsstudien*“ – entsteht somit ein falscher Eindruck. Dass der Unterschied in der Projektbewertung – sieben gegenüber 30 und 33 (Evidenz-)Punkten – ein Artefakt der „Meta-Forschung“ ist, steht daher

außer Frage. Das NZK hätte die Beratungsstelle HAYAT entweder aus dem Evidenzindex herausnehmen oder die Evaluation der „Beratungsstelle Radikalisierung“ des BAMF-Forschungszentrums reflektieren müssen (Uhlmann 2017). Denn im Endeffekt bedürfen „Evaluationssynthesen“ eines Zugangs zu wissenschaftlich aufbereiteten Daten zu Methoden und Standards der Fachpraxis.

Im Hinblick auf Projekte und Maßnahmen der Radikalisierungsprävention ist die Aufgabe der Evaluation scheinbar einfach – sie soll belegen, was wirkt. Denn es zähle nur, was wirke und das, was wirke, bedeute Evidenz. Etwas komplexer wird es, wenn man bedenkt, dass die Kausalität, d.h. das, was Effekte erzeugt, mit kriteriengeleiteten „Beschreibungen von Zusammenhängen zwischen Ursachen und später auftretenden Wirkungen“ zu tun hat (Lüders/Haubrich 2006: 10). Dies wiederum bedeutet, „dass Kausalität eine Zurechnung von Wirkungen auf Ursachen erfordert, also auch eine Selektion aus einem Endloshorizont von in Betracht kommenden Ursachen und einem zweiten Endloshorizont von in Betracht kommenden Wirkungen“ (Luhmann, ebd.: 9). Soll heißen: Im Endeffekt generiert eine Kausalzuschreibung lediglich eines der möglichen Kausalschemas.

Mit Blick auf soziale Prozesse und Problemlagen scheint dies eine Herausforderung zu sein, die jedoch unter Berücksichtigung des jeweiligen konzeptionellen Rahmens überwunden werden kann. Und genau hier liegt die Grundproblematik des angespannten Verhältnisses zwischen der wirkungsorientierten Meta-Forschung und (Evaluation der) Fachpraxis begründet. Denn verschiedene konzeptionelle Rahmen heben teils abweichende Ursache-Wirkung-Zusammenhänge hervor.

Für die Evaluation von Projekten und Interventionen der Radikalisierungsprävention ergeben sich daraus wichtige methodische Konsequenzen. Zum einen muss im Rahmen der Evaluationsplanung die Frage beantwortet werden, welche Wirkungsmechanismen als Analysegegenstand fungieren sollen: „Daher verlangt die Frage nach den Wirkungen von Interventionen eine plausible Beschreibung der Programmelemente und ihres konzeptionellen Zusammenwirkens. Dies erfordert einen klar abgrenzbaren Untersuchungsgegenstand sowie das Vorliegen begründeter Hypothesen und Annahmen über die Struktur des Gegenstandes und die Beziehung zwischen den verschiedenen Programmaktivitäten und angestrebten Ergebnissen“ (Lüders/Haubrich 2006: 16). Auf der Ebene der Projekte und Initiativen bedeutet dies, dass die Annahmen und Konzepte der jeweiligen Programmbeteiligten als Referenzrahmen der Intervention zu analysieren sind.

Zum anderen sollen die Bewertungsverfahren jener Prozesslogik gerecht werden, die durch die jeweiligen Projekte (Strukturen) entsprechend dem konzeptionellen Referenzrahmen zum Tragen kommt und ein Ergebnis generiert. Das seit den 1980er Jahren weit verbreitete „Struktur-Prozess-Ergebnis“ (SPE)-Schema ist für die Abbildung der Interaktionsqualität im Feld und die Qualitätspraxis der Maßnahmen ausschlaggebend (Beywl 2006: 34).

Zugleich stehen die Struktur-, Prozess- und Ergebnisqualität in einem nicht linearen Verhältnis, wobei die Strukturqualität eine notwendige, aber keine hinreichende Bedingung für die Prozessqualität darstellt; ähnlich verhält es sich zwischen der Prozess- und Ergebnisqualität: „Es muss ein Mindestmaß an Strukturqualität gewährleistet sein, um über Prozessqualität die Ergebnisqualität positiv zu beeinflussen. Andererseits führt eine Verbesserung der Strukturqualität nicht zwangsläufig zu besserer Prozess- und Ergebnisqualität. Letztlich ist allerdings von einem positiven Zusammenhang zwischen Struktur- und Ergebnisqualität auszugehen. Neben den beschriebenen Beziehungen zwischen Struktur-/Prozessqualität und Prozess-/Ergebnisqualität existieren Rückwirkungen der Ergebnis- auf die Prozessqualität, der

Ergebnis- auf die Strukturqualität wie auch der Prozess- auf Strukturqualität. Diese *Rückkopplungen* sind nicht etwa als Störvariablen, sondern vielmehr als korrigierende Qualitätsentwicklung zu interpretieren“ (Macsenae 2006: 50).

Vor diesem Hintergrund ist der Vorschlag von Köhler (2017b: 95), auf die Prozess- und Ergebnisevaluation zugunsten einer auf den von ihm vorgeschlagenen Qualitätsstandards¹² basierenden Evaluation der strukturellen Integrität¹³ zu verzichten, *cum grano salis* zu nehmen. Denn so plausibel die Annahme eines positiven Zusammenhangs zwischen der Struktur- und Ergebnisqualität auch ist, würde eine reine Beurteilung der strukturellen Qualitätsstandards eines Programms bzw. Projekts den Evaluationsprozess seines eigentlichen Sinns und Zwecks berauben, der seit eh und je in der Bewertung der Prozess- und Ergebnisqualität besteht (Rogers 2000; Madaus/Kellaghan 2000). Zugleich spielt die Programmqualität hinsichtlich der angestrebten Ergebnisse und Umsetzungsprozesse eine unbestritten wichtige Rolle.¹⁴ Streng genommen resultierte der obige Substitutionsvorschlag aus einem methodischen Irrtum, denn Latessa (2013: 67 f.) kritisierte zwar die weit verbreitete indirekte Messung der Programmqualität über die Bewertung von Prozessergebnissen, betonte jedoch den Mehrwert einer Kombination aus der Konzeptbewertung sowie Output- und Programmintegritätsmessung.¹⁵ Daher bedarf es nach wie vor eines multimodalen Ansatzes, der die Konzept-, Struktur-, Prozess- und Wirkungsqualität von Programmen und Projekten reflektiert.

¹² Auch wenn jegliche Verweise im Handbuch „Strukturelle Qualitätsstandards“ von Köhler (2016) fehlen (Stand: Juli 2020), basieren seine Indikatoren zum überwiegenden Teil auf dem Leitfaden „The Evidence Based Correctional Program Checklist (CPC)“ von Edward J. Latessa und Mitarbeitern (Latessa 2013). In der Monographie „Understanding Deradicalization“ sucht man ebenfalls vergebens nach der Quelle der „Deradicalization program integrity checklist (DPIC)“, die neben den einzelnen Items die CPC-Heuristik ohne jegliche empirische Korrelationsnachweise übernimmt (Köhler 2017a: 296 ff.). Erst in einer späteren Veröffentlichung erfolgte eine transparente Benennung der Quelle der sehr nah am Original gehaltenen strukturellen Qualitätsstandards: „A first step to achieving that gold standard would be to assess the programme’s integrity through a checklist, as suggested by Koehler (2016, 2017), based on the Correctional Programme Checklist (Latessa, 2013)“ (Köhler 2017b: 94).

¹³ „Structural integrity evaluation has so far not been suggested to be used with deradicalisation programmes other than by the author (Köhler, 2017)“ (Köhler 2017b: 96). Es sei an dieser Stelle angemerkt, dass der Vorschlag, Instrumente wie CPAI 2000 zur Bewertung der Qualität von Deradikalisierungsprogrammen zu nutzen, von Mullins (2010: 178) kam. Die CPC-Liste von Edward J. Latessa und Mitarbeitern stellt eine auf weiterführenden Korrelationsforschungen basierende „Verdichtung“ der CPAI-Indikatoren dar. Dabei verwies Mullins (2010: 178) im Abschnitt „Program design, implementation and evaluation“ seines Aufsatzes „Rehabilitation of Islamist terrorists: Lessons from Criminology“ auf frühere kriminologische Forschungen von Paul Gendreau, Claire Groggin und Paula Smith: „Gendreau, Groggin and Smith take an equally detailed approach in offering advice on the implementation of programs in the ‘real world’ and identify four key areas to attend to, relating to organizational factors, the program itself, the agent of change (who is responsible for running the program), and staffing activities. They also emphasize the importance of program evaluation and suggest the use of the CPAI for this purpose. The latest version of this tool, the CPAI-2000, includes 131 items and assesses eight different domains: 1) organizational culture; 2) program implementation/maintenance; 3) management/staff characteristics; 4) client risk–need practices; 5) program characteristics; 6) core correctional practices; 7) interagency communication; and 8) evaluation. CPAI scores have been found to relate to reductions in recidivism, thus adding validity both to this assessment, and by proxy to the ‘what works’ literature on which it is based.“

¹⁴ So berichtete Latessa (2004: 553): „Not only can the ‘failures’ cancel out the ‘successes’, but also by failing to measure program characteristics and fidelity we are often unable to explain some of the programmatic reasons why differences in outcome measures may occur. For example, in a recent study of similar types of residential correctional programs in Ohio, we found a wide range of effects based on the quality of the program“.

¹⁵ Auch der Vordenker der Programmqualitätsforschung, Quay (1977), nennt die Beurteilung der Programmintegrität das „dritte Gesicht der Evaluation“.

Fazit

Es mangelt in der Forschung nicht an fruchtbaren Ansätzen, Modellen und Verfahren der Evaluation in der Kriminal- und Extremismusprävention (Bubenitschek et al. 2014, Gielen 2019, Glaser/Schuster 2007, Hohnstein/Greuel 2015, Widmer et al. 2007). Zugleich entsteht der Eindruck, dass die methodische und institutionelle Ausdifferenzierung des Feldes einen „Kampf“ der Paradigmen¹⁶ in Deutschland befeuert, der einen Dialog zwischen verschiedenen „Evaluationsschulen“ sowie zwischen der Meta- und Praxisforschung teils erschwert. Es fällt zudem auf, dass die Diskurse rund um die Evaluationsmethoden und (Gold-)Standards der Extremismusprävention jene Kontroversen wiederholen, die bereits in der angelsächsischen Soziologie der 1970/80er Jahre einen prominenten Platz einnahmen. Vor diesem Hintergrund hilft ein Blick auf die Historie der Evaluationsforschung, die eine Reihe von brauchbaren Ansätzen bietet. Es sei angemerkt, dass auch in der Evaluation von Ansätzen wirkungsorientierter Sozialer Arbeit innovative Verfahren entwickelt werden konnten, die integriert betrachtet werden sollten (Heil et al. 2001, Merchel 2015, Projekt eXe 2006). Überdies wäre es sicherlich von großem Vorteil, wenn die relevanten Akteure auf dem Gebiet der Praxisforschung, wirkungsorientierten Evaluation und Meta-Forschung einen fachlichen Austausch intensivieren würden, um den Blick für die Handlungslogiken der Anderen zu schärfen und gegenstandsadäquate Bewertungsverfahren (weiter-)entwickeln zu können.¹⁷ Denn neben „schlechter Planung“ und „schlechter Umsetzung“ stellt „schlechte Wissenschaft“ eine weitere Unsicherheitsquelle im Hinblick auf Präventionsprogramme und soziale Interventionen dar.

Das gilt auch für die Frage der politischen und sozialen Intentionalität von Programmen sowie die grundlegenden Verhältnisse zwischen Programmtheorie und Praxis bzw. Programmwirklichkeit. Verschiedene theoretische Modelle, Forschungsgegenstände, Anforderungen und Evaluationsziele können in eine Überspannung in der Wirklichkeit geraten, die oft nur schwer zu korrigieren ist und zu einer „Fehlpraxis“ führen können. Um dies zu vermeiden, ist es notwendig, den Fragen der Evaluationszusammenhänge und -logiken weiter nachzugehen.¹⁸

¹⁶ Siehe bspw. die Diskussion in der Serie „Evaluation“ der Bundeszentrale für politische Bildung.

¹⁷ Vgl. das vom BMI geförderte Projekt PrEval (Evaluationsdesigns für Präventionsmaßnahmen – Multimethodische Ansätze zur Wirkungsermittlung und Qualitätssicherung in der Extremismusprävention sowie den Schnittstellen zur Gewaltprävention und politischen Bildung).

¹⁸ Mit dem Projekt PrEval (Evaluationsdesigns für Präventionsmaßnahmen – Multimethodische Ansätze zur Wirkungsermittlung und Qualitätssicherung in der Extremismusprävention sowie den Schnittstellen zur Gewaltprävention und politischen Bildung) fördert das BMI ein erstes Format dieser Art.

Literatur

- Airasian, Peter W. (1983): Societal Experimentation. In: George F. Madaus, Michael S. Scriven und Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht: Springer Netherlands, S. 163–175.
- Armborst, Andreas (2018): Einführung: Merkmale und Abläufe evidenzbasierter Kriminalprävention. In: Maria Walsh, Benjamin Pniewski, Marcus Kober und Andreas Armborst (Hg.): *Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis*. Wiesbaden: Springer, S. 3–19.
- Armborst, Andreas; Biene, Janusz; Coester, Marc; Greuel, Frank; Milbradt, Björn; Nehlsen, Inga (2018): *Evaluation in der Radikalisierungsprävention. Ansätze und Kontroversen*. Frankfurt am Main.
- Baruch, Ben; Ling, Tom; Warnes, Rich; Hofman, Joanna (2018): Evaluation in an emerging field: Developing a measurement framework for the field of counter-violent-extremism. In: *Evaluation* 24 (4), S. 475–495. DOI: 10.1177/1356389018803218.
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation - Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Karin Haubrich (Hg.): *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis*. München, S. 25–48.
- Beywl, Wolfgang; P, Susanne (2012): Nutzungsfokussierte Evaluation am Beispiel eines multizentrischen Programms. In: Rainer Strobl, Olaf Lobermeier und Wilhelm Heitmeyer (Hg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden, S. 101–125.
- Björgero, Tore (2016): *Preventing Crime. A Holistic Approach*. New York.
- Blamey, Avril; Mackenzie, Mhairi (2007): Theories of Change and Realistic Evaluation. In: *Evaluation* 13 (4), S. 439–455. DOI: 10.1177/1356389007082129.
- Brandt, Tasso (2009): *Evaluation in Deutschland. Professionalisierungsstand und -perspektiven*. Münster.
- Breuer, Erica; Lee, Lucy; Mary De Silva, Crick Lund (2016): Using theory of change to design and evaluate public health interventions: a systematic review. In: *Implementation Science* (2016) 11:63. DOI 10.1186/s13012-016-0422-6.
- Bubenitschek, Günther; Greulich, Reiner; Wegel, Melanie (2014): *Kriminalprävention in der Praxis*. Heidelberg.
- Chen, Huey T. (2012): Theory-driven evaluation: Conceptual framework, application and advancement. In: Rainer Strobl, Olaf Lobermeier und Wilhelm Heitmeyer (Hg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden, S. 17–40.
- Coester, Marc (2018): Das Düsseldorfer Gutachten und die Folgen. In: Maria Walsh, Benjamin Pniewski, Marcus Kober und Andreas Armborst (Hg.): *Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis*. Wiesbaden, S. 37–57.
- Connell, James P.; Kubisch, Anne C. (1998): *Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects, and Problems*. Online: <https://www.dmeforpeace.org/sites/default/files/080713%20Applying+Theory+of+Change+Approach.pdf> (24. Juni 2020).

- Cronbach, Lee J. (1982). *Designing evaluations of educational and social programs*. San Francisco.
- Cronbach, Lee J. (1983): *Ninety-five Theses for Reforming Program Evaluation*. In: George F. Madaus, Michael S. Scriven und Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 405–412.
- Davolio, Miryam Eser/Drilling, Matthias (2007): *Vergleichende Evaluation von Interventionen gegen Rechtsextremismus: Perspektiven professionalisierter wissensgenerierender Evaluationen*. In: Michaela Glaser; Silke Schuster (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle, S. 53–68.
- Dollinger, Bernd (2018): *Die Konstruktion von Evidenz in der Präventionsarbeit*. In: Maria Walsh, Benjamin Pniewski, Marcus Kober und Andreas Armbrorst (Hg.): *Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis*. Wiesbaden, S. 187–203.
- Duriez, Stephanie A.; Sullivan, Carrie; Latessa, Edward J.; Lovins, Lori Brusman (2018): *The Evolution of Correctional Program Assessment in the Age of Evidence-Based Practices*, *Corrections*, 3 (2), S. 119–136, DOI: [10.1080/23774657.2017.1343104](https://doi.org/10.1080/23774657.2017.1343104).
- 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: Rainer Strobl, Olaf Lobermeier und Wilhelm Heitmeyer (Hg.): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden, S. 69–97.
- Farrington, David P.; Ttofi, Maria M.; Lösel, Friedrich A. (2016): *Developmental and Social Prevention*. In: David Weisburd, David P. Farrington, Charlotte Gill (Hg.): *What Works in Crime Prevention and Rehabilitation*. New York, S. 15–75.
- Fink, Naureen Chowdhury; Romaniuk, Peter; Barakat, Rafia (2013): *Evaluating Countering Violent Extremism Programming. Practice and Progress*. New York.
- Gansewig, Antje (2018): *Prävention von politischem Extremismus in Deutschland*. In: Maria Walsh, Benjamin Pniewski, Marcus Kober und Andreas Armbrorst (Hg.): *Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis*. Wiesbaden, S. 465–488.
- Giel, Susanne (2013): *Theoriebasierte Evaluation: Konzepte und methodische Umsetzungen*, Münster.
- Gielen, Amy-Jane (2018): *Exit programmes for female jihadists: A proposal for conducting realistic evaluation of the Dutch approach*. In: *International Sociology*, 33(4), 454–472. DOI: [10.1177/0268580918775586](https://doi.org/10.1177/0268580918775586).
- Gielen, Amy-Jane (2019): *Countering Violent Extremism: A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How?* In: *Terrorism and Political Violence* 31 (6), S. 1149–1167. DOI: 10.1080/09546553.2017.1313736.
- Gill, Charlotte (2016): *Community Interventions*. In: David Weisburd, David P. Farrington, Charlotte Gill (Hg.): *What Works in Crime Prevention and Rehabilitation*. New York, S. 77–109.
- Glaser, Michaela; Schuster, Silke (Hg.) (2007): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle.
- Glock, Birgit; Lüter, Albrecht; Schroer-Hippel, Miriam (2018): *Jugendgewaltprävention und Wirkungsorientierung: Monitoring, Evaluation und Transfer durch die Berliner Arbeitsstelle Jugendgewaltprävention*. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst

(Hg.): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden, S. 167–184.

Groeger-Roth, Frederick; Marks, Erich; Meyer, Anja (2018): Erfahrungen des Landespräventionsrates Niedersachsen (LPR) bei der Umsetzung einer qualitätsorientierten und evidenzbasierten Präventionspraxis. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst (Hg.): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden, S. 145–165.

Guerette, Rob (2009): Pull, Push, and Expansion of Situational Crime Prevention Evaluation: An Appraisal of Thirty-Seven Years of Research. In: Johannes Knutsson, Nick Tilley (Hg.): Evaluating Crime Reduction Initiatives. New York, S. 29–58.

Heil, Karolus; Heiner, Maja; Feldmann, Ursula (Hg.) (2001): Evaluation sozialer Arbeit, Frankfurt a.M. 2001.

Helmus, Todd C. et al. (2017): RAND Program Evaluation Toolkit for Countering Violent Extremism. Santa Monica.

Hemert, Dianne van; Helma van den Berg; Tony van Vliet; Maaïke Roelofs; Mirjam Huis in 't Veld (2014): Synthesis report on the state-of-the-art in evaluating the effectiveness of counter-violent extremism interventions Brussels.

Hohnstein, Sally; Greuel, Frank; Glaser, Michaela (2015): Einstiege Verhindern, Ausstiege begleiten. Pädagogische Ansätze und Erfahrungen im Handlungsfeld Rechtsextremismus. Halle.

Horgan, John; Braddock, Kurt (2010): Rehabilitating the Terrorists?: Challenges in Assessing the Effectiveness of De-radicalization Programs. In: *Terrorism and Political Violence*, 22(2), 267–291. DOI: [10.1080/09546551003594748](https://doi.org/10.1080/09546551003594748).

House, Ernest R. (1983): Assumptions Underlying Evaluation Models. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 45–64.

Jackson, Suzanne F.; Kolla, Gillian (2012): A New Realistic Evaluation Analysis Method. In: *American Journal of Evaluation* 33 (3), S. 339–349. DOI: [10.1177/1098214012440030](https://doi.org/10.1177/1098214012440030).

Kellaghan, Thomas; Madaus, George F. (2002): Outcome Evaluation. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 97–112.

Knutsson, Johannes (2009): *Evaluating crime reduction Initiatives*. New York.

Kober, Marcus; Frevel, Bernhard; van den Brink, Henning; Wurtzbacher, Jens (2018): Evidenz in der Kommunalen Kriminalprävention – Zur Wirksamkeitsanalyse von Kooperationsstrukturen. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst (Hg.): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden, S. 729–741.

Köhler, Daniel (2016): Strukturelle Qualitätsstandards in der Interventions- und Präventionsarbeit gegen gewaltberbeiteten Extremismus. Hg. v. Ministerium für Inneres, Digitalisierung und Migration Baden-Württemberg. Stuttgart.

Koehler, Daniel (2017a): *Understanding Deradicalization. Methods, tools and programs for countering violent extremism*. New York.

Koehler, Daniel (2017b): Preventing Violent Radicalisation: Program Design and Evaluation. Online: https://www.cidob.org/en/articulos/monografias/resilient_cities/preventing_violent_radicalisation_programme_design_and_evaluation (25. Juni 2020).

Korn, Judy (2016): European CVE Strategies from a Practitioner's Perspective. In: *The ANNALS of the American Academy of Political and Social Science* 668 (1), S. 180–197. DOI: 10.1177/0002716216671888.

Kromrey, Helmut (2007): Begleitforschung und Evaluation – fast das Gleiche, und doch etwas Anderes! In: Michaela Glaser, Silke Schuster (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle, S. 113–135.

La Chaux, Marlen de; Kober, Marcus; Nabo, Mitra Moussa (2018): Überlegungen zur Schaffung einer Evidenzgrundlage für die Präventionsarbeit im Bereich islamistischer Extremismus. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst (Hg.): *Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis*. Wiesbaden, S. 489–509.

Latessa, Edward J. (2013): Evaluating Correctional Programs. Online: https://www.unafei.or.jp/english/pdf/RS_No88/No88_11VE_Latessa_Evaluating.pdf (11. Februar 2018; nicht mehr abrufbar, liegt dem Verfasser vor).

Lee, Barbara (2004): Theories of Evaluation. In: Reinhard Stockmann (Hg.): *Evaluationsforschung*. Wiesbaden, S. 127–164.

Leeuw, Frans L. (2000): Evaluation in Europe. In: Reinhard Stockmann (Hg.): *Evaluationsforschung*. Wiesbaden, 57–76.

Lüders, Christian; Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe: Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Karin Haubrich (Hg.): *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis*. München, S. 5–24.

Lüders, Christian/Haubrich, Karin (2007): Evaluation in wenig formalisierten pädagogischen Settings. In: Michaela Glaser, Silke Schuster (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle, S. 136–149.

Lum, Cynthia; Kennedy, Leslie W. (2012): *Evidence-Based Counterterrorism Policy*. New York.

Lützinger, Saskia; Gruber, Florian (2017): *Extremismusprävention in Deutschland – Herausforderungen und Optimierungspotential*. Modulabschlussbericht. Wiesbaden.

Macsenaere, Michael (2006): Wirkungen der Kinder- und Jugendhilfe sind messbar! Methoden, Ergebnisse und Empfehlungen der Jugendhilfe-Effekte-Studie (JES) und weiterer darauf beruhender wirkungsorientierter Evaluationen. In: Karin Haubrich (Hg.): *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis*. München, S. 49–80.

Madaus, George F.; Kellaghan, Thomas (2000): Models, Metaphors, and Definitions in Evaluation. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 19–31.

Madaus, George F.; Stufflebeam, Daniel L. (2000): Program Evaluation: A Historical Overview. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 3–18.

- Madaus, George F.; Stufflebeam, Daniel; Scriven, Michael S. (1983): Program Evaluation. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): Evaluation Models. Dordrecht, S. 3–22.
- Merchel, Joachim (2019): Evaluation in der Sozialen Arbeit. Wien.
- Mertens, Donna M. (2004): Institutionalizing Evaluation in the United States of America. In: Reinhard Stockmann (Hg.): Evaluationsforschung. Wiesbaden, S. 41–56.
- Molenkamp, Merel et al. (2018): Guideline Evaluation of PCVE Programmes and Interventions. Online: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-we-do/networks/radicalisation_awareness_network/ranpapers/docs/ms_workshops_guidelines_evaluation_of_pcve_programmes_and_interventions_july_2018_en.pdf (25. Juni 2020).
- Möller, Kurt; Neuscheler, Florian (2018): Abschlussbericht zur Evaluation der Beratungsstelle Hessen – Religiöse Toleranz statt Extremismus. Esslingen. Online: <https://violence-prevention-network.de/wp-content/uploads/2019/02/Abschlussbericht-Evaluation-Beratungsstelle-Hessen.pdf> (25. Juni 2020).
- Mullins, Sam (2010): Rehabilitation of Islamist terrorists: Lessons from criminology, Dynamics of Asymmetric Conflict, 3 (3), 162–193. DOI: [10.1080/17467586.2010.528438](https://doi.org/10.1080/17467586.2010.528438).
- Müller, Christoph Emanuel (2017): Kausale Wirkungsevaluation zwischen methodischem Anspruch und empirischer Praxis. In: Reinhard Stockmann, Wolfgang Meyer (Hg.): Die Zukunft der Evaluation. Münster, S. 205–222.
- Otto, Hans-Uwe; Polutta, Andreas; Ziegler, Holger (Hg.) (2010): What works – welches Wissen braucht die soziale Arbeit? Zum Konzept evidenzbasierter Praxis. Opladen.
- Palloks, Kerstin (2007): „Große Erwartungen“ – zur Wirkungsfrage bei der Evaluation von Modellprogrammen. In: Michaela Glaser, Silke Schuster (Hg.): Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen. Halle, S. 14–31.
- Patton, Michael Q. (2003): Utilization-Focused Evaluation. In: Kellaghan, T., Stufflebeam, D.L. (Hg.): International Handbook of Educational Evaluation, Wiesbaden, S. 223-244.
- Pawson, Ray/Tilley, Nick (2004): Realist Evaluation. Online: http://www.communitymatters.com.au/RE_chapter.pdf (25. Juni 2020).
- Pniewski, Benjamin; Walsh, Maria (2018): Schwierigkeiten bei der Implementierung von Evaluationsforschung in der Praxis. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst (Hg.): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden, S. 253–268.
- Projekt eXe (Hg.) (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis, Augsburg.
- Quay, Herbert C. (1977): The three faces of evaluation: What can be expected to work. Criminal Justice and Behavior, 4 (4), 341–354. DOI: [10.1177/009385487700400402](https://doi.org/10.1177/009385487700400402).
- Rieker, Peter (2009): Rechtsextremismus: Prävention und Intervention. Ein Überblick über Ansätze, Befunde und Entwicklungsbedarf. München.

- Rogers, Patricia J. (2000): Program Theory: Not Whether Programs Work but How They Work. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 209–232.
- Schuster, Silke (2007): Improvisation, Partizipation und die Frage der Wirkungen – Eine Untersuchung zu Evaluationen der pädagogischen Arbeit gegen Rechtsextremismus. In: Michaela Glaser, Silke Schuster (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle, S. 169–181.
- Scriven, Michael (1983): Evaluation Ideologies. In: George F. Madaus, Michael S. Scriven und Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 229–260.
- Schuurman, Bart; Bakker, Edwin (2015): Reintegrating jihadist extremists: evaluating a Dutch initiative, 2013–2014. In: *Behavioral Sciences of Terrorism and Political Aggression*, 8 (1), S. 66–85, DOI: [10.1080/19434472.2015.1100648](https://doi.org/10.1080/19434472.2015.1100648).
- Sherman, Lawrence W. (Hg.) (2006): *Evidence-based crime prevention*. London.
- Stake, Robert E. (2000): Responsive Evaluation. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 63–68.
- Steinmetz, Andrés (1983): The Discrepancy Evaluation Model. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 79–99.
- Stockmann, Reinhard (2004): *Evaluation in Deutschland*. In: Reinhard Stockmann (Hg.): *Evaluationsforschung*. Wiesbaden, S. 11–40.
- Stockmann, Reinhard (Hg.) (2007): *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster.
- Stockmann, Reinhard; Meyer, Wolfgang (2017): *Die Zukunft der Evaluation. Trends, Herausforderungen, Perspektiven*, Münster.
- Strobl, Rainer; Lobermeier, Olaf; Heitmeyer, Wilhelm (Hg.) (2012): *Evaluation von Programmen und Projekten für eine demokratische Kultur*. Wiesbaden.
- Stufflebeam, Daniel L. (1983): The CIPP Model for Program Evaluation. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 117–141.
- Stufflebeam, Daniel L. (2000): Foundational Models for 21st Century Program Evaluation. In: Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan (Hg.): *Evaluation Models*, Bd. 49. Dordrecht, S. 33–83.
- Stufflebeam, Daniel L.; Madaus, George F. (1983): The Standards for Evaluation of Educational Programs, Projects, and Materials. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 395–404.
- Stufflebeam, Daniel L.; Madaus, George F.; Kellaghan, Thomas (Hg.) (2002): *Evaluation Models*. Dordrecht.
- Stufflebeam, Daniel L.; Webster, William J. (1983): An Analysis of Alternative Approaches to Evaluation. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): *Evaluation Models*. Dordrecht, S. 23–43.
- Stufflebeam, Daniel L. (2003): The CIPP Model for Evaluation. In: Kellaghan, T., Stufflebeam, D.L. (Hg.): *International Handbook of Educational Evaluation*. Wiesbaden, S. 31–62.

Tilley, Nick (2002): Evaluation for crime prevention. New York.

Tyler, Ralph W. (1983): A Rationale for Program Evaluation. In: George F. Madaus, Michael S. Scriven, Daniel L. Stufflebeam (Hg.): Evaluation Models. Dordrecht, S. 67–78.

Uhlmann, Milena (2017): Evaluation der Beratungsstelle "Radikalisierung". Abschlussbericht. Nürnberg.

Ullrich, Simone et al. (2019): Evaluationskriterien für die Islamismusprävention. Bonn.

Vedung, Evert (2010): Four Waves of Evaluation Diffusion. In: Evaluation, 16 (3), 263–277. DOI: [10.1177/1356389010372452](https://doi.org/10.1177/1356389010372452).

Walsh, Maria; Pniewski, Benjamin; Kober, Marcus; Armbrorst, Andreas (Hg.) (2018): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden.

Weisburd, David; Farrington, David P.; Gill, Charlotte (2016): Introduction: What Works in Crime Prevention? In: David Weisburd, David P. Farrington, Charlotte Gill (Hg.): What Works in Crime Prevention and Rehabilitation. New York, S. 1–13.

Weiss, Carol H. (1995): Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In: The Aspen Institute (Hg.): New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts, The Aspen Institute, 65–92.

Widmer, Thomas (2012): Wirkungsevaluation zu Maßnahmen der Demokratieförderung. In: Rainer Strobl, Olaf Lobermeier, Wilhelm Heitmeyer (Hg.): Evaluation von Programmen und Projekten für eine demokratische Kultur. Wiesbaden, S. 41–68.

Widmer, Thomas; Blaser, Cornelia; Falk, Chantal (2007): Evaluating Measures Taken Against Right-Wing Extremism. In: *Evaluation* 13 (2), S. 221–239. DOI: 10.1177/1356389007075225.

Yngborn, Annalena; Hoops, Sabrina (2018): Das Logische Modell als Instrument der Evaluation in der Kriminalitätsprävention im Kindes- und Jugendalter. In: Maria Walsh, Benjamin Pniewski, Marcus Kober, Andreas Armbrorst (Hg.): Evidenzorientierte Kriminalprävention in Deutschland. Ein Leitfaden für Politik und Praxis. Wiesbaden, S. 349–368.

Der Autor ist wissenschaftlicher Mitarbeiter im [Forschungs- und Informationsdienst Extremismus und Militanz \(FIDEM\)](#) der ZDK Gesellschaft Demokratische Kultur gGmbH.

Dieser Beitrag stellt in stark komprimierter Form die im Rahmen eines Teilprojekts des vom Bundesministerium des Innern, für Bau und Heimat (BMI) geförderten Forschungs- und Informationsdienstes Extremismus und Militanz (FIDEM) gewonnenen Erkenntnisse über Ansätze und Methoden der Evaluation sowie Implikationen für die Praxis der Radikalisierungsprävention dar.

Gefördert durch:



Bundesamt
für Migration
und Flüchtlinge