



# KAPITEL 4

## Zur Evaluation von Praxisansätzen der Extremismusprävention

Kurt Möller, Jan Buschbom, Thomas Pfeiffer

### Prof. Dr. Kurt Möller

Erziehungswissenschaftler und Professor für Soziale Arbeit an der Hochschule Esslingen. Prof. Möller ist Evaluator diverser Projekte und Mitglied zahlreicher Netzwerke in den Bereichen Demokratieentwicklung, Jugend und Rechtsextremismus. Ferner ist er Mitglied der Sachverständigenkommission, die den 16. Kinder- und Jugendbericht der Bundesregierung zum Thema „Förderung demokratischer Bildung im Kindes- und Jugendalter“ erstellt.

### Jan Buschbom

Historiker und Gründungsmitglied von Violence Prevention Network e. V. in dessen Vorstand er bis Ende 2017 Mitglied war. Er ist zudem Gründer und Vorstand des Vereins juvenes e. V.

### Dr. Thomas Pfeiffer

Studierte Journalistik und Politikwissenschaft. Seit 2002 ist er als wissenschaftlicher Referent in der Abteilung Verfassungsschutz des Innenministeriums Nordrhein-Westfalen tätig.

Sie planen, unterstützen oder fördern ein Angebot zur Extremismusprävention? Und Sie fragen sich, wie Sie herausfinden können, was es bringt und wie gut es funktioniert? Dann sollten Sie über eine Evaluation Ihres Vorhabens nachdenken. In diesem Kapitel erfahren Sie, was Sie dazu wissen müssen und bedenken sollten.

## Bedarfe und Zwecke von Evaluation

### Evaluation? Wozu?

Wo sind Bedarfe an Evaluation auszumachen und was haben Auftraggeber, Präventionsträger, haupt- und ehrenamtlich Mitarbeitende, potenzielle Kooperationspartner und Adressaten und Adressatinnen sowie ggf. deren Angehörige und befreundete Personen davon?

Bedarfe an Evaluation (begrifflich abgeleitet aus dem Lateinischen: *valere* = stark sein, etwas wert sein, Geltung beanspruchen können) liegen grundsätzlich immer dann vor, wenn Interesse an der Aus- und Bewertung der Arbeit von Organisationen, der Funktionen von Produkt(teil)en und insbesondere auch von bestimmten Handlungspraxen und den damit zusammenhängenden Sachverhalten besteht. Sie sind insbesondere dann stark vorhanden, wenn Evaluationsgegenstände wie die genannten innovativen Charakter haben, (oft auch deshalb) bislang nur unzureichend untersucht wurden und optimiert werden sollen. Zudem werden sie dort aufgeworfen, wo Entscheidungen über ihre Ein- bzw. Weiterführung und ihre Verbreitung bzw. ihren Ausbau anstehen und Verantwortliche sich dabei in der Pflicht sehen, Rechenschaft darüber abzulegen. Evaluation kann allerdings auch dann sinnvoll sein, wenn es um langjährig erprobte, vermeintlich oder tatsächlich bewährte Ansätze geht. Je länger eine Maßnahme zum eigenen Repertoire gehört, desto weniger gelingt es möglicherweise, selbst Verbesserungspotenziale zu entdecken. Auch das Phänomen, dem die Maßnahme entgegenwirken soll, könnte sich zwischenzeitlich derart verändert haben, dass selbst die beste Prävention angepasst werden muss. Bei der Evaluation geht es nicht darum, ein Angebot einmalig zu testen und für mehr oder weniger gut zu befinden – sie ist eine Daueraufgabe.

Auf den Gebieten der Extremismusprävention sind Bedarfe an Evaluation zumeist nicht nur vereinzelt und isoliert voneinander, sondern gleich in vielfältiger Kombination gegeben. Dies gilt umso mehr, als extremistische Tendenzen unterschiedlicher Herkunft und Kontur sich weltweit, aber auch national in Deutschland bedrohlich ausbreiten, die Prinzipien und Strukturen demokratischen Zusammenlebens angreifen, ihre Erscheinungsweisen sowie ihre Strategien zunehmend rasch an neue Gegebenheiten anpassen und dabei in hohem Maße die Aufmerksamkeit von Öffentlichkeit, Medien und Politik auf sich ziehen. Sie bilden damit unübersehbare Herausforderungen, denen im Interesse des Erhalts von Demokratie, wechselseitigem Respekt und Gewaltferne dringlich mit erfolgversprechenden, aktualitätsbezogenen und daher nahezu zwangsläufig auch kontinuierlich neu auszurichtenden Vorgehensweisen zu begegnen ist. Daher propagiert die Bundesregierung auch die Evaluation von extremismuspräventiven Angeboten, um insbesondere deren „Qualitätssicherung, Nachhaltigkeit und Weiterentwicklung“ zu gewährleisten.<sup>1</sup> Sie will daher „Evaluation und Qualitätssicherung weiter voranbringen“ und „ihre Forschungsförderung auf diesen Gebieten weiter verstärken“.<sup>2</sup> Und das Bundesinnenministerium ergänzt im „Nationalen Präventionsprogramm gegen islamistischen Extremismus“: „Zudem wird die Bundesregierung die praxisorientierte

1 Die Bundesregierung 2016, 26.

2 Bundesministerium für Familie, Senioren, Frauen und Jugend 2017, 50.

Forschung zur Weiterentwicklung der präventiv-pädagogischen Arbeit sowie zu Ansätzen der Ausstieges- [sic!] und Distanzierungsarbeit fördern“.<sup>3</sup> Ein entsprechender Forschungsauftrag ist in der zweiten Julihälfte 2018 ergangen, bewilligte Projekte – nicht alle im engeren Sinne evaluativ ausgerichtet – sind im März 2019 gestartet.

Wenn also damit der Bedarf an Evaluationen, insbesondere für Programme, Projekte und Einzelmaßnahmen, die sich auf (relatives) Neuland begeben, geradezu auf der Hand liegt und als unbestritten gelten kann, so werden doch die Zwecke, die im Einzelnen mit Evaluationen verfolgt werden sollen, standort- und perspektiveabhängig meist sehr unterschiedlich gesehen:

## Evaluation – Wer hat welche Interessen?

Für politisch Verantwortliche, Auftraggeber und Finanziers steht meist im Vordergrund, entweder ex ante (also vor der Durchführung eines Programms, eines Projekts oder einer einzel-

nen Maßnahme) evaluativ abzuklären, ob das Angebot zum Bedarf passt, durchführbar ist und die angestrebte Wirkung zu erwarten ist. Oder es geht ihnen darum – dies faktisch weitaus häufiger –, ex post (etwa nach Ablauf einer Pilotphase) nachvollziehbare Entscheidungen über die weitere Unterstützung bzw. Förderung von Angeboten oder auch über deren Ausbau oder Beendigung zu treffen und dies legitimieren zu können. Die wissenschaftliche Evaluation kann also für den Träger auch ein strategisches Instrument sein: Er kann auf das scheinbar neutrale Urteil der Wissenschaft verweisen und damit möglicher Kritik den Wind aus den Segeln nehmen. Falls hier Fragen der Mittelgewährung und des Ressourceneinsatzes sowie deren Rechtfertigung im Mittelpunkt stehen, ist das vornehmliche Interesse dann darauf gerichtet, ob und wie weit welche Wirkungen in welchem Ausmaß zu erzielen sind. Entsprechend stark konzentriert sich die Aufmerksamkeit auf das (hier primär monetär-ökonomisch verstandene) Kosten-Nutzen-Verhältnis. Wie und warum Wirkungen zustande kommen (oder wieso ggf. auch nicht), erscheint demgegenüber ebenso zweitrangig wie eine Auseinandersetzung darüber, ob und wie (ökonomisch gefasste) Kosten und sozialer sowie politischer Nutzen überhaupt berechenbar sind.



**Ex ante:** Die „Ex ante Evaluation“ ist eine Evaluation, die bereits vor der Durchführung eines wie auch immer gearteten Angebots durchgeführt wird, beispielsweise um zu erheben, ob das Angebot zum Bedarf passt, es finanziell durchführbar ist und die angestrebte Wirkung zu erwarten ist.

**Ex post:** Die „Ex post Evaluation“ ist eine nachträgliche Evaluation, die nach der Durchführung eines wie auch immer gearteten Angebots oder nach Ablauf einer Projektphase, beispielsweise der Pilotphase, stattfindet. Dadurch können Entscheidungen über die weitere Unterstützung bzw. Förderung von Angeboten oder auch über deren Ausbau oder Einstampfen getroffen und legitimiert werden.

Aus der Perspektive von Angebotsdurchführenden, also von beim Angebotsträger beschäftigten Hauptamtlichen, Honorarkräften und Ehrenamtlichen, stellen sich Evaluationsanliegen im Allgemeinen anders dar, besonders dann, wenn diese über Selbstevaluationsaufgaben hinausgehen und eine externe Evaluation beauftragt wird. Auf der einen Seite stellt sich dann bei ihnen nicht selten eine Reihe von Befürchtungen ein: Kommen jetzt Kontrolleure von außen, die uns in die Karten schauen, unsere Arbeit vermessen und unsere institutionelle und persönliche Leistung bewerten? Sind dadurch nicht letztendlich unsere Stellen zumindest teilweise gefährdet oder könnte nicht zumindest die Außenwirkung unserer Einrichtung und darüber hinaus vielleicht auch die unseres Trägers negativ beeinflusst

werden? Ist es nicht zu erwarten, dass die evaluierenden Personen abgehobene Wissenschaftlerinnen und Wissenschaftler sind, die von den Realitäten der Praxis kaum Kenntnis haben, und birgt nicht schon deshalb eine Bewertung durch sie erhebliche Verzerrungen und Lücken und ihr Empfehlungskatalog für die Ausrichtung der künftigen Arbeit manch Unrealistisches? Ist der Aufwand, der im Zusammenhang mit der Evaluation verlangt wird – Zusammenstellung von Daten, Offenlegen von Dokumenten, Bereitstehen für Interviews, Vermittlung von Angebotsnutzerinnen und -nutzern, organisatorische Absprachen etc. –, überhaupt für uns leistbar und steht er wenigstens halbwegs in angemessener Relation zum vermutlichen Nutzen? Und gerade in der Arbeit mit Radikalisierten kommt die Problematisierung auf: Sind womöglich die Anonymität und/oder der Daten- und Vertrauensschutz unserer Klientinnen und Klienten, vielleicht sogar unsere eigene Sicherheit als Mitarbeitende durch zu viel Offenlegung gefährdet? Wie weit gefährdet der Eingriff der Evaluation den höchst fragilen Arbeitsprozess, der auf Aufbau und Erhalt vertrauensgeleiteter Arbeitsbeziehungen basiert? Ist das Vertrauen zwischen Praktiker bzw. Praktikerin und Klientin bzw. Klient erst einmal erschüttert, ist es schwierig, wenn nicht unmöglich, so hervorgerufenen Fehlentwicklungen erfolgreich zu begegnen. Will Evaluation Zugang zu unserer Klientel? Welche Wirkungen wird das auf sie haben? Und wie stelle ich ihr gegenüber Transparenz her, ohne dass das Vertrauensverhältnis gestört wird?

Solche Bedenken können sich mit einer generellen Evaluationsskepsis mischen, die davon ausgeht, dass Erfolg und Wirkung pädagogischen Handelns ganz und gar nicht messbar sind und das Vorhaben insofern wenig Mehrwert verspricht. Das gilt nicht zuletzt für die primäre bzw. universelle Prävention, zu der Maßnahmen der politischen Bildung zählen – hier sind die Grenzen tatsächlich besonders eng, in denen Erfolg im Sinne der Wirkung eines Projekts nachgewiesen werden kann. Klaus Ahlheim und Bardo Heger sprechen davon, dass allenfalls „Wirkungsspuren“ gesucht und festgestellt werden können.<sup>4</sup> Evaluation beschränkt sich in diesem Feld gelegentlich auf die Messung der Teilnehmendenzufriedenheit durch Feedbackbögen. Gerade bei unsicher Beschäftigten in der politischen Bildung kann das mit Ängsten verbunden sein, künftige Aufträge zu verlieren.

Auf der anderen Seite sind bei Mitarbeitenden zu evaluierender Programme, Projekte und Einzelmaßnahmen und ihren Anstellungsträgern erfahrungsgemäß aber durchaus auch Erwartungen nach positivem Nutzen vorhanden. Oftmals sind es aus Sicht der Mitarbeitenden vorrangig zwei Hoffnungen, die dann mit Evaluation verbunden werden: Wenn sie gut gemacht wird, wird sich zeigen, wie wichtig und zielführend die Arbeit ist, die wir in unserer Einrichtung bzw. in unserer Trägerschaft leisten und wie tragfähig das ist, was jeder Mitarbeitende persönlich leistet. Und: Wir werden durch eine genauere Untersuchung unserer Arbeitsprozesse, -konzepte und -strukturen herausfinden können, wo unsere Stärken und unsere Schwächen liegen. Wir werden dann in der Lage sein, Fehlentwicklungen vorzubeugen, Stabilisierungen unserer Arbeit an den richtigen Stellen anzugehen und Optimierungen als Weiterentwicklungen vorzunehmen – und das vielleicht sogar schon während der Laufzeit der Evaluation. Die Erwartung an die Evaluation wird dann sein, möglichst konkrete, unmittelbar umsetzbare Impulse für die eigene Praxis zu geben. In dieser durchaus widersprüchlichen Ausgangslage in der Sekundär- und insbesondere Tertiärprävention, geprägt von Befürchtungen und Hoffnungen, spiegelt sich das Verhältnis der Praktiker und Praktikerinnen zu ihren Klientinnen und Klienten im Verhältnis der evaluierenden Personen zu den Praktizierenden wider: Es kommt darauf an, ob es gelingt, die Methoden der Evaluation,

<sup>4</sup> Ahlheim/Heger 2006, 199; vgl. Hufer 2018 und die weiteren Beiträge im Themenheft „Wirkungen politischer Bildung“ des Journals für politische Bildung 3/2018.

nicht nur auf die Evaluationsbedarfe, sondern auch entlang der Charakteristika der eigentlichen Adressatengruppierung und der daraus abgeleiteten pädagogischen Arbeitsweisen und Methoden abzustimmen. Angesichts der Gefahr, die von Radikalisierten ausgeht, kommt es also in hohem Maße darauf an, die Balance zwischen den belastbaren Methoden der Evaluationsforschung und denen des untersuchten Arbeitsprozesses herzustellen.

Nach den Bedarfen, Zwecksetzungen und darüber hinausgehenden Erwartungen tatsächlicher und potenzieller Nutzerinnen und Nutzer von Angeboten, ihren alltagsweltlichen Bezugspersonen und nach den Bedarfen und Erwartungen von Kooperationspartnern wird im Allgemeinen im Zusammenhang mit Evaluationsvorhaben eher selten gefragt. Dabei können auch sie als wichtige Beteiligte im Evaluationszusammenhang betrachtet werden. Es ist erforderlich, Bezugspersonen und Kooperationspartner nicht nur als Außenstehende zu betrachten, sondern auch die Vorteile im Blick zu haben, die sich durch bestimmte Zwecke der Evaluation und die Anlage der Untersuchung auch für sie ergeben können.

## Zwischenfazit 1

Alles in allem bleibt hinsichtlich Bedarfen und Zwecken von Evaluation (nicht nur) in der Extremismusprävention zunächst festzuhalten: Evaluation im Sinne einer Auswertung und Überprüfung dessen, was inhaltlich, methodisch und strukturell an Arbeit geleistet wird, ist prinzipiell immer anzuraten, wenn produktive und auf Dauer tragfähige professionelle Leistungen erbracht werden sollen. Ihre Notwendigkeit drängt sich aber in erster Linie dort auf, wo Neuland betreten wird. Unabhängig davon, ob es sich um zivilgesellschaftliche Initiativen oder staatliche Stellen handelt, erscheint es dabei angezeigt, die Interessen der unterschiedlichen Stakeholder, etwa von Auftraggebern über Mitarbeitende bis hin zu Adressierten und ggf. ihren Bezugspersonen, aufzunehmen, die bei ihnen bestehenden Befürchtungen und Erwartungen zu berücksichtigen und diese Aspekte in ein Evaluationsdesign einfließen zu lassen, das ebenso gegenstandsangemessen wie nutzenorientiert angelegt ist.

# Grundlegende Konzepte und Ziele von Evaluation

## Theoretische und methodologische Ansätze und ihre Zielsetzungen im groben Überblick

Evaluation = Evaluation – diese Gleichung ist falsch. Von Evaluation existiert eine große Vielfalt an Verständnissen. Mehr noch: Diese Verständnisse und die mit ihnen verknüpften theoretischen Annahmen und Modelle werden in der Fachliteratur auch unterschiedlich systematisiert. Kategorisieren die einen etwa eher danach, ob mehr das Messen, das Beschreiben, das Bewerten oder das Aushandeln im Vordergrund des Evaluationsinteresses steht – so wie dies z. B. Guba/Lincoln<sup>5</sup> mit Blick auf die Geschichte von Evaluationsansätzen in ihrem Generationenmodell tun –, gruppieren andere die Ansätze entlang von Hauptpfaden der Entwicklung, die durch ihre jeweilige Fokussierung – stärker auf Methodisches, stärker auf Nutzenorientierung oder stärker auf Wertungsaspekte – unterschieden werden.<sup>6</sup> Wieder andere favorisieren ein Ordnungsmodell, das entlang von Schwerpunkt-

<sup>5</sup> Guba/Lincoln 1989.

<sup>6</sup> Vgl. Alkin/Christie 2004.

setzungen klassifiziert, die auf vier Funktionen von Evaluation Bezug nehmen: Erkenntnisgewinn, Kontrollerwartungen, Entwicklungsinteresse am zu evaluierenden Gegenstand und (nachträgliche) Legitimation.<sup>7</sup> Weitere Übersichtsversuche stellen Gesichtspunkte der praktischen Nutzensausrichtung von Evaluationen ins Zentrum der Kategorisierung. So nehmen etwa Fitzpatrick, Sanders und Worthen<sup>8</sup> eine Sortierung nach ziel-, management-, konsumenten-, experten- und partizipationsorientierten Ansätzen vor.

Methodologisch lassen sich zwei wissenschaftsphilosophisch und erkenntnistheoretisch gegensätzlich begründete Hauptrichtungen unterscheiden:

Eine klassische, lange Zeit in der wissenschaftlichen Forschung insgesamt, aber auch in der Evaluationsforschung vorherrschende Position propagiert das naturwissenschaftlich inspirierte Interesse von Wissenschaft, Sachverhalte im Sinne von naturgesetzlich-analogen Ursache-Wirkungszusammenhängen zu erklären.<sup>9</sup> Demgegenüber steht die Auffassung, insbesondere in den Sozial- und Geisteswissenschaften gehe es (zumindest eher) darum, die Bedeutungen des Handelns und die Sinnzusammenhänge, in denen gehandelt wird, zu verstehen.<sup>10</sup> Das dem erstgenannten Paradigma folgende deduktiv-nomologische Erklärungsschema<sup>11</sup> versucht, das zu Erklärende (etwa ein bestimmtes Ergebnis von Extremismusprävention) nach dem Muster eines Wenn-Dann-Verhältnisses aus dem Erklärenden (etwa einem bestimmten Verhalten von extremistischen Personen) „objektiv“ abzuleiten. Es betrachtet, damit dies gelingen kann, alle anderen möglichen Einflüsse als „Störvariablen“, die aufgrund dessen aus der Untersuchungssituation mittels statistischer Kontrolle auszuschalten sind.

Das Verstehens-Paradigma bestreitet dagegen, dass Realität (etwa die Wirkung eines extremismuspräventiven Angebots) in einer unbezweifelbaren Beschaffenheit gegeben und aus strikt neutraler Position erkennbar und bewertbar ist, hält sie ganz wesentlich für sozial konstruiert und versucht deshalb, die Perspektiven und darauf basierenden Interpretationen aller an der Realitätskonstruktion Beteiligten in Relation zueinander zu setzen. Die Komplexität und Dynamik von Realität ist demnach nur dann forschersich adäquat zu erfassen, wenn sie nicht experimentell-laborartig versimpelt wird. Aus dieser Sicht müssen zwingend anstelle bzw. neben statisch eingefangenen abstrakt-logischen und quantitativen Verhältnissen zwischen relevanten Faktoren die Qualitäten und Prozesse ihres Zusammenspiels (oder auch ihres Gegeneinanders) rekonstruiert werden.

Beschäftigen wir uns spezifischer mit Konzepten der Evaluation von Extremismusprävention, so findet sich im internationalen Raum eine Reihe von Vorschlägen, wie einschlägige Evaluationen angelegt sein können. Sie nehmen jeweils bestimmte Gesichtspunkte dieser theoretisch-inhaltlichen und methodologischen Dimensionen auf, lehnen andere hingegen ab oder stellen sie zumindest hinten.

Beachtenswert ist etwa die Multi Attribute Utility Technology (MAUT), die Horgan & Braddock<sup>12</sup> angelehnt an Edwards/Newmann<sup>13</sup> vorschlagen. Sie bezieht sich allerdings

7 Vgl. Stockmann/Meyer 2014.

8 Fitzpatrick/Sanders/Worthen 2011.

9 Vgl. z. B. Campbell 1969.

10 Vgl. z. B. schon Dilthey 1910.

11 Vgl. v. a. Hempel/Oppenheim 1948.

12 Horgan/Braddock 2010.

13 Edwards/Newmann 1982.

auf den (auch internationalen) Vergleich mehrerer Programme und deren Effektivität durch mathematisch gestützte Berechnungen. Der Ansatz ist allein schon insofern sehr voraussetzungsvoll, als unterschiedliche Zielsetzungen und Aufgabenstellungen der Programme angeglichen bzw. gewichtet werden müssen, um sie in der Gesamtbilanz ihrer Erreichung einigermaßen vergleichbar zu machen. Die Akzeptanz verschiedener Stakeholder in unterschiedlichen Praxisfeldern dafür zu erhalten, dürfte bislang ein wesentliches Hindernis gewesen sein, MAUT umzusetzen.

Ebenfalls eher abstrakte Vorschläge als evaluationspraktische Umsetzungserfahrungen und -verweise bieten die Überlegungen von Romaniuk & Fink.<sup>14</sup> Sie propagieren multidimensionale Evaluationen, die in Kombination verschiedener Methoden diverse Level von Erkenntniserwartungen erfüllen sollen:

(1.) Erkenntnisse darüber, ob Projekte überhaupt adäquat auf die Erzielung von Wirkungen und Fähigkeiten ausgerichtet sind, (2.) Erkenntnisse darüber, ob sie ihre Ziele erreichen, (3.) Erkenntnisse darüber, ob Projekte zu den erhofften Programmwirkungen beitragen, (4.) Erkenntnisse darüber, ob die Resultate wirklich die gewünschten Gewinne mit sich bringen und (5.) Erkenntnisse darüber, ob registrierte Gewinne auf Dauer die mit ihnen verbundenen Visionen realisieren helfen. Ferner unterscheiden die Autoren vertikale Evaluationen, die von Beginn an bis hin zur Ergebnisauswertung spezifische Projekte begleiten, und horizontale Evaluationen, die die Umsetzung von Strategien der Prävention in mehreren Einheiten bzw. Maßnahmenbündeln zugleich untersuchen. Romaniuk & Fink betonen die Relevanz von Vorher-Nachher-Vergleichen und Vergleichsgruppeneinsatz für möglichst valide Ergebnisse, unterstreichen die Bedeutung des Einbezugs von Stakeholdern, den Bedarf an Pragmatismus und die Flexibilität bei der Datensammlung, die hilfreiche Funktion von Praxis- und Kontextkenntnissen auf Seiten der evaluierenden Personen sowie die Triangulation, also die Kombination verschiedener quantitativer und qualitativer Methoden. Bei der Datenanalyse heben sie die Bedeutung der Mitwirkung von Praktikerinnen und Praktikern bei der Zuschreibung von Wirkungskausalitäten hervor und plädieren für eine „Fusion“ unterschiedlicher Sichtweisen und Erkenntnisquellen,<sup>15</sup> um der Komplexität sozialer Realität besser als durch ein rein deduktives Verfahren gerecht werden zu können.

Williams & Kleinmann<sup>16</sup> entwickeln eine nutzenfokussierte Anleitung für die Durchführung von Evaluationen zur Verminderung des Terrorismus-Risikos, die durchaus sehr praxistauglich Checklisten für evaluierende Personen und auch einen ausführlichen Selbstevaluationsbogen für die Fachkräfte enthält. Auch sie verweisen auf die zentrale Bedeutung der Stakeholder und ihrer Vorstellungen vom Nutzen ihres Angebots sowie über ihm angemessene Bewertungsmaßstäbe. Deshalb zählen sie die Identifizierung der relevanten Stakeholder und die Beratung mit ihnen über Evaluationsgegenstand und -ziele zu den ersten Schritten einer Evaluation. Im Zuge dessen suchen sie die dem Angebot unterliegende Theorie, also die theoretischen Begründungen für die angenommenen Funktionsmechanismen, zu klären; sie wählen zu deren Untersuchung und zur Untersuchung der real ablaufenden Prozesse, also zur „process evaluation“, die sie als unverzichtbaren Teil der beabsichtigten Untersuchung von Wirkungen verstehen, passende Methoden, vorzugsweise sog. Logische Modelle (s. u. und grundlegend W. K. Kellogg Foun-

<sup>14</sup> Romaniuk/Fink 2012.

<sup>15</sup> Ebd., 17.

<sup>16</sup> Williams/Kleinmann 2013.

dation).<sup>17</sup> Vergleichsmaßstäbe und -gruppen einzubeziehen, wird als wünschenswert betrachtet.

In Absetzung von diesen Ansätzen und unter Verweis auf die mit ihnen verbundenen Schwierigkeiten, Wirkung mit hinreichender Sicherheit festzustellen (siehe dazu auch unten), propagiert Köhler<sup>18</sup> in jüngerer Zeit eine „structural integrity evaluation“. Ihr Ziel ist darauf gerichtet, Angebote der Extremismusprävention daraufhin zu untersuchen, ob sie strukturellen Standards entsprechen, die gut funktionierende Programme erfüllen müssen. Es handelt sich um insgesamt 64 Punkte, die sich auf die sechs Bereiche (1) der Programmleitung und Entwicklung, (2) der Organisation, (3) der Klienteneinstufung, (4) der Beratung und Betreuung, (5) der Qualitätssicherung und (6) der Transparenz verteilen.<sup>19</sup>

Nicht ohne Grund wird es in den vergangenen Jahren als zunehmend wichtiger betrachtet, Evaluationen auf einen konkreten Nutzen auszurichten. Auch die Leserinnen und Leser dieses Handbuchs dürften in erster Linie einen solchen praktischen Nutzen erwarten. Im Weiteren werden daher nicht in akademischer Manier verschiedene Evaluationstheorien, -modelle und -methodologien weiter ausdifferenziert und erörtert; stattdessen wird der Blick auf Konzepte gerichtet, die für die Ziele von Evaluationen im Bereich der Extremismusprävention besonders bedeutsam sind und unmittelbare Praxisorientierung versprechen. Der Fokus liegt auf Konzepten, die die Aussicht bieten, Wirkungsaspekte nicht ganz aus den Augen zu verlieren und mehr als nur Strukturaspekte von Angeboten zu untersuchen. Dafür sind zunächst jedoch einige grundlegende begriffliche Klärungen unumgänglich. Sie betreffen Gegenstände und Zielsetzungen von Evaluation, die Zeitpunkte ihrer Durchführung und die Frage, wer sie ausführt.

## Evaluationsgegenstand: Programm, Projekt, Einzelmaßnahme? – primär-, sekundär- oder tertiärpräventiv?

Ein erster Fragenkomplex, der zu klären ist, bezieht sich auf das Evaluationsobjekt – solche Fragen stellen sich hier mindestens in zweifacher Weise:

Erstens ist zu berücksichtigen, ob das zu evaluierende extremismuspräventive Angebot eine einzelne Veranstaltung, ein Projekt mit einem Bündel von Maßnahmen und/oder sonstigen Elementen oder ein ganzes Programm mit verschiedenen Modulen bzw. Schwerpunkten ist, die wiederum aus einzelnen Projekten und Maßnahmen zusammengesetzt sind. Der in der Fachwelt gängige, wohl aus dem angelsächsischen Raum entlehnte, diese Differenzen ignorierende und vereinheitlichende Evaluations-Sprech von „Programmen“, wenn von Evaluationsgegenständen die Rede ist, darf nicht darüber hinwegtäuschen, dass sich konkrete Aufgaben, Möglichkeiten und Grenzen ganz anders darstellen, je nachdem, ob man es mit einem einzelnen Typus von personenbezogener Dienstleistung zu tun hat oder mit komplexen Zusammenhängen unterschiedlicher Angebotsformen.

Die Erwartungen, die realistischer Weise an eine Evaluation gestellt werden können, differieren dementsprechend zwangsläufig. Schon an dieser Stelle sind ganz praktisch ge-

<sup>17</sup> W. K. Kellogg Foundation 2001.

<sup>18</sup> Köhler 2017.

<sup>19</sup> Vgl. Köhler 2016.



wisse Richtungsentscheidungen dahingehend zu treffen, ob – einmal ganz abgesehen von den vordringlichen Erkenntnis- und Verwertungsinteressen sowie den damit verbundenen Zwecken – allein schon die Anzahl der einzubeziehenden Fälle bzw. Personen quantitative Vermessungen nahelegt oder nicht eher eine qualitative Methodologie, vielleicht aber auch die Kombination von beidem im Sinne von „mixed methods“. Es gilt hier bspw. zu entscheiden, welcher Stellenwert prozentualen Aussagen oder Mittelwerten bei einer Gesamtstichprobe von geringer Größe überhaupt beigemessen werden kann und wie sich diese Frage bei Interesse an genaueren Untergruppen-Differenzierungen beantworten lässt. Und nicht nur, wenn Programme besonders komplex sind, stellt sich das Problem, wie sich neben einem zahlenmäßigen Überblick eine wünschenswerte Auswertungstiefe erreichen lässt.

Zum zweiten gilt es zu beachten, ob der Evaluationsgegenstand in Feldern der primären bzw. universellen (General-)Prävention liegt, ob selektiv, also gezielt bei Personen(gruppierungen) mit erhöhtem Risiko oder indiziert bzw. sekundärpräventiv, also bei Personen mit bereits aufgetretenen Problemscheinungen, angesetzt wird, oder ob Klientinnen und Klienten, die bereits sicherheitsbehördlich auf- und straffällig geworden sind, tertiärpräventiv behandelt werden.

Im tertiären und mit gewissen Abstrichen auch im sekundärpräventiven Bereich drängt sich im Zusammenhang mit dem erstgenannten Punkt bei Projektevaluationen schon deshalb ein qualitativ-rekonstruktives Vorgehen auf, weil die Fallzahlen hier vergleichsweise gering sind und zahlengestützte Auswertungen daher entweder wenig Sinn machen oder nur den Rang quantifizierender Fall-Übersichten einnehmen (können).

Anders verhält es sich mit Programm-Evaluationen, bei denen quantitative Befunde allerdings bislang in Deutschland kaum erzielt wurden – im Islamismusbereich zumindest auch aufgrund der relativen Neuigkeit des Arbeitsfeldes, im Bereich der Distanzierungsarbeit sowie spezieller Ausstiegshilfen aus dem Rechtsextremismus dagegen vermutlich aus Gründen nur schwach gegebener Vergleichbarkeit verschiedener Bundes- und Landesprogramme und zivilgesellschaftlicher Initiativen (Ausnahmen im Bereich der Beratung zu Islamismusphänomenen;<sup>20</sup> zum Themenbereich Rechtsextremismus-Aus- bzw. -Umstiege).<sup>21</sup> Falls doch Zahlen ermittelt werden, liefern sie bislang im Allgemeinen nur sehr grobe, in ihrem Zustandekommen wenig transparente und nicht sonderlich aussagefähige Daten.<sup>22</sup> In diesen Präventionsfeldern muss ohnehin weniger die Zahl (sowieso hochgradig definitionsabhängiger) realisierter Aus- und Umstiege interessieren. Vielmehr geht es um Antworten auf die Frage, warum und wie Distanzierungen von un- und antidemokratischen Haltungen erreicht werden können. Denn die hier anzusiedelnden Arbeitsfelder der Beratung und unterstützenden Begleitung können – auch international betrachtet – nur äußerst unzureichend auf entsprechende Erfahrungen zurückgreifen und erst recht kaum auf darauf bezogenen wissenschaftlichen Analysen und Systematisierungen aufbauen. Die einschlägige Forschung ist sich mehr oder weniger einig: Quantitativ verfahrenende Untersuchungen allein können auf diesen Präventionsgebieten keine differenzierten Erkenntnisse für die Anlage von Erfolg versprechenden Praxisprozessen hervorbringen. In der universellen Extremismusprävention hingegen kann quantitativen Studien deutlich höherer Stellenwert zu-

<sup>20</sup> Uhlmann 2017.

<sup>21</sup> Vgl. Deutscher Bundestag 2018; Einhorn et al. 2012, 2013; Becker/Brunsen/Einhorn 2014.

<sup>22</sup> Ebd.

gesprochen werden. Hier sind Stichprobenumfänge zu erzielen, die differenzierte Auswertungen erlauben.

Aus Sicht der Fachkräfte bleibt freilich offen, wie viele Gewinne für eigene konkrete Arbeitsabläufe aus solchen Daten gezogen werden können und wie eng nachverfolgt werden kann, wie(so) ein registrierter positiver Effekt mit bestimmten Ausgangsanalysen, spezifischen Konzeption(sentwicklungen), einzelnen Prozessaktivitäten, verlaufsbezogenen Rahmenbedingungen und darauf einwirkenden professionellen Orientierungen in Zusammenhang zu bringen ist. Dessen ungeachtet wird unter Expertinnen und Experten die Grundsatzfrage nach dem Sinn von demokratiefördernden Angeboten als universelle Extremismusprävention diskutiert. Auf der einen Seite wird dabei argumentiert, über rechtsextreme Diskurse und Inszenierungen sowie über Islamismus müsse breit, kritisch und vorbeugend aufgeklärt werden. Eine solche Prävention versteht sich als politische Bildung, ihr Kerngeschäft sind demnach Schlüsselqualifikationen der Demokratie: Hinterfragen, Differenzieren, auch das Trennen von Inszenierung und Inhalt. Sie zielt zum Beispiel auf Medienkompetenz<sup>23</sup> und „inhaltliche Auseinandersetzungskompetenz“ mit menschenfeindlichen Diskursen.<sup>24</sup> Dazu zählt die Auseinandersetzung mit Hate Speech und Fake News im Netz. Extremismusprävention ist demnach ein Element der Demokratiepädagogik und kann mit handlungsorientierten, Zivilcourage stärkenden Methoden verbunden werden, zum Beispiel Argumentationstrainings gegen Parolen und Populismus.<sup>25</sup> Die andere Seite fragt, ob nicht statt universeller Extremismusprävention, die mit ihrer Verhinderungslogik nahezu zwangsläufig in den Geruch von Etikettierung und Stigmatisierung bestimmter (jugend-)kultureller Ausdrucksweisen (z. B. Skinheadsein) oder auch ganzer Bevölkerungsgruppierungen (z. B. von Musliminnen und Muslimen) kommen könne, eher (Intentionen der) Demokratiebildung und Demokratieförderung die grundlegende Zielrichtung sein sollten, um ein „Rationalitätsschema“ des „aktivistischen Negativismus“<sup>26</sup> und die Ausweitung einer Verdächtigungskultur ohne konkrete Anlässe mit unter Umständen sogar kontra-produktiven Folgen des Rückzugs in zur Gesamtgesellschaft abgeschottete Diskurszusammenhänge und Gemeinschaftsformen zu vermeiden.

## Ziele: Wirkung oder was?

Spätestens, wenn der Evaluationsgegenstand in seinen Umrissen geklärt erscheint, stellt sich die Frage, mit welchen Zielen Evaluation durchgeführt werden soll. In den meisten Fällen drängt sich dann das Interesse an Wirkungsevaluation in den Vordergrund. Konkreter: Es soll herausgefunden werden, welche Folgen, genauer: welche Ergebnisse, welchen Erfolg bzw. welchen Nutzen und welche gesellschaftlichen Gewinne ein extremismuspräventives Angebot hat. Mehr noch: Es soll herausgefunden werden, welche über einen gewissen, jeweils festzulegenden Zeitverlauf hinweg registrierten Veränderungen tatsächlich auf welche Elemente des Angebots zurückgeführt werden können, was also nachweislich von ihnen und nicht von irgendwelchen anderen Faktoren bewirkt wurde. So naheliegend der Wunsch sein mag, hier Eindeutigkeiten zu schaffen, so schwierig ist es, die Wirkung belastungsfähig feststellen zu können.

<sup>23</sup> Vgl. z. B. Glaser/Pfeiffer 2017, 15.

<sup>24</sup> Elverich 2011, 420; vgl. auch Heinrich 2017, 169 f.

<sup>25</sup> Hufer 2017.

<sup>26</sup> Bröckling 2017, 75 u. 86.

Zu bedenken ist, dass Maßnahmen und Projekte der Extremismusprävention und die damit verbundenen Bildungsprozesse nicht im „luftleeren Raum“, d. h. nicht in von der Umwelt hermetisch abgeschotteten Laborkonstellationen, stattfinden. Außer den Angeboten selbst wirkt erfahrungsgemäß eine Vielzahl von äußeren Faktoren auf die evaluativ interessierenden Prozesse ein, zumal dann, wenn sie sich über einen längeren Zeitraum erstrecken und nicht nur auf den Effekt von Kurzzeitpädagogik vertrauen. Noch erheblich mehr gilt dies für die Einflusspotenziale, denen ganze Programme mit ihren diversifizierten Kontexten über ihre Laufzeit hinweg und in ihren möglichen Resultaten bei den Nutzerinnen und Nutzern auch noch nach ihrer Beendigung ausgesetzt sind.

Hinzu kommen weitere Schwierigkeiten, Wirkung festzustellen: Zu klären ist, worin Wirkung eigentlich besteht. Inwieweit ist sie als objektives Faktum registrierbar und inwieweit hängt ihre Definition von denen ab, die sie registrieren und einschätzen? Gerade beim beabsichtigten Abbau schon bestehender extremistischer Haltungen offenbart sich diese Problematik auch als methodische Schwierigkeit besonders deutlich: Wenn z. B. jemand nach Durchlaufen einer Maßnahme, etwa durch entsprechend gesetzte Kreuzchen in einem Fragebogen, zu erkennen gibt, nunmehr weniger extremistisch eingestellt zu sein als vorher, ist er oder sie es dann tatsächlich? Sind solche oder andere Instrumente der Einstellungsmessung lebensweltnah genug, um nicht nur in einer künstlichen Befragungssituation (bestenfalls) valide Befunde hervorzubringen? Besitzen sie auch Aussagekraft für Positionierungen im Alltag oder gar in speziellen Konfliktlagen, und haben sie wenigstens kurzfristig einen gültigen prognostischen Wert? Vor allem aber: Sind die angezeigten Veränderungen ursächlich auf die Teilnahme an der Maßnahme zurückzuführen, handelt es sich bei ihnen also nicht nur um Brutto-, sondern um Netto-Wirkungen, also Effekte, die ohne Teilnahme nicht eingetreten wären? Wann kann von ideologischer Distanzierung gesprochen werden? Und ist es eine „innere“ Deradikalisierung, oder nur eine äußerlich sichtbare Verhaltensänderung, also ein komplettes oder wachsendes „Disengagement“ und eine Demobilisierung, die offenlässt, ob auch ein „Umdenken“ stattgefunden hat? Zudem: Wenn andere Personen, z. B. Betreuende oder pädagogische Begleitende dieser Person, den Sachverhalt bei ihr ganz anders sehen, liegt dann die zunächst sich andeutende Wirkung doch nicht vor?

Offenbar ist Wirkung, will man überhaupt weiter davon sprechen, vor allem etwas, worauf man sich intersubjektiv verständigen muss. Es ist daher sinnvoll, wirkungsorientierte Evaluation so anzulegen, dass Wirkannahmen und Wirkungseinschätzungen unterschiedlicher Beteiligter an einem Angebot (z. B. Teilnehmende, Referierende, Angebotsplanerinnen und -planer) und ggf. noch zusätzlich von Externen (etwa privaten Bezugspersonen, Lehrpersonen und sonstigen Fachkräften, die mit Teilnehmerinnen und Teilnehmern außerhalb des Angebots selbst zu tun haben) eingeholt werden, um das, was sich als Wirkung darstellt, im Abgleich solcher Sichtweisen multiperspektivisch zu erschließen. Dabei gilt es auch, zu entscheiden, ob es ausreichend ist, Wirkungseinschätzungen mittels standardisierter Erhebungs- und Auswertungsverfahren zu analysieren, oder ob es nicht angezeigt ist, die subjektiven Relevanzen und Wertungen, die die jeweiligen Wirkungsauffassungen sowie die Annahmen von Wirkungsbedingungen bestimmen, qualitativ-rekonstruktiv zu erfassen, um nicht nur zu überprüfen, ob und in welchem messbaren Ausmaß (eine wie auch immer definierte) Wirkung eingetreten ist, sondern auch, warum und wie sie zustande kam (vgl. zu Lösungsansätzen der aufgezeigten Problematiken weiter unten in diesem Beitrag).

So hoch auch das Interesse an eindeutigen Wirkungsaussagen und sich unumstößlich präsentierenden Kausalitätszuweisungen vor allem auf Seiten von politisch Verantwortlichen und Financiers sein mag: Evaluation muss nicht unbedingt Wirkungsevaluation sein.

Denkbar und praktisch weiterführend können auch ganz andere Bewertungskriterien sein, z. B.: Werden Leistungen erbracht, die den Ansprüchen der Adressatengruppierungen oder auch den Erwartungen anderer Stakeholder gerecht werden? Wie sind Arbeitsprozesse anzulegen, damit sie aus fachlicher Sicht sachlich geeignet, zielführend und effizient sind? Welche konzeptionellen Grundlegungen lassen die Arbeit Erfolg versprechend erscheinen? Wie beeinflusst die Professionalität involvierter Fachkräfte die Angebotsplanung, -anbahnung, -durchführung und -auswertung? Bei welchen strukturellen Bedingungen werden welche Resultate erzielt<sup>27</sup> und welche Kooperationen sind ihnen zuträglich? Wie kann ein Angebot Nachhaltigkeit entfalten (zu Fragen wie diesen siehe unten)? Gerade Träger, Einrichtungen und Mitarbeitende in der Prävention und nicht zuletzt ihre Adressatinnen bzw. Adressaten und deren Angehörige, aber auch große Teile einer wohlinformierten Öffentlichkeit dürften an Antworten auf diese Fragestellungen besonders interessiert sein. Auch kann manchmal eine wissenschaftlich fundierte externe Begleitung Fehlentwicklungen viel früher entdecken, als die Akteure selbst dazu in der Lage wären. Sie kann dann helfen, mit dem Blick über den engen Tellerrand des eigentlichen Arbeitsprozesses der Evaluierten hinaus Fehlentwicklungen frühzeitig gegenzusteuern und Lösungen zu erarbeiten. In diesem Sinne bietet Evaluation die Chance, selbst präventive Wirksamkeit zu entfalten, vorausgesetzt, sie ist in der Lage, Arbeitsprozesse intensiv und kritisch, aber konstruktiv und vertrauensgeleitet zu begleiten. Das setzt ferner auch voraus, dass Evaluation sowohl ihr Selbstverständnis als auch ihr Verhältnis zum Untersuchungsgegenstand entsprechend justiert (dazu unten).

## Evaluationszeitpunkte: vorher – nachher – begleitend?

Wirkungsevaluation legt ein Untersuchungsdesign nahe, das mit Vorher-/Nachher-Erhebungen operiert und ggf. auch „on-going“ während des Angebotsverlaufs Datensammlung im Forschungsfeld betreibt, um Entwicklungen in bestimmten Praxisphasen und -stadien nachzeichnen zu können. Hat man nicht mindestens zwei Vergleichszeitpunkte, lassen sich Veränderungen, also etwa die Abnahme extremistischer Haltungen, im Nachhinein nur sehr eingeschränkt und unter Inkaufnahme von Wahrnehmungs- und Beurteilungsverzerrungen identifizieren. Auch wenn – nicht zuletzt aus pragmatischen Gründen – solche Verfahren, zumeist dann von Trägern bzw. Einrichtungen selbst durchgeführt, oftmals als Evaluationen ausgegeben werden, sind bloße Ex-post-Analysen für Wirkungsaussagen wenig verlässlich, umso mehr, wenn nur durchführende Fachkräfte befragt werden. Geht es nur um die Teilnehmerzufriedenheit, so mögen Ex-post-Befragungen ausreichend erscheinen. Nichts desto weniger haben nahezu jegliche evaluatorischen Analysen, die nach Beendigung eines Programms, eines Projekts oder einer Maßnahme stattfinden, summative bzw. bilanzierende Intentionen: Sie geben – mit welcher Aussagekraft und Reichweite auch immer – eine Gesamtbeurteilung des Evaluationsgegenstands vor dem Hintergrund von Ergebnisfeststellungen ab und ziehen ein Fazit, das als Entscheidungshilfe für den weiteren Umgang mit ihm betrachtet werden kann. In Bezug auf die Evaluation von Bereichen der Extremismusprävention

<sup>27</sup> Vgl. z. B. den Vorschlag von Köhler 2017; s. o.

klafft hier eine besonders große Lücke in Hinsicht auf die Möglichkeit, wissenschaftlich belastbare Aussagen über Langfristwirkungen und Nachhaltigkeitssicherungen zu treffen. Es fehlt eklatant an Erhebungen, die in sinnvollen Abständen vom Ende der jeweils evaluierten Maßnahmen, Projekte und Programme, z. B. etwa sechs oder zwölf Monate oder noch später nach ihrem Abschluss, durchgeführt und ausgewertet werden.

Reine Ex-ante-Evaluationen mit präformativem Charakter kommen hingegen infrage, wenn in der Phase der Planung Aufgaben der Programmentwicklung unterstützt werden sollen. Derartige klärende Funktionen erfüllen sie dort, wo geeignete Rahmenbedingungen zu eruieren sind, die wissenschaftlichen Verankerungen der Angebotsanlage untersucht oder erwartbare Langfristeffekte bzw. Nachhaltigkeitschancen abgeschätzt und noch Korrekturen an Angebotsentwürfen ermöglicht werden sollen. Evaluationskonzepte verstehen sich hier als aktiv gestaltende, konstruktiv oder geradezu evolutiv wirksame Hilfen für Vorab-Steuerungen. Einige Evaluationskonzepte beginnen noch früher, indem sie proaktiv und dabei auf vorgängige Evaluationen und sonstige wissenschaftliche Erkenntnisse gestützt in der Vorlaufphase von Praxis Rahmenkonzepte für sie anbieten, innerhalb derer sich dann in der Folge Praxis und Evaluation bewegen.<sup>28</sup>

Ebenfalls unter Umständen summativ und bilanzierend, vor allem aber formativ, also das Angebot mitgestaltend, gehen Evaluationen vor, die angebotsbegleitend angelegt sind; sie nehmen kontinuierlich während der Laufzeit des zu evaluierenden Programms, Projekts oder Einzelangebots Durchführungs- und oft auch schon Planungs- und später von der Praxis vorgenommene Auswertungsprozesse in den Blick und geben auf dieser Basis Feedbacks. Solche Konzepte galten lange Zeit aus manchen wissenschaftlichen Blickwinkeln als nicht unproblematisch, weil sie den Untersuchungsgegenstand noch während des Untersuchungsprozesses durch Interaktion mit Praxisverantwortlichen beeinflussen. Wer wissenschaftliche Qualität daran bemisst, in welchem Maße es Forschung gelingt, aus einer strikt eingehaltenen Außenperspektive möglichst wertneutral ihre Studienobjekte zu erfassen, um nicht durch eigene Eingriffe Störfaktoren in die mit ihnen zusammenhängenden Sachverhalte einzubringen und so mehr oder minder zwangsläufig (auch) eigenes Handeln zum Forschungsgegenstand zu machen, warnt vor solchen Konzepten der Evaluation. Wer hingegen als Voraussetzung für ein Höchstmaß an praktischer Wertbarkeit der zu erzielenden Erkenntnisse Praxis und Wissenschaft als notwendigen Interaktions-, ja Kooperationszusammenhang denkt und konsequent auf Anwendungsorientierung setzt, kann ohne Dialog- und Partizipationsangebote an die Nutzerinnen und Nutzer von Evaluation nicht auskommen. Er hält es größtenteils sogar für ethisch zweifelhaft, z. B. sie nur deshalb nicht auf bereits während der Laufzeit festgestellte Fehlentwicklungen der Praxis hinzuweisen, damit die wissenschaftliche Distanz gewahrt bleibt und sie infolgedessen gleichsam „sehenden Auges“ den „Karren vor die Wand fahren“ zu lassen.



**Summative Evaluation:** Die summative Evaluation erfolgt am Ende eines durchgeführten Programmes/Projektes. Der Evaluationsgegenstand wird einer Gesamtbeurteilung unterzogen und mit Hinblick auf die Ergebnisse bilanzierend beurteilt.

**(Prä)formative Evaluation:** Die formative Evaluation dient der Bewertung eines noch stattfindenden Prozesses, sie ist mitgestaltend und greift dadurch verbessernd in Prozesse ein. Beispielsweise unterstützt die formative Evaluation bei der Angebotskonzeptualisierung und bei Aufgaben zur Programmentwicklung.

<sup>28</sup> Vgl. z. B. Möller et al. 2017.

Bieten sich aus einer wissenschaftlichen Sicht formativ-gestaltungsorientierte Designs primär dann an, wenn zu evaluierende Programme, Projekte oder Maßnahmen großenteils Innovationen darstellen, noch mit deutlichen Unwägbarkeiten belastet sind, eher experimentellen Charakter besitzen und deshalb noch keinen Reifegrad erreicht haben, der eine Konzentration auf summativ-bilanzierende Bewertungen angezeigt erscheinen lässt, so ist vor dem Hintergrund von Praxis-Interessen vom explorativen Funktionspotenzial solcher Designs am ehesten zu erwarten, dass sie leicht handhabbare, anwendungsorientierte Empfehlungen abgeben und zugleich zukunftsfähige Handlungslogiken kollaborativ mitkonstruieren können, ohne gänzlich auf Bilanzierungen verzichten zu müssen. Im Bereich der Extremismusprävention, zumal dort, wo es um sekundär- und tertiärpräventive Aufgaben geht, kann beides als gegeben gelten.

## Evaluationsorte: Intern – extern, Selbst- und/oder Fremdevaluation?

i

**Selbstevaluation:** Eine Selbstevaluation kann in unterschiedlichen Formen erfolgen, beispielsweise als Dokumentation von Arbeitsabläufen, als Selbsteinschätzung mithilfe eines standardisierten Fragebogens, als Praxis-Tagebuch oder als kollegiales Reflexionsgespräch.

**Externe Evaluation:** Eine externe Evaluation wird (im Gegensatz zu der internen Evaluation) nicht vom eigenen Personal oder dem des Projektträgers durchgeführt, sondern durch ein Forschungsteam, welches extra dafür herangezogen wird.

Grundsätzlich steht bei jedem Evaluationsvorhaben die Entscheidung darüber an, ob intern, also durch Personal der eigenen Einrichtungen bzw. des eigenen Trägers, oder extern durch hinzuzuziehende Forschungsteams evaluiert werden soll. Als Vorteil selbst durchgeführter Evaluationen gegenüber entsprechenden Auftragsvergaben nach außen kann angeführt werden, dass sie im Regelfall auf den ersten Blick preiswerter erscheinen, die Praxis aus größerer Nähe betrachten können, im Falle schlechter Ergebnisse weniger Aufmerksamkeit erregen und vielleicht auch innerhalb der Organisation

bei den Mitarbeiterinnen und Mitarbeitern einen Vertrauensvorschluss für sich verbuchen können. Allerdings kann das, was als augenscheinlicher Vorteil erscheint, auch schnell in sein Gegenteil umschlagen: Scheinbare Ersparnisse ergeben sich unter Umständen nur, weil die Kosten durch interne Abrechnung verdeckter gehalten werden und sich in anders ausgewiesenen Ausgabenposten verstecken; die Nähe zur zu evaluierenden Praxis kann zu Betriebsblindheit, ungerechtfertigten Selbstverständlichkeitsannahmen und Ausblendungen ungewöhnlicher, aber genau deshalb weiterführender Sichtweisen führen. Negative Ergebnisse als Interna zu behandeln, positive dagegen in die Öffentlichkeit zu tragen wirft kein gutes Licht auf die Organisation und kann in der Außendarstellung zu der Unterstellung führen, man habe sich schlichtweg selbst Gefälligkeitsbeurteilungen ausgestellt; Mitarbeitenden des organisationsinternen Qualitätsmanagements wird womöglich mit nicht minder viel Skepsis oder sogar Misstrauen begegnet, als völlig Fremden – vielleicht gerade deshalb, weil sie bekannt sind und ihnen möglicherweise heimliche Kollaboration mit der Geschäftsleitung und Weitergabe als vertraulich eingestuft Information an sie unterstellt werden. Ferner ist die steigende Arbeitsbelastung gerade für kleinere Träger eine Herausforderung. Sie bereits in der Planungsphase einzukalkulieren, kann durchaus selbst als Qualitätskriterium gelten. Manche Träger befürchten mit Blick auf die zu beantragenden Mehrkosten negative Folgen für die Beantragung. Verringern die höheren Kosten die Chancen auf positive Bescheide? Widerstände gibt es über die Arbeitszeithorizonte hinaus seitens des Fachpersonals manchmal gegen ein allzu elaboriertes Dokumentationswesen: Nicht jeder gute Praktiker ist auch ein guter Schreiber.

In jedem Fall lässt sich wohl davon ausgehen, dass einer Evaluation, die von einer unabhängigen Instanz, etwa einem Forschungsteam einer Hochschule, geleistet wurde, in der (Fach-)Öffentlichkeit mehr Objektivität zugeschrieben wird, als einer internen Aus- und Bewertung.

Wenn also vieles für externe Evaluationen spricht, so bedeutet dies keinesfalls, dass Selbstevaluation damit überflüssig wird. Im Gegenteil: Aktiv Qualitätskriterien und -standards, Methoden und Prozessabläufe zu diskutieren und zu gestalten (anstatt sie gewissermaßen von außen vorgesetzt zu bekommen), kann nicht nur eine attraktive Aufgabe sein, sondern zwingt zur intensiveren Auseinandersetzung mit Arbeitsfeld und -abläufen als Querschnittsaufgabe. Insofern ist die Selbstevaluation von Mitarbeitenden als ein wichtiger Baustein innerhalb eines größeren Komplexes von evaluatorischen Verfahren anzusehen.<sup>29</sup> In welcher Form sie auch immer geschieht (z. B. als bloße Dokumentation von Arbeitsabläufen, als Selbsteinschätzung mithilfe eines standardisierten Fragebogens, als Praxis-Tagebuch, als kollegiales Reflexionsgespräch), sie bringt eine wichtige Perspektive auf das Geschehen, das zu evaluieren ist, und dessen Produkte ein. Bedeutung hat sie jedoch nicht nur innerhalb eines befristeten Evaluationsprojekts, sondern auch als arbeitsalltagsbezogenes Verfahren der Vergewisserung, des Überdenkens und ggf. der Neuausrichtung professionellen Handelns.

## Zwischenfazit 2

Konzepte und Ziele von Evaluation können theoretisch, methodologisch und ganz praktisch erheblich variieren. Auch wenn deshalb abhängig von der jeweiligen Ausrichtung die Gütekriterien, unter denen sie bewertet werden, variieren, bleibt konsensfähig zu konstatieren, dass Evaluationsansätze vor allem eines sein müssen: gegenstandsangemessen und zielorientiert. Je stärker dabei Nutzungsinteressen unmittelbar Beteiligter befriedigt werden sollen (und weniger Kontrolle und Legitimation durch Mittelgeber als vielmehr gut praxisverwertbare Erkenntnisse und Entwicklungsinteressen von Programmen, Projekten und Maßnahmen im Fokus von Funktionserwartungen stehen), umso mehr sind Konzepte gefragt, die gestaltungs- und prozessorientiert angelegt sind und nicht nur auf (summative) Bilanzierungen setzen, sondern, ohne Praxisverstrickungen zu riskieren, über aufmerksame kontinuierliche Begleitung Potenziale für Qualitätssicherung und -entwicklung aufzudecken vermögen. Soweit sie dabei Interaktion mit und Partizipation der Evaluierten praktizieren, müssen sie ihr Design so flexibel halten, dass auf neu auftretende Herausforderungen responsiv reagiert werden kann.

## Formen und Vorgehensweisen

Es liegt auf der Hand, dass über Formen und konkrete Vorgehensweisen von Evaluationsvorhaben, etwa Erhebungsinstrumente und Auswertungsverfahren, nicht unabhängig von der Verfolgung spezifischer inhaltlicher Interessen, dafür grundlegenden konzeptionellen Überlegungen und den mit ihnen verknüpften Zielorientierungen entschieden werden kann. Unbeschadet dessen kann ein Spektrum aufgezeigt werden, das die untersuchbaren Dimensionen und methodischen Herangehensweisen absteckt und dies mit dem Blick auf aktuelle Erfordernisse der Evaluation von Extremismusprävention tut.

<sup>29</sup> Vgl. auch Williams/Kleinman 2013.



## Evaluationsdimensionen – Ergebnisse, Prozesse, Konzeptionen, was noch?

Stellen wir in Rechnung, dass die Extremismusprävention insgesamt, und hier verschärft die vergleichsweise noch junge sog. Islamismusprävention, nicht nur bislang unbefriedigend evaluiert ist und wenig um ihre tatsächlich erzielten Wirkungen weiß, sondern auch ihre Folgen nur höchst unzureichend auf deren Entstehungs- und Entwicklungsfaktoren zurückführen kann,<sup>30</sup> so ist nicht nur die qualitative Beschaffenheit ihrer Ergebnisse fraglich, sondern auch ihre jeweilige Genese. In der Konsequenz erscheint es prinzipiell sinnvoll, Evaluationsdimensionen nicht nur auf Ergebnisqualitäten auszurichten. Vielmehr sind auch die Ausgangsbedingungen und ihre Einschätzungen bzw. Analysen durch Präventionsverantwortliche, die zur Anwendung kommenden Konzeptionen (bzw. Konzepte; zur Unterscheidung dieser Begriffe s. u.), die konkreten Arbeitsprozesse und die Strukturbedingungen extremismuspräventiver Angebote zu analysieren. Schon aus pragmatischen und ökonomischen, und nicht nur aus inhaltlichen Gründen können dabei in konkreten Evaluationsprojekten Schwerpunktsetzungen innerhalb dieses Spektrums erfolgen.

Soll (unter anderem oder primär) auf die Ergebnisse der Präventionsanstrengungen geblendet werden, so empfiehlt es sich, nicht allein auf den leicht sicht- und zählbaren Output – Teilnehmerzahlen, entstandene Produkte etc. – zu achten. Es müssen vor allem die Outcomes fokussiert werden, also die zielkonformen Veränderungen, die sich bei den Haltungen, d. h. in den Orientierungen und Aktivität(sbereitschaft)en, von erreichten Adressatinnen und Adressaten zeigen (vgl. hierzu wie zum Folgenden auch Abb. 1 weiter unten). Nicht weniger interessant als solche Zielerreichungsüberprüfung dürfte ein systematischer Blick auf nicht-intendierte Resultate sein: Werden Haltungsveränderungen erreicht, die gar nicht beabsichtigt waren, jedoch positiv einzuschätzen sind? Aber auch: Gibt es kontra-produktive (oder als ambivalent zu betrachtende) Ergebnisse zu verzeichnen? Insbesondere dann, wenn daneben auch die Auswirkungen von Angeboten auf die Angebotskontexte von Interesse sind, gilt es auch, ihre eventuell vorhandenen Abstrahl-effekte, die Impacts, zu erfassen. So kann man ergründen, in welcher Weise das indirekt über den Einfluss von Teilnehmenden Erreichte, der eigene Träger, aber auch andere vergleichbare Träger oder ähnlich gelagerte Programme, Projekte und Maßnahmen in seinem Umfeld von den Erfahrungen des evaluierten Angebots berührt werden und womöglich von ihm profitieren können.



### PROZESSEVALUATION

*Die Prozessevaluation eignet sich, um Rückschlüsse auf Effektivität, Effizienz und Geeignetheit einer Maßnahme ziehen zu können, sowie um Arbeitsprozesse auf ihre fachliche Angemessenheit zu überprüfen. Dabei werden Prozesse analysiert, um festzustellen, ob das Programm/Projekt das erfüllt oder erreicht, wofür es entwickelt wurde.*

Ohne pädagogischen Optimismus in der Extremismusprävention überzustrapazieren, ist anzunehmen, dass Resultate im Regelfall nicht zufällig zustande kommen, sondern durch bestimmte Praktiken der Angebotsdurchführenden (mit-)verursacht werden. Um hierzu Rückschlüsse auf Effektivität, Effizienz und Geeignetheit ermöglichen zu können, auch jedoch, um Arbeitsprozesse auf ihre fachliche Angemessenheit überprüfen zu können, lohnt sich Prozessevaluation. Diese kann dadurch

30 Vgl. aktuell zusammenfassend z. B. Gansewig 2018; Walsh et al. 2018; Gruber/Lützing 2017; Kober 2017; Kober/Armborst 2017.



strukturiert werden, dass nach den sachlichen, zeitlichen, räumlichen und sozialen Aspekten der Prozesse sowie deren Verflechtung und Funktionalität geschaut wird. Zum Einsatz kommende Praktiken sind als Umsetzungen konzeptioneller Planungen versteh- und analysierbar. Insofern macht es Sinn, zu eruieren, inwieweit sich hier Konzeptionelles realisiert – und umgekehrt, inwieweit nicht. Zum einen geht es dann darum, herauszufinden, ob und inwiefern sich darin institutionelle Konzeptionen widerspiegeln, zum anderen ist zu untersuchen, ob und in welcher Weise sich individuelle Konzepte verwirklichen, die einzelne Angebotsdurchführende besitzen und deren Elemente sich nicht unbedingt mit der institutionellen Konzeption decken müssen, ja vielleicht ihr sogar entgegenlaufen.<sup>31</sup>

Konzept(ions)analysen können aber auch ganz unabhängig von Umsetzungsfragen Evaluationsaufgaben bilden, etwa dann, wenn es darum geht, Wirkungs- und Handlungsziele zu bestimmen und sie in einen sinnhaften Zusammenhang mit Inhalten und Methoden zu bringen. Erfahrungsgemäß ist in der Anfangsphase von Evaluationsprojekten das Zielsystem (nicht nur) der Präventionspraxis nicht immer in sich konsistent und zudem wenig ausdifferenziert. Wirkungsannahmen sind dann entsprechend diffus und die Mechanismen möglicher Wirkung nicht detailliert zu benennen. Meist fehlt es auch an Angaben zu Indikatoren, mit deren Hilfe Zielerreichung später überprüft werden kann. Entsprechende Festlegungen sind aber notwendig, wenn Ergebnisauswertungen zum Abschluss von (Präventions-)Vorhaben nicht nach Gusto, sondern mit systematischem Bezug auf Zielsetzungen vorgenommen werden sollen. Je nachdem, wie partizipativ das Evaluationsverständnis der Evaluierenden ausfällt, wird man hier nur auf bereits Vorhandenes, vornehmlich in dokumentierter Form, Bezug nehmen oder zusätzlich auch Prozesse initiieren (müssen), die zu entsprechenden konzeptionellen Ausarbeitungen führen.

Die Qualität von Konzepten und Konzeptionen wie auch diejenige der Arbeit, der sie Orientierung geben sollen, hängt essentiell davon ab, in welcher Weise die Ausgangsbedingungen, unter denen sie entworfen werden, analysiert werden. Unter dem Gesichtspunkt von Analysequalität ergeben sich Fragen wie: Wird auf aktuelle wissenschaftliche Befunde, den state of the art, aber auch vorhandenes Praxiswissen Bezug genommen? Ist genügend Wissen über die Incomes, also die Haltungen und sonstigen relevanten Merkmale der Adressierten angesammelt und verarbeitet worden? Sind die konkreten sozialräumlichen Kontextbedingungen Gegenstand einer möglichst gründlichen Analyse gewesen? Sind die Inputs, also die Kompetenzen und Ressourcen, die Fachkräfte und andere Mitarbeitende einbringen können, in ihrem Umfang und in ihren Qualitäten geklärt und nutzbar? Wurden die strukturellen Bedingungen, die durch die Arbeit in der Trägerorganisation und ggf. mit weiteren Kooperationspartnern gegeben sind, ausreichend erhellt und bewertet?

## Verfahren und Instrumente

Analyse-, Konzept(ions)-, Prozess-, Struktur- und Ergebnisqualität sind also bedeutsame Evaluationsdimensionen. Unter anderem von ihrer Gewichtung vor dem Hintergrund zu verfolgender Evaluationsziele hängt es ab, welche Untersuchungsverfahren und -instrumente methodisch zum Einsatz kommen.

Die wichtigsten und am meisten zur Anwendung gelangenden Erhebungsinstrumente sind folgende: Literatursichtungen und Auswertungen vorhandener statistischer Daten gelten als bread-and-butter-Aufgaben, ohne die es nicht geht. Standardisierte Fragebögen für Teil-

<sup>31</sup> Vgl. v. Spiegel 2013.

nehmende eignen sich am ehesten, wenn vor allem Ergebnisse und Prozesswahrnehmungen interessieren, ein größerer Stichprobenumfang vorhanden ist und quantitative Überblicke gewünscht werden. Qualitative Interviews unterschiedlicher Art kommen zum Zuge, wo vertiefend der subjektiven Bedeutungshaftigkeit von Phänomenen in allen denkbaren Dimensionen nachgegangen werden soll. Dies kann bei Teilnehmenden, Durchführenden, Kooperationspartnern o. a. für die Fragestellung relevanten zu Befragenden der Fall sein. Dokumentenanalysen (von Konzeptpapieren, Leitbildern, Internetauftritten, Werbeflyern etc.) sind zumindest bei Konzeptionsanalysen unverzichtbar. Gruppendiskussionen sind das Mittel der Wahl, wo kollektiv geteiltes Wissen und typische Orientierungsfiguren (etwa von Fachkräften, die Angebote durchführen) eruiert werden sollen. Teilnehmende Beobachtungen am Präventionsgeschehen ermöglichen einen unmittelbaren Einblick in die realisierte Praxis. Workshops mit Angebotsverantwortlichen und -durchführenden werden veranstaltet, wo dialogisches und partizipatives Evaluationsverständnis vorherrscht und schon während der Laufzeit des Evaluationsprojekts Einflüsse auf den Evaluationsgegenstand zur Geltung gebracht werden können/sollen. Die Auswertung von Selbstevaluationen der Fachkräfte kann nicht nur hier gut brauchbare Ergänzungen liefern.

Wo immer es möglich und angezeigt erscheint, ist ein Methodenmix aus quantitativen und qualitativen Techniken zu bevorzugen – so auch die aktuelle Schlussfolgerung, die aus den vergleichsweise umfangreichen Evaluationserfahrungen des Deutschen Jugendinstituts zu Wirkungsevaluationen in der pädagogischen Präventionsarbeit gezogen wird.<sup>32</sup> Und: Je mehr Perspektiven auf den Gegenstand eingeholt werden können, umso besser kann er ausgeleuchtet werden und umso deutlicher sind auch die wechselseitigen Sichtweisen von Stakeholdern untereinander erkennbar zu machen.

Eine Rangfolge von Auswertungsverfahren und -techniken lässt sich abstrakt nicht aufstellen. Auch hier gilt: Sinnhaft ist das, was gegenstandsangemessen und zielführend ist. Dies können je nach favorisiertem Evaluationsmodell prinzipiell z. B. quantitative Auswertungen von Daten mittels Prozent- und Mittelwertberechnungen oder auch qualitative Verfahren aus unterschiedlichen methodologischen und methodischen Zusammenhängen wie inhalts- oder sequenzanalytische Herangehensweisen sein.

### Zwischenfazit 3

Spezifische Evaluationsbedarfe und -zwecke legen – verbunden mit verfolgten Evaluationskonzepten und -zielen – bestimmte Entscheidungen über Eingrenzungen von Evaluationsgegenständen wie über Formen und Vorgehensweisen ihrer Untersuchung nahe. Hierbei kann – im Rahmen nachweislicher Vereinbarkeit verschiedener methodologischer Ansätze – aus einem Fundus von quantitativen und qualitativen Instrumenten und Verfahren geschöpft werden. Ihre Verwendbarkeit und Kombinierbarkeit ist nicht beliebig, wird aber letztlich neben pragmatischen Erwägungen von dem Erkenntnisgewinn, den sie im Hinblick auf die zentralen Fragestellungen und die dafür bearbeiteten Evaluationsdimensionen des jeweiligen Projekts versprechen, bestimmt.

Grundsätzlich gilt, je transparenter die Vorgehensweise, je stärker die Praxis bereits in der Evaluationskonzeption eingebunden wird und je besser es gelingt, die Chancen von Evaluation vor den befürchteten Risiken zu konturieren, desto fruchtbarer wird für alle Beteiligten der Evaluationsprozess ausfallen.

<sup>32</sup> Vgl. Bischoff u. a. 2018.

## Zum Stand der Anlage von Evaluationen extremismuspräventiver Praxisansätze

Der Stand des auf Grundlagenforschung und Evaluationen gestützten wissenschaftlichen Wissens über extremistische Bestrebungen und deren Prävention wird in seinen inhaltlichen Aspekten in den voranstehenden Kapiteln dieses Handbuchs detailliert dargestellt; er muss hier nicht wiederholt werden, ist aber in einem kurz gehaltenen Überblick auch bei Möller & Neuscheler<sup>33</sup> nachlesbar. Stattdessen ist im Zusammenhang mit den hier entfalteten Überlegungen nach konkreten Anlagemerkmalen vorhandener Evaluationen zu fragen. Insbesondere sind ihre Qualität und deren Standards zu erörtern.

### Merkmale

Eine phänomenübergreifende Sichtung der Präventionslandschaft durch eine Forschungsgruppe des Bundeskriminalamts kommt auf der Basis von Eigenaussagen von 721 im Internet recherchierten, vom 01.01.2014 bis 31.05.2015 gelaufenen Projekten hinsichtlich deren Evaluation zu folgenden Befunden:<sup>34</sup> 59 % der Projekte machen Angaben zu geplanten oder durchgeführten evaluativen Maßnahmen; davon beziehen sich 77 % auf Prozessevaluation und 67 % auf Wirkungsevaluation. Zu 92 % sind externe Stellen mit solchen Evaluationen befasst. Zieht man allerdings die Evaluationen der „Lokalen Aktionspläne“ (inzwischen modifiziert zu „Partnerschaften für Demokratie“) ab, so lassen sich demnach überhaupt „nur in Einzelfällen“ Evaluationsergebnisse finden. Sie attestieren den jeweiligen Projekten nahezu allesamt, (zumindest überwiegend) zielführend zu sein, geben aber in ihrer Mehrzahl nicht zu erkennen, „was konkret im Rahmen der Evaluation gemessen bzw. bewertet wurde“.<sup>35</sup> Insofern verwundert das Gesamturteil des Berichts nicht: „Die vorliegenden Informationen zu Evaluationsmaßnahmen müssen sowohl hinsichtlich Quantität als auch Qualität als äußerst dürftig bezeichnet werden“.<sup>36</sup>

Zu einem vergleichbaren Schluss kommt das Gutachten von Beelmann aus dem Herbst 2017<sup>37</sup> zum Stand der Rechtsextremismusprävention: „Hauptmanko der Radikalisierungs- und Extremismusprävention ist das nahezu vollständige Fehlen aussagekräftiger Evaluationsergebnisse.“

Die Forschungssynthese von Kober und Armbrorst<sup>38</sup> zum Stand der Evaluation von Ansätzen und Maßnahmen islamistischer Radikalisierung in Deutschland kommt nur auf sieben thematisch einschlägige Projekte, zu denen überhaupt frei zugängliche Evaluationsberichte vorliegen, von denen jedoch nur einer im wissenschaftlichen Kontext publiziert ist. Dabei ist fraglich, inwieweit überhaupt von Evaluation gesprochen werden kann, wenn mehr oder weniger nur „einzelne Merkmale der Maßnahmen“<sup>39</sup> benannt werden, diese wiederum zu-

33 Möller/Neuscheler 2019.

34 Vgl. Gruber/Lützing 2017, bes. 16 ff.

35 Ebd., 20.

36 Ebd.; unter internationaler Perspektive liefern eine aktuelle, allerdings kaum in die Tiefe gehende Übersicht zur Anlage einschlägiger Evaluationen in englischer, niederländischer und deutscher Sprache auch Bellasio et al. 2018.

37 Beelmann 2017, 38.

38 Kober/Armbrorst 2017.

39 Ebd., 1.

meist „diffuse Ziele“<sup>40</sup> verfolgen, Berichte allenfalls Teilnehmerzufriedenheiten wiedergeben und „nur für eine der sieben Präventionsmaßnahmen vage Anhaltspunkte über deren Wirksamkeit“<sup>41</sup> angegeben werden können. Auch wenn die Bewertungskriterien, mit denen die Autoren operieren – es handelt sich um die Kriterien des Evidenzindex von WESPE (s. u.) –, höchst strittig sind, bleibt das Resümee zu den in den Berichten vorgelegten Erkenntnissen nachvollziehbar: „Diese Befunde lassen kaum belastbare Rückschlüsse darüber zu, ob die evaluierten Maßnahmen Radikalisierung verhindern, begünstigen oder wirkungslos bleiben.“<sup>42</sup>

Nichtsdestoweniger sind in den letzten Monaten immerhin einige Evaluationen zu islamismuspräventiven Ansätzen vorgelegt worden, die den Anspruch erheben können, weitergehenden wissenschaftlichen Standards zu entsprechen. So informiert die Evaluation der Beratungsstelle „Radikalisierung“ des BAMF<sup>43</sup> einerseits überblicksartig über die Entwicklung des quantitativen Aufkommens von Anrufen bei der dort geschalteten bundeszentralen Hotline (insgesamt 3.163 Anrufe zwischen 01.01.2012 und 31.12.2016) und die Zahl der sich daraus ergebenden Beratungsfälle (854). Andererseits geht sie differenziert und systematisch als Prozess- und Netzwerkanalyse den organisatorischen Eckdaten – wie etwa Zielen, Standards und Prozessen – der vier kooperierenden zivilgesellschaftlichen Partnerorganisationen (Zentrum Demokratische Kultur gGmbH (ZDK) mit der Beratungsstelle „Hayat“ in Berlin und Bonn, Initiative zur Förderung ausländischer Kinder e. V. (IFAK) mit dem „Beratungsnetzwerk Grenzgänger“, Bochum, Verein zur Förderung akzeptierender Jugendarbeit e. V. (VAJA), Bremen mit dem „Beratungsnetzwerk kitab“ und Violence Prevention Network e. V. (VPN), Berlin und Beratungsstellen in Baden-Württemberg, Bayern und Hessen) nach. Eine umfassende Wirkungsanalyse bleibt jedoch künftigen Evaluationsarbeiten beim BAMF vorbehalten.

Noch tiefer in die Analyse der praktischen Beratungsarbeit selbst steigt der Evaluationsbericht zur VPN-Beratungsstelle Hessen ein.<sup>44</sup> Als Resultate präsentiert er umfangreiche Befunde zu Ausgangsbedingungen, Strukturen, konzeptionellen Grundlagen, Prozessen und Ergebnissen der Arbeit der Beratungsstelle und zieht daraus anwendungsorientierte Konsequenzen. Im Kern bestätigt er die bisherige Ausrichtung der Arbeit. Methodisch verfolgt die Untersuchung einen Methodenmix, der partizipationsorientiert ein prozessbegleitendes rekonstruktiv-qualitatives Design formativer Evaluation mit einer Vielzahl von Instrumenten einsetzt, aber auch Daten aus Fallstatistiken einbezieht. In ähnlicher Weise geht die Evaluation der Hamburger Beratungsstelle „legato“ vor, die seit einigen Monaten vorliegt und schwerpunktmäßig eine Struktur-, Konzept- und Prozessevaluation vornimmt.<sup>45</sup> Sie kommt im Wesentlichen u. a. zu dem Ergebnis, dass der systemisch und sozialräumlich angelegte und dabei nicht unmittelbar auf die Erreichung von Indexpersonen zielende Ansatz der evaluierten Praxis aussichtsreich erscheint und weitergeführt werden sollte. Weitere vergleichbare Evaluationen sind zwar abgeschlossen, aber noch nicht veröffentlicht, zur Veröffentlichung nicht vorgesehen oder noch in Arbeit. Detailliertere, umfangreichere, belastbarere und stärker verallgemeinerbare Erkenntnisse sind insofern erst in den nächsten Jahren zu erwarten.

40 Ebd., 10.

41 Ebd., 6.

42 Ebd., 1.

43 Vgl. Uhlmann 2017.

44 Vgl. Möller/Neuscheler 2018.

45 Vgl. Schuhmacher o. J.

Für den internationalen Bereich sieht die Gesamtbilanz nicht wesentlich besser aus. So kommt eine Metaevaluation von Feddes/Gallucci<sup>46</sup> zur Radikalisierungsforschung zwar auf ein Sample von insgesamt 135 durchgeführten Evaluationsstudien zwischen 1990 und Juli 2014 – die meisten davon ab 2008 publiziert –, muss aber feststellen, dass davon nur 12 % mit Primärdaten arbeiteten. Während in 50 der einbezogenen 135 Studien (= 37 %) die eingesetzten Methoden nicht einmal spezifiziert wurden, nutzte der Löwenanteil von ihnen nur eine Methode und gab es nur bei 3 % einen Multi-Methoden-Einsatz. In der Konsequenz werden nur wenige Evaluationsanlagen als vielversprechend eingeschätzt;<sup>47</sup> diese gerade deshalb, weil sie auf Veränderungstheorien gestützte Untersuchungen darstellen und Methoden-Mixes nutzen.

Die Gründe für diese insgesamt betrachtet geringe Elaboration von Evaluationsprojekten und -berichten sind vielfältig, aber wohl auch darauf zurückzuführen, dass: (1.) das Arbeitsfeld – zumindest gilt dies für die sog. Islamismusprävention – noch relativ jung ist, (2.) thematische Zuordnungsprobleme von Angeboten zu Rubriken wie „Radikalisierungs-“ und „Extremismus-“ oder eben auch „Gewalt-“, „Kriminalitäts-“ und „Vorurteilsprävention“ vorliegen, (3.) Intransparenz aufgrund von Trägerinteressen, wenig ausgefeilten Praxisstandards, Sicherheitsbedenken und Datenschutz gegeben ist, (4.) grundlegende Begriffe – wie „Extremismus“ und „Radikalisierung“ – noch nicht hinreichend konsentiert sind, (5.) damit zusammenhängend Übereinkünfte über realistische Zielsetzungen – Abbau von extremistischer Gewalt und Straffälligkeit oder auch in jedem Fall möglichst vollständige Reduktion von extremistischen Orientierungen – nicht vorhanden sind, (6.) die Rekrutierung von radikalisierten Personen oder von im Ausstieg Begriffenen schwierig, die von Abbrecherinnen und Abbrechern oder Rückfälligen nahezu unmöglich ist, (7.) vorhandene theoretische Konzepte und die sie erstellenden Personen leider oftmals relativ viel Distanz zu den Gegebenheiten und zum Personal der Praxis aufweisen und (8.) aufgrund kaum entwickelter Evaluationskultur in diesen Bereichen Kriterien für Evaluationsstandards noch wenig diskutiert, geschweige denn zugrunde gelegt werden. Genau diesen ist aber verstärkt Beachtung zu schenken.

## Wirkungskontrolle und Qualitätsstandards

Wird im internationalen Raum von Standard gesprochen, so ist damit im Regelfall, angelehnt an die Definition der „International Organization for Standardization“, ein „document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose“<sup>48</sup> gemeint.

Beziehen wir dieses sehr allgemeine Verständnis auf Evaluationsprojekte, dann lässt es sich im Sinne der Deutschen Gesellschaft für Evaluation (DeGEval) konkretisieren. Demnach sind Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit die bedeutsamsten Gütekriterien für Evaluationsvorhaben; sie sind in der gegenwärtigen Fassung in weitere 21 Unterkriterien gegliedert.<sup>49</sup>

<sup>46</sup> Feddes/Gallucci 2015.

<sup>47</sup> Es handelt sich um Möller et al. 2015 im Bereich Rechtsextremismus- und Schuurman/Bakker 2015 im Bereich Islamismusprävention.

<sup>48</sup> ISO, the International Organization for Standardization.

<sup>49</sup> Siehe hierzu und kurz zu ihrer historischen Entwicklung, orientiert an den älteren amerikanischen „Standards for Educational Evaluation“.

Speziell für die Evaluation von Extremismusprävention und der in ihrem Rahmen verfolgten Programme, Projekte und Maßnahmen hat die Diskussion über Qualitätsstandards erst in jüngster Zeit begonnen. Der am prägnantesten ausgearbeitete Entwurf dazu resultiert aus Überlegungen beim „Nationalen Zentrum Kriminalprävention“ (NZK), einem Anfang 2016 gegründeten und beim „Deutschen Forum Kriminalprävention“ angesiedelten Fachdienst.

Konzentriert auf die evaluatorische Feststellung des Wirksamkeitsgrades von Prävention wird hier die Strategie verfolgt, Angebote qualitativ danach einzustufen, wie gut sie evaluiert wurden. Die Qualität von Evaluationen wiederum wird an einem Punktesystem festgemacht, das unter der Bezeichnung WESPE wissenschaftliche Erkenntnisse zu Sicherheits- und Präventionsmaßnahmen durch Evaluation sortiert und in Rangfolge bringt. Im Ergebnis entsteht angelehnt an ein Bewertungsschema von Beelmann und Hercher<sup>50</sup> ein „Evidenz-Index“<sup>51</sup>, in dem Studien bis zu 100 Punkte für wissenschaftliche Exzellenz erreichen können. Entlang von neun „Bewertungsmerkmalen“ werden dabei Evaluationsstudien bestimmte, jeweils mit Punkten belegte Stufen zugeschrieben, die als erreicht gelten – oder eben nicht.

Die neun Bewertungsmerkmale beziehen sich im Einzelnen: (1.) auf die Klarheit und Überprüfbarkeit, mit der die jeweiligen Ziele der Studie beschrieben werden, (2.) auf die Geeignetheit des methodischen Zugangs, (3.) auf die Benennung von theoretischen Annahmen und deren Bezug auf empirische Überprüfbarkeit, (4.) die interne Validität (= Gültigkeit), die durch die Anzahl von Erhebungszeitpunkten, deren zeitliche Lagerung (Ex-ante und Ex-post) und das Vorhandensein einer zufällig zusammengestellten (randomisierten) Vergleichsgruppe definiert wird, (5.) die externe Validität, die aufgrund des Grades statistischer Repräsentativität zugesprochen wird, (6.) die Messvalidität, die sich auf die Aussagekraft der Indikatoren zur Messung von Effektgrößen bezieht, (7.) die Qualität der Datenauswertung und ihrer Dokumentation, (8.) die Angemessenheit und Reflexivität der Ergebnisinterpretation und (9.) die Vermeidung möglicher Interessenkonflikte, am besten durch die Beauftragung eines unabhängigen, externen Evaluationsteams.

Nach diesem Index besteht der lange in der Evaluationsforschung als „Goldstandard“ betrachtete Lösungsweg für das Problem, registrierte Effekte eindeutig auf spezifische Maßnahmen zurückführen und jegliche anderen Ursachen dafür ausschließen zu können, vorzugsweise darin, sog. „randomisierte Kontrollgruppen-Designs“ (randomized controlled trials; RCTs) zu entwerfen, also neben der/den das Angebot nutzenden eigentlichen Probandengruppe/n („Interventionsgruppe“) noch eine (oder mehrere) ebenfalls möglichst zufällig zusammengestellte, aber doch in Kernmerkmalen (wie z. B. Alters- und Geschlechterzusammensetzung) vergleichbare Kontrollgruppe/n zu installieren, die das Angebot nicht durchläuft/durchlaufen, um dann durch Vergleiche Rückschlüsse auf Effekte ziehen zu können.

Abstrakt kann ein solches Vorgehen zu Zwecken des Wirkungsnachweises nachvollziehbar und überzeugend erscheinen. Allerdings sind gerade im Bereich der Extremismusprävention komplizierte und auch sehr schwerwiegende Probleme mit ihm verbunden. Zum einen stellen sich Fragen der experimentellen Praktikabilität: Lassen sich randomisierte Zuord-

<sup>50</sup> Beelmann/Hercher 2006.

<sup>51</sup> Evidenz bezieht sich im Rahmen von Evaluationsforschung – anders als oftmals im Alltagssprachgebrauch – nicht auf das selbstverständlich und offensichtlich Erscheinende, sondern hat hier die Bedeutung „Beweis“, „Nachweis“, „Beleg“. Evidenzorientierung bzw. -basierung zielt also darauf ab, Entscheidungen bestmöglich informiert und auf robuster wissenschaftlicher Grundlage treffen zu können.

nungen von Interventions- und Kontrollgruppen innerhalb realer Settings (z. B. bei Untersuchungen mit Schulklassen oder Untersuchungen mit Besucherinnen und Besuchern von Präventionsveranstaltungen) überhaupt praktisch und sinnfällig herstellen? Kann man hier etwa auslösen, wer in welchen Untersuchungsgruppentyp hineinkommt? Und wenn man auf die Randomisierung der Interventionsgruppe (deshalb) verzichtet, ist es dann möglich, eine Kontrollgruppe zusammenzustellen, die die Vielzahl und Beschaffenheit von individuellen Voraussetzungen und externen Einflüssen, die für die Mitglieder der Interventionsgruppe gegeben sind, in dem Maße teilt, dass sinnvolle Vergleiche zustande kommen können? Hinzu kommt: Wenn dies schon für die Dauer der Angebotsdurchführung kaum möglich ist, wie können dann über einen längeren Zeitraum hinweg, also etwa auch noch für Monate oder Jahre nach der Durchführung zu Zwecken der Prüfung von längerfristigen Effekten, relevante Variablen kontrolliert werden?

Viel bedeutsamer ist aber zum anderen ein weiteres Gegenargument – wie auch NZK-Mitarbeitende selbst einräumen:<sup>52</sup> Ethisch ist es in höchstem Maße bedenklich, Menschen, die man für präventionsbedürftig hält, Präventionsangebote vorzuenthalten, nur um sie als Vergleichsgruppe nutzen zu können. Verschärft gilt dies für die Bereiche der sekundären bzw. selektiven und tertiären Prävention. Distanzierungsoffenen und Ausstiegswilligen nur aus kausalitätsanalytischen Gründen entsprechende Hilfen nicht zukommen zu lassen und noch dazu dadurch Sicherheitsrisiken für durch sie Gefährdete einzugehen, ist moralisch nicht vertretbar, grenzt an den rücksichtslosen Einsatz von Menschenversuchen für wissenschaftliche Zwecke und läuft dem Kernanliegen von Prävention, nämlich weitestmöglich über die vorbeugende Behandlung von (potenziellen) Täterinnen und Tätern Opferschutz zu betreiben, entgegen. Das Vorgehen widerspricht zudem geltendem Recht zumindest dort, wo Bereiche der Kinder- und Jugendhilfe berührt sind. Denn § 1 SGB VIII legt fest, dass „jeder junge Mensch (...) ein Recht auf Förderung seiner Entwicklung und auf Erziehung zu einer eigenverantwortlichen und gemeinschaftsfähigen Persönlichkeit (hat)“.

Zudem ist zu bedenken: Pädagogische und soziale Arbeit sind nur sehr bedingt standardisierbar und manualisierbar. Ihr Setting ist vielfach – etwa in der Offenen und aufsuchenden Kinder- und Jugendarbeit – seitens der Fachkräfte faktisch wenig planbar oder bis in alle Einzelheiten strukturierbar. Ihre Konstellationen sind im Allgemeinen divers und dynamisch. Adressatinnen bzw. Adressaten und Fachkräfte konstruieren sie als Ko-Produzentinnen und -produzenten. Daher erscheint aus Praxissicht ein bloßer Transfer von (scheinbar noch so gelungenen) „Treatments“ auf ihre jeweilig konkret gegebenen Zusammenhänge als ein technologisches Missverständnis und als eine Folge funktionalistischer Übertragbarkeits-Logik. Diese wiederum scheint das Resultat von Defiziten an Praxisnähe und Ignoranz von Praxiserfahrungen zu sein, wenn nicht gar einer Geringschätzung von Praxiswissen gegenüber vermeintlich höherwertigem wissenschaftlichem Wissen.<sup>53</sup>

Darüber hinaus erscheint es höchst unangemessen, qualitative und quantitative Untersuchungsanlagen „über einen Kamm zu scheren“, indem man für beide ein und dasselbe Begutachtungsschema zugrunde legt. Das im Evidenzindex stark gewichtete Bewertungskriterium der statistischen Repräsentativität und überhaupt die positive Beurteilung des Einsatzes statistischer Verfahren – etwa bei der Vergleichsgruppen-Konstruktion, die die interne Validität entscheidend bestimmt – können an qualitativ-rekonstruktive Studien gar nicht angelegt werden, weil sie weder Vermessungen vornehmen, noch Repräsentativität

<sup>52</sup> Vgl. Ullrich et al. 2018.

<sup>53</sup> Vgl. auch Bellmann/Müller 2011, bes. 24 ff.; Herzog 2011.

erreichen wollen und können, sondern stattdessen auf theoretische Repräsentanz setzen. Gegenüber qualitativen Studien enthält der Index mithin einen Abwertungsautomatismus, der sachlich inadäquat ist.

Für ihre Beurteilung sind daher spezifische Gütekriterien qualitativer Forschung heranzuziehen: (1.) die Relevanz von Fragestellungen und theoretischen Aussagen, (2.) die intersubjektive Nachvollziehbarkeit, (3.) die Indikation des Forschungsprozesses und der Bewertungskriterien, (4.) die empirische Verankerung der Theoriebildung, (5.) die Begrenzung der Reichweite von Aussagen, (6.) die Reflexion der Subjektivität der forschenden Person, (7.) die Kohärenz von Aussagen und (8.) die Problemorientierung als Praxisorientierung.<sup>54</sup>

Speziell für die qualitative Evaluationsforschung sind sie um weitere Prüfsteine zu ergänzen:<sup>55</sup> (1.) die Vertrauenswürdigkeit hinsichtlich Versprechungen, Erwartungen und Aufwand, (2.) die Durchführbarkeit im Sinne von Datenzugänglichkeit und Einhaltung forschungsethischer Standards, (3.) die Angemessenheit von evaluatorischen Belastungen im Verhältnis zum erhofften Nutzen, (4.) der Nutzen im Sinne der Evaluationsziele und aus der Perspektive der relevanten Stakeholder, nicht zuletzt auch der Adressatinnen und Adressaten und (5.) die Genauigkeit im wissenschaftlichen Vorgehen. Evidenz ergibt sich in diesem Kontext realistischer Weise eher dann, wenn die Beziehungen zwischen dem Kontext des Untersuchungsgegenstands, den Wirkmechanismen und den Ergebnissen erhellt und dabei die Interessen und emotionalen Befindlichkeiten der Akteure ebenso wie äußere situativ auftretende Faktoren berücksichtigt werden.<sup>56</sup> Für Evidenzsynthesen, die auf qualitative Studien bezogen sind, können darüber hinaus die CERQual-Kriterien herangezogen werden: (1.) die kritische Bewertung der methodologisch bedingten Aussagebegrenzungen einbezogener Studien, (2.) die Relevanz, die der jeweiligen Studie in Hinsicht auf die Fragestellung des Reviews beikommt, (3.) die Kohärenz der Reviewbefunde und (4.) die Adäquanz der Daten, die die Reviewbefunde stützen.<sup>57</sup>

## Wirkungsorientierung und Wirkungsmodellierung

Mehr Alltagstauglichkeit als konventionelle evidenzorientierte RCT-Designs, die auf nachträgliche Wirkungskontrolle zentriert sind, versprechen Ansätze der Wirkungsorientierung, die neben Wirkungsfeststellung und -identifizierung<sup>58</sup> das innovations-, optimierungs- und lösungsbezogene Potenzial von Evaluation zur Geltung bringen.

Wirkungsorientierte Evaluation nach der Gruppe um Beywl<sup>59</sup> verläuft idealerweise in einem fünfstufigen Verfahren. Auf Stufe 1 wird das Verhältnis zwischen Input und gezähltem Output im Sinne einer Kosten-Leistungs-Rechnung bestimmt. Stufe 2 unternimmt die Messung und Bewertung der Zielerreichung, also die Outcome-Evaluation. Auf Stufe 3 wird das Verhältnis zwischen Incomes und Outcomes gemessen und bewertet, also der Nutzenzuwachs, den das Angebot für die Adressatengruppierungen erbracht hat. Die Feststellung und Bewertung der Wirksamkeit des Angebots mittels Wirkungseinschätzungen Externer zu den Outcomes, Untersuchungen zur Funktionslogik des Angebots mittels Wirkungsmodellierung

<sup>54</sup> Vgl. Steinke 2013.

<sup>55</sup> Vgl. auch ausführlicher Patton 2003.

<sup>56</sup> Vgl. auch Pawson/Tilley 1997.

<sup>57</sup> Vgl. Lewin et al. 2018.

<sup>58</sup> Vgl. zu diesen Begriffen Beywl/Niestroj 2009, insb. 106 ff.

<sup>59</sup> Vgl. 2006.



dellierung und/oder statistisch abgesicherte Messungen mit Interventions- und Kontroll- bzw. Vergleichsgruppen passieren auf Stufe 4. Stufe 5 ist der Bestimmung und Bewertung der Wirtschaftlichkeit des Angebots gewidmet.

Zugleich macht Beywl jedoch deutlich, dass ein solcher Stufenaufbau in den realen Beschaffenheiten von Feldern Sozialer Arbeit – z. B. explizit in der Kinder- und Jugendhilfe, die auch einen höchst bedeutsamen Bereich von Extremismusprävention darstellt – in seiner Gänze kaum zu realisieren ist. Designs mit Interventions- und Kontroll- bzw. Vergleichsgruppen eignen sich demnach nur für Einzelfälle der evaluativen Bildungsforschung, nämlich dort, wo „stark standardisierte, fokussierte Programme (...) über längere Zeit stabil angewandt werden“, wie dies etwa beim Mathematikunterricht angenommen werden könnte, „da dort die Outcomes mit breitem Fachkonsens festgelegt werden können“.<sup>60</sup> Das von Dynamik und starken Wertekonflikten bestimmte und zudem noch ressourcenschwache Gebiet Sozialer Arbeit dagegen braucht demgemäß Evaluationen, die in Abstimmung und Kooperation mit Fachleuten aus der Praxis „mindestens auf Stufe 2 der Wirkungsorientierung angelegt werden“, nach Möglichkeit auch Stufe 3 erreichen und auf Stufe 4 Wirkungsfeststellungen mittels Wirkungseinschätzung oder Wirkungsmodellierung vornehmen.<sup>61</sup>

Für wirkungsorientierte Evaluation plädiert z. B. auch der Stockmannsche Ansatz,<sup>62</sup> der allerdings nicht speziell auf soziale und pädagogische Arbeit oder Extremismusprävention zugeschnitten ist, sondern ausdrücklich ein weites Anwendungsspektrum offeriert. Er kann auf experimentelle und quasi-experimentelle Elemente verzichten und sucht dies durch Multiperspektivität und Multi-Methodeneinsatz zu kompensieren, wobei vor allem auch auf die Triangulation quantitativer und qualitativer Methodiken gesetzt wird. Kern der Konzeption sind fünf Analysebereiche: Erstens werden die Angebotskonzeption mit ihren Zielsetzungen, Wirkannahmen und zu deren Realisierung bereitgestellten Ressourcen analysiert, die Zielgruppen identifiziert sowie die Felder beschrieben, in denen sich die beabsichtigten Wirkungen entfalten sollen. Zweitens wird der gesamte Planungs- und Durchführungsprozess sowie seine Steuerung analysiert, um die kausalen Verkettungen einzelner Phasen miteinander in den Blick nehmen zu können. Drittens werden die zentralen Parameter der durchführenden Organisation untersucht, wie personelle, räumliche, organisatorische, technische und kommunikative Strukturen und Ressourcen, damit die Leistungsfähigkeit interner Wirkungsbedingungen bestimmbar wird. Viertens werden die intendierten und nicht-intendierten Brutto- und Netto-Wirkungen auf Seiten der Adressatinnen und Adressaten, bei Mitgliedern sozialer Gruppen, die mit ihnen in Kontakt sind und innerhalb der gesellschaftlichen Felder, in denen die Aktivitäten des Angebots stattfinden, identifiziert, bewertet und einem Soll-Ist-Vergleich unterzogen. Zum fünften wird die Nachhaltigkeit der registrierten Wirkungen in ihren sozialen, ökonomischen und ökologischen Aspekten untersucht. Der Ansatz erfordert eine partizipative Einbindung von Evaluierten, die auch dadurch realisiert wird, dass in Workshops vorab Abstimmungen über die Evaluationsziele und den Evaluationsaufbau erfolgen, Zwischenergebnisse zur Verfügung gestellt und Ergebnisbewertungen auch unter Einbezug von Evaluierten vorgenommen werden.

<sup>60</sup> Ebd., 41.

<sup>61</sup> Ebd., 42; ähnliche Schlussfolgerungen bei Bischoff u. a. 2018.

<sup>62</sup> Vgl. Stockmann 2006.

## Wirkungsmodellierung

Um Analyse-, Konzept-, Struktur-, Prozess- und Ergebnisqualität einfangen, aber auch so weit wie möglich (dabei) Wirkungen und Wirkungsweisen umfassend und differenziert evaluativ feststellen zu können, bedarf es der systematisch dokumentierenden Beschreibung und ggf. auch Visualisierung der angestrebten Verbindungen zwischen dem beabsichtigten und auch dem realisierten Präventionsgeschehen einerseits und den Wirkungen, die damit erzielt werden (sollen), andererseits. Im Rahmen von Angebotstheorien (evaluatorisch: „Programmtheorien“) werden Wirkungspfade und -mechanismen erläutert und begründet und auf diese Weise Interventionslogiken dargelegt. In ihnen drückt sich aus, welche vor und während der Umsetzung des Angebots vermuteten oder anschließend registrierten Wirkungen auf welche Wirkungsbedingungen zurückgeführt werden. Sie stellen sich, je nachdem, wer sie anstellt und oft auch je nachdem, zu welchem Zeitpunkt sie vorgebracht werden, zumeist durchaus unterschiedlich dar. Insofern erscheint es angebracht, zu ihrer Erfassung auf ein Vorgehen zu setzen, das sowohl multiperspektivisch als auch prozesshaft angelegt ist. Die Konstruktion Logischer Modelle bietet dafür ein gut handhabbares, z. B. auch von Beywl<sup>63</sup> favorisiertes und bereits seit mehreren Jahren in der Evaluation von Extremismus- sowie Kriminalitäts- und Gewaltprävention auf Maßnahmen-, Projekte- wie Programmebene bewährtes Verfahren.<sup>64</sup>

Bei Logischen Modellen handelt es sich um Abbildungen der Deutungs- und Funktionslogiken von Projekten und damit u. a. auch der Wirkungsannahmen der in ihnen Mitarbeitenden – seien sie explizit vorhanden

oder als „tacit knowledge“ implizite, atheoretische und pragmatische „konjunktive Wissensbestände“ im Sinne Karl Mannheims.<sup>65</sup>

Sie stellen – in Abb. 1 angelehnt an den Programmbaum von Univation,<sup>66</sup> ihn jedoch erweiternd – dar, in welcher Weise Verbindungen zwischen Ausgangsbedingungen, konzeptionellen Planungen, ihren realisierten Umsetzungen und Ergebnissen konstruiert werden und wie verlaufsbezogenen Rahmenbedingungen und die Professionalität der Fachkräfte darauf einwirken (vgl. Abb. 1). In diesem Sinne beanspruchen sie, sowohl institutionell als auch subjektiv repräsentierte Theorien mit nicht nur deskriptiver, sondern auch as-riptiver und explikativer Funktion

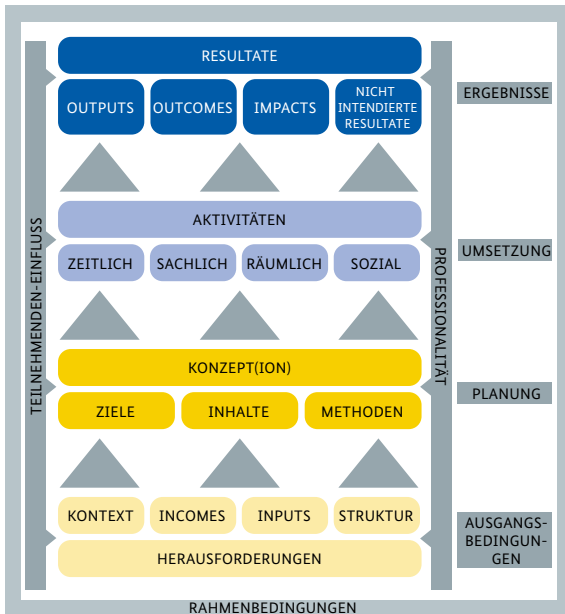


Abb. 1: Logisches Modell; eigene Darstellung

63 Vgl. z. B. Beywl 2006.

64 Vgl. z. B. Klingelhöfer 2007; Haubrich 2009; Lüders 2010; Armbrorst u. a. 2018; Yngborn/Hoops 2018; Möller/Neuscheler 2016, 2017, 2018.

65 Vgl. Mannheim 1980; Bohnsack 2011.

66 Univation, Institut für Evaluation.

aufzeigen zu können und sind deshalb mit sog. „Programmtheorien“<sup>67</sup> vergleichbar, auch wenn sie im Unterschied zu ihnen nicht zwingend eine forschungsbasierte Begründung besitzen.<sup>68</sup>

Bezogen auf solche Logiken können auch andere Stakeholder nach ihren Wirkannahmen und Wirkungseinschätzungen befragt werden – dies idealerweise ebenfalls vor, während und nach der Angebotsdurchführung. Dadurch wird ein Prozesse widerspiegelndes Spektrum verschiedener, an manchen Stellen womöglich aber auch kongruenter Sichtweisen auf die Elemente des Angebots und deren Zusammenspiel verfügbar. In Ergänzung der eigenen Beobachtungen und Analysen des Evaluationsteams können weitere Perspektiven die Aussagen über den Evaluationsgegenstand bereichern und zugleich deren Validität befördern.

Freilich bedarf es dazu eines konsequent partizipatorischen Vorgehens mit einer engen und vertrauensvollen Wissenschaft-Praxis-Kooperation, in dem Interaktion mit allen Beteiligten, Dialogizität und Diskursivität „großgeschrieben“ und Workshops zur Diskussion von geplanten und umgesetzten Zielen, Inhalten und Methodiken sowie zur Auswertung von Ergebnissen veranstaltet werden. Daneben sollten vielfältige Rückkoppelungen erfolgen und insgesamt ein responsives Design ermöglicht werden, das neu auftauchende Bedarfe sowie sonstige Anforderungen aufgreifen kann. Evaluation dieser Kontur ist über weite Strecken eine Form wissenschaftlicher Begleitung, erschöpft sich jedoch nicht darin: Sie erfüllt auch i. e. S. begleitungsunabhängige Aufgaben wie etwa Erstellung eines Forschungsdesigns, Konzeptionsanalyse, Ergebnisbewertung und Berichtlegungen an den Auftraggeber. Sie kann zwar seitens der in sie einbezogenen Fachkräfte und sonstigen Angebotsdurchführenden als Beratung empfunden werden, geht allerdings nicht in einer solchen Funktion auf. Sie stellt bspw. auch wissenschaftliche Nutzenerwartungen an sich und/oder formuliert Schlussfolgerungen für angebotsexternes fachliches Handeln sowie (präventions-)politische Folgerungen aus ihren Resultaten. Sie ist (z. T. deshalb) auch keine Supervision, weil nicht das individuelle oder kollektive Erleben der Berufsrolle im Mittelpunkt steht, sondern die an sachlichen, sozialen, zeitlichen und räumlichen Rahmungen ausgerichtete, angebotsbezogene Logik von professionellen Deutungen und Handlungen.

Förderlich ist eine Auffassung von Evaluation, die sich weniger der Leitidee von Wirkungskontrolle im Sinne einer unabhängig von den Evaluierten vorzunehmenden Kosten-Nutzen-Abwägung verschreibt. Stattdessen sieht sie sich vielmehr dem Motiv verpflichtet, sozialpolitische Innovationen zu befördern<sup>69</sup> und verfolgt etwa im Sinne der Strategie einer „realist evaluation“<sup>70</sup> die Frage: „Was funktioniert wieso für welche Adressatinnen und Adressaten unter welchen Bedingungen?“. Dabei kann sie sich an dem Verständnis und der Praxis gestaltungsorientierter Transformationsforschung orientieren.

## Skizze: Evaluation als gestaltungsorientierte Transformationsforschung

Ausgangspunkt gestaltungsorientierter Transformationsforschung<sup>71</sup> ist die in den letzten Jahrzehnten beobachtbare Veränderung der Modi der Wissen(schaft)produktion und die Notwendigkeit, sie auch in der Evaluationsforschung zu berücksichtigen. Dieser Wandel

<sup>67</sup> Vgl. grundlegend Chen 2004.

<sup>68</sup> Vgl. näher dazu auch Haubrich 2009.

<sup>69</sup> Vgl. Haubrich 2009.

<sup>70</sup> Vgl. Pawson/Tilley 1997, 2004.

<sup>71</sup> Vgl. Möller 2012; ders. et al. 2017.

drückt sich darin aus, dass für die scientific community zunehmend auch Adressierte und Akteure wichtig werden, die außerhalb der klassischen Wissenschaftssphäre arbeiten. Praxisakteure und -institutionen sowie die Arenen von Politik, Öffentlichkeit und Internet mit der Open-Science-Bewegung (Open Access, Open Data, Open Source u. a.) erhalten für die Sozialforschung in dem Maße gesteigerten Stellenwert, wie sie sich zunehmend – gerade in der Evaluationsforschung – mit der Erwartung von mehr Praxisbezug konfrontiert sieht bzw. sich selbst als anwendungsorientierte, sozial innovative, gestaltungsorientierte<sup>72</sup> und/oder öffentliche Wissenschaft – etwa im Burawoyschen Sinne der „öffentlichen Soziologie“<sup>73</sup> – versteht.

Die Herausforderung, die mit einer entsprechenden Schwerpunktverlagerung der Wissens-erzeugung und -vermittlung einhergeht, betrifft nicht nur die Überwindung von „Transfer“-Problemen wissenschaftlicher Erkenntnisse im Allgemeinen und wissenschaftlicher Evaluationserkenntnisse im Besonderen. Sie besteht vielmehr in der Notwendigkeit der transformativen Neugestaltung von Prozessen der Wissensproduktion und -verwendung; mehr noch: auch von fachlichen und transdisziplinären Diskursen über normative und ethische Grundlagen. Der Einbezug von Akteuren, deren berufliche und/oder zivilgesellschaftliche Tätigkeitsfelder außerhalb von Hochschulen und Forschungsinstitutionen liegen, spielt dabei eine Schlüsselrolle.

Die These der zunehmenden Bedeutung des sogenannten „mode 2“ der wissenschaftlichen Wissensproduktion<sup>74</sup> verweist auf stattgefundene und sich weiter vollziehende Veränderungen in den gesellschaftlichen Erwartungshaltungen an die Sphäre der Wissenschaft. In mode 1 findet danach Wissenserzeugung in einem vom Rest der Gesellschaft getrennten Subsystem der Wissenschaft, damit auch in abgeschlossenen Institutionen und Räumen – eben in Universitäten oder in speziellen Forschungseinrichtungen – statt, läuft differenziert nach Fachspezifiken ab, und ihre Erkenntnisprodukte werden in homogenen Kreisen von Fachpublika per Fachveröffentlichung, Fachkonferenzen etc. weitergegeben. In mode 2 dagegen werden die institutionellen, räumlichen und disziplinären Grenzen gesprengt. Es entstehen offene Systeme der Wissensproduktion, die eher netzwerkartig als hierarchisch strukturiert, eher transdisziplinär geöffnet als disziplinär beschränkt sind, eher global und leicht zugänglich über Informations- und Kommunikationstechniken zur Verfügung stehen als in Fachbibliotheken eingesperrt sind, eher heterogene als homogene Anlagen besitzen und in einer interdependenten Mischung von Zusammenspiel und Konkurrenz unterschiedlicher Quellen der Wissensproduktion zustande kommen. Diese Entwicklungen führen dazu, dass: (1) die Relevanz und Qualität von Wissensinhalten immer weniger durch Wissenschaftlerinnen und Wissenschaftler selbst als durch Nutzende des geöffneten Wissensmarktes bestimmt wird, (2) Wissenserzeugung selbst immer stärker im Kontext von Anwendungen und in Verbindung und in Rückkopplung mit Verwendenden erfolgt, ja erfolgen muss, (3) Interaktion und Kommunikation zwischen den unterschiedlichen Wissensträgerinnen und -trägern sowie mit Verwendenden zunehmende Bedeutung erhält, (4) Wissenschaft in stärkerem Maße gesellschaftlich rechenschaftspflichtig wird, sich an Medien orientiert und mehr als früher politische Zielsetzungen und soziale Werte zu berücksichtigen hat, (5) sich sowohl durch wechselseitige Informierungen und Beeinflussungen als auch durch persönliche Mobilität wissenschaftlich Ausgebildeter zwischen den Sphären der Wissenschaft und

72 Vgl. Howaldt/Schwarz 2010.

73 Vgl. 2012.

74 Vgl. Gibbons et al. 1994; Bender 2001; Novotny et al. 2001.

der Praxis die Grenzen zwischen Wissensträgern hier und Verwendern da teilweise bis zur Unkenntlichkeit verwischen und (6) sich dabei ein Korridor zu Projekten gemeinsamer Gestaltung von Innovationen öffnet.

Die genannten Entwicklungen fordern verstärkt praxisbezogene Forschung ein, die nomoprägnantes Wissen generiert, also etwa Handlungsheuristiken, die ihre Nützlichkeit bei der Bearbeitung praktischer Probleme erweisen oder Deutungspluralitäten für Praxisphänomene bieten. Sie muss dabei mit Praxis Kooperationszusammenhänge eingehen, in denen beide Sphären sich begegnen und wechselseitig befruchten können, ohne ihre spezifischen Eigenlogiken negieren zu müssen. Es bedarf sozusagen eines „Dritten Raums“, in dem Wissenschaft und Praxis sich begegnen, sich bidirektional austauschen und sich wechselseitig zueinander in Beziehung setzen, ohne ihre jeweilige Eigenständigkeit einzubüßen.

Die zentrale Funktion dieses Raums ist nicht dadurch bestimmt, dass sich hier Wissens„transfer“ im klassischen Sinne, nur jetzt als wechselseitiger Austausch, vollzieht. Sie liegt vielmehr in der Rahmung für eine „Transformation“ des hier repräsentierten Bestandes an Wissen, Vorstellungen und Wertsetzungen der beteiligten Parteien.<sup>75</sup> Es geht also um eine Umformung, eine „nicht-identische Reproduktion“<sup>76</sup> wissenschaftlichen und praktischen Wissens und eine Neukontextuierung von Vorstellungen und Werthaltungen, die aus den jeweiligen Bezugssystemen mitgebracht werden. Der „Dritte Raum“ dient nicht nur wechselseitiger Information „auf gleicher Augenhöhe“ und „im Verhältnis der Komplementarität“.<sup>77</sup> Über seine Atmosphäre „wechselseitiger Resonanz“<sup>78</sup> fungiert er auch als ein Ort der Emergenz und bewussten Kreation von Forschungs- und Praxisideen in einem Prozess der kontrastierenden Reflexion, der Ideen- und Wissensproduktion als rekursive Angelegenheit konzipiert. Die intensive Diskussion Logischer Modelle ist geradezu dafür prädestiniert, die Zentralaktivität innerhalb eines solchen Dritten Raums und seiner Funktion zur Innovationskonstruktion zu bilden.

Eine solche Begegnungs-Sphäre kann freilich nur unter mindestens fünf Bedingungen ihre Funktionen erfüllen: (1.) die Bezugnahmemöglichkeit auf gemeinsame Interessen; (2.) die Relevanzeinschätzung von Seiten der Fachkräfte von Forschung generell und von konkreten Forschungsergebnissen für die tägliche Arbeit; (3.) die Glaub- und Vertrauenswürdigkeit der Forschenden für die in der Praxis Tätigen, die wiederum u. a. und vor allem auf der Kontinuität und Intensität des schon vorhandenen, auch informellen und persönlichen Kontakts beruht und aus Erfahrungen von Vertrauenswürdigkeit der Forschungspersönlichkeiten in der Vergangenheit resultiert; (4.) das Finden einer gemeinsamen und von Symmetrie gekennzeichneten sprachlichen und habituellen Verständigungsbasis und (5.) die Beteiligung der Nutzer und Nutzerinnen am Forschungsprozess selbst.

Gerade der letztgenannte Aspekt setzt weitere Gelingensbedingungen voraus. Zu ihnen gehören neben den schon genannten kooperationsförderlichen Kriterien u. a.: (1) transparente Regelungen für den Ein- und Ausschluss von Personen bzw. Personengruppen, (2) funktionierende Prozesssteuerung, (3) Ergebnisoffenheit, (4) Übereinkunft über den Umgang mit dem gewonnenen Wissen und bei der Definition von Erfolg sowie (5) vorhandene Umsetzungschancen.

<sup>75</sup> Vgl. auch Dewe 2005.

<sup>76</sup> Luhmann 2005, 379 ff.

<sup>77</sup> Dewe et al. 1992, 79.

<sup>78</sup> Ebd.

## Abschließende Empfehlungen

In Konsequenz des bislang Erörterten ist aktuell und mindestens für die nähere Zukunft zunächst die „Dresdner Erklärung des 23. Deutschen Präventionstages“ und seiner Veranstaltungspartner, darunter das BMFSFJ, vom Juni 2018 orientierungsgebend, wonach „ein Schwerpunkt auf die wissenschaftliche Evaluation von Präventionsmaßnahmen und Programmen zu legen“ ist – ein Postulat, das, wie Baier<sup>79</sup> richtig anmerkt, sowohl Bereitschaft von Präventionsakteuren voraussetzt, „Extraaufwände bzgl. der Datengewinnung“ einzukalkulieren, als auch die „Bereitschaft der diese Maßnahmen finanzierenden Stellen“ einschließt, verstärkt „unabhängige Evaluationen als einen Bestandteil einer Projektdurchführung mit zu fördern“.

Bedeutsam ist es, dass bei der Auftragsvergabe nicht mit jener fatalen, praxisfernen und daher realitätsfremden Einseitigkeit, wie dies das Federal Department of Education in den USA (2005, 3586) gemacht hat, „random assignment and quasiexperimentell designs“ als „most rigorous methods“ präferiert werden. Vielmehr ist der Anwendungsorientierung von Evaluation zuträglich, wenn seitens des Personals wissenschaftlicher Evaluation Verständnis für die Anliegen und Handlungslogiken der Praxis aufgebracht wird und dieses konzeptionell Berücksichtigung findet.

Im Einzelnen ist für Evaluationsvorhaben im Bereich der Extremismusprävention zu empfehlen:

- Wenn Sie extremismuspräventive Angebote planen, kalkulieren Sie von vornherein organisatorisch, zeitlich und finanziell deren Evaluation mit ein und werden Sie sich darüber im Klaren, welche Erwartungen Sie damit verbinden (können). Dies gilt auch, falls Sie bereits Angebote durchführen und nunmehr deren Evaluation beabsichtigen. Geht es Ihnen (nur) um Wirkungsnachweise oder (auch) um Auskünfte über Konzept-, Prozess- und Strukturqualitäten sowie deren Funktionszusammenhänge mit feststellbaren Resultaten?
- Um sachgerecht entscheiden zu können, ob Selbstevaluation ausreichend ist oder eine externe wissenschaftliche Evaluation notwendig und möglich erscheint, empfiehlt es sich, sich als Praxisträger an das Personal von Hochschulen und Forschungseinrichtungen mit einschlägiger Expertise in anwendungsorientierter Forschung und Entwicklung sowie zugleich möglichst auch mit thematischen Kenntnissen und praxisnahen Erfahrungen, vorzugsweise in der Region, zu wenden und das Vorhaben damit ausführlich zu erörtern.
- Anwendungsorientierte Evaluation von Programmen, Projekten und Maßnahmen der Extremismusprävention versteht sich nicht als Controlling. Ihr Wert bemisst sich schon gar nicht danach, wie viel Einsparpotenzial sie entdeckt. Sie erkennt die Arbeit der Akteure an und sucht gemeinsam mit ihnen nach Verbesserungspotenzialen, um die selbstgesetzten Ziele in geeigneter, zielführender und effizienter Weise zu erreichen.
- Gelingende Evaluation hält nicht nur die o. g. Standards für diese Arbeit ein, sie lebt auch vom kontinuierlichen Austausch mit möglichst allen, die an dem evaluierten Angebot beteiligt sind: Träger, Beschäftigte, Klientinnen und Klienten, Bezugspersonen, Kooperationspartner. Sie würdigt das Praxis- und Alltagswissen dieser Gruppierungen in seiner

<sup>79</sup> Baier 2018, 72.

Vielfalt und bezieht es als wichtige Erkenntnisquellen in ihre Arbeit ein. Darüber hinaus erkennt sie die besondere Notwendigkeit des Schutzes von Vertrauensverhältnissen an und sucht durch das Angebot verlässlicher Zusammenarbeit eventuelle anfängliche Befürchtungen zu zerstreuen.

- Auf der Seite von Trägern und deren Beschäftigten setzt dies die Bereitschaft voraus, Evaluation als Teil eines eigenen Lern- und Entwicklungsprozesses zu verstehen und in ihn entsprechend zu investieren. In diesem Sinne sind nicht nur Erfolgsmeldungen, sondern auch kritische Hinweise wertvolle Impulse für die künftige Arbeit.
- Evaluation im Bereich der Extremismusprävention sollte sich nicht von dem Glauben leiten lassen, Wirkung in naturwissenschaftlicher Analogie bemessen und beziffern zu können. Der Anspruch, ihr in geeigneter Weise nachzuspüren, muss deshalb allerdings keineswegs aufgegeben werden. Dazu ist es dienlich, Wirkung intersubjektiv – aus der Perspektive möglichst aller Beteiligten an dem Angebot und der des Evaluationsteams – einzuschätzen.
- Einen überall einsetzbaren methodischen Königsweg gibt es nicht. Die Methoden müssen dem jeweiligen Angebot, den Evaluationszielen und den Untersuchungspersonen angemessen sein. Mit Bischoff<sup>80</sup> u. a. lässt sich formulieren: Gegenstandsangemessenheit ist der „neue“ Goldstandard“. Wo möglich und sinnvoll, ist ein Mix aus quantitativen und qualitativen Verfahren anzustreben.
- Die Verwendung von Evaluationsergebnissen und die Streuweite sowie Formen ihrer Distribution sollten von Anfang an mitbedacht werden. Ihr gestaltungsorientierendes Potenzial lässt sich nur dann faktisch umsetzen, wenn rechtzeitig die Spielräume für ggf. erforderliche neue Schwerpunktsetzungen oder optimierende Umsteuerungen ausgelotet werden.

---

80 Bischoff 2018, 153.

## Literatur

- Ahlheim, Klaus/Heger, Bardo (2006): Wirklichkeit und Wirkung politischer Erwachsenenbildung. Eine empirische Untersuchung in Nordrhein-Westfalen. Schwalbach/Ts., Wochenschau.
- Alkin, Marvin C./Christie, Christina A. (2004): An Evaluation Theory Tree. In: Alkin, M. C. (Ed.): Evaluation Roots. Tracing Theorists Views and Influences. Thousands Oaks, Sage, S. 12-66.
- Armborst, Andreas/Biene, Janusz/Coester, Marc/Milbradt, Björn/Nehlsen, Inga (2018): Evaluation in der Radikalisierungsprävention: Ansätze und Kontroversen. PRIF Report 11/2018. Report-Reihe Gesellschaft extrem. Frankfurt, Leibniz Institut Hessische Stiftung Friedens- und Konfliktforschung. <https://www.hsfk.de/publikationen/publikationssuche/publikation/evaluation-in-der-radikalisierungspraevention-ansaeetze-und-kontroversen/>.
- Baier, Dirk (2018): Gutachten für den 23. Deutschen Präventionstag am 11. & 12. Juni 2018 in Dresden. In: in: Kerner, Hans-Jürgen & Marks, Erich (Hrsg.): Internetdokumentation des Deutschen Präventionstages. Hannover 2017. [www.praeventionstag.de/dokumentation.cms/4094](http://www.praeventionstag.de/dokumentation.cms/4094).
- Becker, Carsten/Brunsen, Hendrik/Einhorn, Annika (2014): Evaluation des XENOS-Sonderprogramms „Ausstieg zum Einstieg“. Abschlussbericht mit einer zusammenfassenden Betrachtung des Förderzeitraums 2010-2013 im Auftrag des Bundesministeriums für Arbeit und Soziales (BMAS). Berlin, GIB.
- 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 2015/2016 – Aktuelle Beiträge aus Wissenschaft und Praxis, S. 97-116.
- Bellasio, Jacopo/Hofman, Joanna, Ward, Antonia/Nederveen, Fook/Knack, Anna/Arya Sofia Meranto/Hoorens, Stijn (2018): Counterterrorism evaluation. Taking stock and looking ahead. RAND Europe. Online: [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR2600/RR2628/RAND\\_R2628.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR2600/RR2628/RAND_R2628.pdf).
- Bellmann, Johannes/Müller, Thomas (2011): Evidenzbasierte Pädagogik – ein Déjà-vu? In: Dies. (Hrsg.): Wissen, was wirkt. Kritik evidenzbasierter Pädagogik. Wiesbaden, VS, S. 9-32.
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation - Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Deutsches Jugendinstitut (Hrsg.): Wirkungsevaluation in der Kinder- und Jugendhilfe - Einblicke in die Evaluationspraxis. München, DJI, S. 25-46.
- Beywl, Wolfgang/Niestroj, Melanie: Das A-B-C der wirkungsorientierten Evaluation. Glossar – Deutsch / Englisch – der wirkungsorientierten Evaluation. 2., vollständig bearbeitete und ergänzte Auflage. Köln, Univention Institut für Evaluation.
- Bischoff, Ursula/König, Frank/Zimmermann, Eva (2018): Pädagogik wirkt?! Wirkungsevaluationen in der pädagogischen Präventionsarbeit. In: Glaser, Michaela, Frank, Anja & Herding, Maruta (Hrsg.): Gewaltorientierter Islamismus im Jugendalter. Perspektiven aus Jugendforschung und Jugendhilfe. Weinheim und Basel, Beltz Juventa, S. 147-157.
- Bundesministerium des Innern, für Bau und Heimat Bund, [https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/praeventionsprogramm-islamismus.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/praeventionsprogramm-islamismus.pdf?__blob=publicationFile&v=2).
- Bohnsack, Ralf (2011): Hauptbegriffe qualitativer Sozialforschung. Opladen, Budrich (3. durchgesehene Aufl.).
- Bröckling, Ulrich (2017): Gute Hirten führen sanft. Über Menschenregierungskünste. Berlin, Suhrkamp.
- Bundesministerium für Familien, Senioren, Frauen und Jugend (BMFSFJ) (2017): Bericht der Bundesregierung über Arbeit und Wirksamkeit der Bundesprogramme zur Extremismusprävention. Berlin, BMFSJ.
- Campbell, Donald T. (1969): Reforms as Experiments. In: American Psychologist, 24, 4, S. 409-429.
- Chen, Huey-Tseh (2004): The roots of theory-driven evaluation: current views and origins. In: Alkin, Marvin C. (Ed.): Evaluation Roots. Tracing theorists' views and influences. Thousand Oaks, Sage, S. 132-152.
- DeGEval, [https://www.degeval.org/fileadmin/Publikationen/DeGEval\\_Standards\\_fuer\\_Evaluation\\_-\\_Erste\\_Revision\\_2016\\_.pdf](https://www.degeval.org/fileadmin/Publikationen/DeGEval_Standards_fuer_Evaluation_-_Erste_Revision_2016_.pdf).



- Deutscher Bundestag (2018): Drucksache 19/1134.
- Deutscher Präventionstag (2018): Dresdner Erklärung des 23. Deutschen Präventionstages (vom 07.06.2018). online-Bezug über: <http://www.praeventionstag.de/nano.cms/news/details/2951>, 23.06.2018.
- Die Bundesregierung (2016): Strategie der Bundesregierung zur Extremismusprävention und Demokratieförderung. Berlin, Bundesministerium für Familie, Senioren, Frauen und Jugend/Bundesministerium des Innern, für Bau und Heimat.
- Dilthey, Wilhelm (1981): Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften. Frankfurt a. M., Suhrkamp (Orig. 1910).
- Edwards Ward/Newman Robert J. (1982): Multiattribute Evaluation. Beverly Hills, Sage.
- Einhorn, Annika/Goldkamp, Sina/Grebe, Tim/Kroos, Daniela/Popp, Sandra (2012): Evaluation des „XENOS-Sonderprogramms Ausstieg zum Einstieg. Kurzbericht. Berlin, GIB.
- Einhorn, Annika/Lietzmann, Anja/Meyer, Stefan (2013): Evaluation des XENOS- Sonderprogramms „Ausstieg zum Einstieg“. 2. Zwischenbericht. Ergebnisse der 2. Online-Befragung & Präsentation von Good-Practice-Ansätzen. Berlin, GIB.
- Elverich, Gabi (2011): Demokratische Schulentwicklung. Potenziale und Grenzen einer Handlungsstrategie gegen Rechtsextremismus. Wiesbaden, VS Springer.
- Feddes, Allard R./Gallucci, Marcello (2015): A Literature Review on Methodology used in Evaluating Effects of Preventive and De-radicalisation Interventions. In: JD Journal for Deradicalization, Winter 2015/16 (5), S. 1-27. (PDF) online unter: <http://journals.sfu.ca/jd/index.php/jd/article/view/33> (30.04.2018).
- Fitzpatrick, Jody L./Sanders, James R./Worthen, Blaine R. (2010): Program Evaluation. Alternative Approaches und Practical Guidelines. 4th ed., Upper Saddle River, NJ, Pearson.
- Gansewig, Antje (2018): Prävention von politischen Extremismus in Deutschland. Eine Betrachtung zur Bedarfs- und Angebotslage. In: Walsh, Maria, Pniewski, Benjamin, Kober, Marcus & Armbrorst, Andreas (Hrsg.): Evidenzorientierte Kriminalprävention. Ein Leitfaden für Politik und Praxis. Wiesbaden, VS Springer, S. 465-488.
- Glaser, Stefan/Pfeiffer, Thomas: Erlebniswelt Rechtsextremismus. modern – subversiv – hasserfüllt, 5. Aufl. Schwalbach/Ts., Wochenschau.
- Gruber, Florian/Lützing, Saskia (2017): Extremismusprävention in Deutschland – Erhebung und Darstellung der Präventionslandschaft. Modulabschlussbericht. Wiesbaden, BKA.
- Guba, Egon G./Lincoln, Yvonna S. (1989): Fourth Generation Evaluation. Newbury Park, CA, Page.
- Haubrich, Karin (2009): Sozialpolitische Innovation ermöglichen. Die Entwicklung der rekonstruktiven Programmtheorie-Evaluation am Beispiel der Modellförderung in der Kinder- und Jugendhilfe. Münster/New York/München/Berlin, Waxmann Internationale Hochschulschriften.
- Hempel, Carl G./Oppenheim, Paul (1948): Studies of the Logic of Explanation. In: Philosophy of Science, 15, 2, S. 135-175.
- Heinrich, Gudrun (2017): Unterstützung von außen und langer Atem. Strategien gegen Rechtsextremismus im Sozialraum Schule, in: Glaser, Stefan & Pfeiffer, Thomas: Erlebniswelt Rechtsextremismus. modern – subversiv – hasserfüllt, 5. Aufl. Schwalbach/Ts., Wochenschau, S. 165-175.
- Herzog, Werner (2011): Eingeklammerte Praxis – ausgeklammerte Profession. Eine Kritik der evidenzbasierten Pädagogik. In: Bellmann, Johannes & Müller, Thomas (Hrsg.): Wissen, was wirkt. Kritik evidenzbasierter Pädagogik. Wiesbaden, VS, S. 123-145.
- Horgan, John/Braddock, Kurt (2010): Rehabilitating the Terrorists?: Challenges in Assessing the Effectiveness of De-radicalization Programs. In: Terrorism and Political Violence, 22, S. 267-291. doi: 10.1080/09546551003594748.
- Hufer, Klaus-Peter (2017): Argumente gegen Parolen und Populismus. Aktualisierter Sonderdruck aus „Argumente am Stammtisch“. Schwalbach/Ts., Wochenschau.
- Hufer, Klaus-Peter: Wirkungslose Wirkungsforschung? Rückblick und Ausblick. In: Journal für politische Bildung, 3/2018, 10-14.

- Innenministerkonferenz, [https://www.innenministerkonferenz.de/IMK/DE/termine/to-beschluesse/2016-11-29\\_30/nummer%20zu%20anlage%201\\_evaluationsbericht\\_apr\\_nrw.pdf?\\_\\_blob=publicationFile&v=2](https://www.innenministerkonferenz.de/IMK/DE/termine/to-beschluesse/2016-11-29_30/nummer%20zu%20anlage%201_evaluationsbericht_apr_nrw.pdf?__blob=publicationFile&v=2). Joint Committee on Standards for Educational Evaluation (1994): The program evaluation standards. How to assess evaluations of educational programs (2. ed.). Thousand Oaks, CA, Sage.
- Joint Committee on Standards for Educational Evaluation (Hrsg.) (2000): Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation". (2. Auflage). Opladen, Springer VS.
- Joint Committee on Standards for Educational Evaluation & Sanders, J. R. (2006): Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation" (3. erweiterte und aktualisierte Auflage). Übersetzt und für die dt. Ausgabe erweitert von Wolfgang Beywl und Thomas Widmer. Wiesbaden, Springer VS.
- 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. Halle, Deutsches Jugendinstitut, S. 32-52.
- Kober, Marcus (2017): Zur Evaluation von Maßnahmen der Prävention von religiöser Radikalisierung in Deutschland. In: Journal for Deradicalization, 11.
- Kober, Marcus/Armborst, Andreas (2017): Forschungssynthese: Evaluation von Ansätzen und Maßnahmen zur Prävention Islamistischer Radikalisierung. Bericht zur Veröffentlichung im Portal für wissens- und evidenzbasierte Sicherheit und Prävention (WESPE). Verfügbar unter: [https://www.nzkrim.de/fileadmin/user\\_upload/Praevention\\_Extremismus\\_Islamismus\\_20170521.pdf](https://www.nzkrim.de/fileadmin/user_upload/Praevention_Extremismus_Islamismus_20170521.pdf), 23.06.2018.
- Köhler, Daniel (2016): Strukturelle Qualitätsstandards in der Interventions- und Präventionsarbeit gegen gewaltbereiten Extremismus: ein Handbuch für Praktikerinnen, Praktiker und staatliche Koordinationsstellen sowie zivilgesellschaftliche Projektträger in Deutschland. Stuttgart, Ministerium für Inneres, Digitalisierung und Migration.
- Köhler, Daniel (2017): Preventing violent radicalization. Programme Design and Evaluation. Collection Monografias, S. 91-98; online: [www.cidob.org/en/articulos/monografias/resilient\\_cities/preventing\\_violent\\_radicalisation\\_programme\\_design\\_and\\_evaluation](http://www.cidob.org/en/articulos/monografias/resilient_cities/preventing_violent_radicalisation_programme_design_and_evaluation), 27.06.2018.
- Lewin Simon/Booth Andrew/Glenton Claire/Munthe-Kaas Heather/Rashidian Arash/Wainwright Megan/Bohren Meghan A./Tunçalp Özge/Colvin Christopher J./Garside Ruth/Carlsen Benedicte/Langlois Etienne V./Noyes Jane (2018): Implement Sci. 2018 Jan 25;13 (Suppl 1):2. doi: 10.1186/s13012-017-0688-3., 20.07.2018.
- Lüders, Christian (2010): Neue Wege der Evaluation gewalt- und kriminalpräventiver Maßnahmen und Projekte. Das Logische Modell als Instrument der Evaluation in der Kinder- und Jugendkriminalitätsprävention. In: Berliner Forum Gewaltprävention Nr. 41: Evaluation und Qualitätsentwicklung in der Gewalt- und Kriminalitätsprävention. Dokumentation des 10. Berliner Präventionstages, S. 127-139.
- Mannheim, Karl (1980): Strukturen des Denkens. Frankfurt a. M., Suhrkamp.
- Möller, Kurt (2012): Der Dritte Raum. Möglichkeiten zu gestaltungsorientierter Grundlagenforschung in Wissenschaft-Praxis-Kooperation. In: Unzicker, Kai/Hessler, Gudrun (Hrsg.): Öffentliche Sozialforschung und Verantwortung für die Praxis. Zum Verhältnis von Sozialforschung, Praxis und Öffentlichkeit. Wiesbaden, Springer VS, S. 85-108.
- Möller, Kurt/Küpper, Beate/Buchheit, Frank/Neuscheler, Florian (2015): Evaluation des Aussteigerprogramms für Rechtsextremisten des Landes Nordrhein-Westfalen (APR NRW). Abschlussbericht. Esslingen. [www.innenministerkonferenz.de/IMK/DE/termine/to-beschluesse/2016-11-29\\_30/nummer%20zu%20anlage%201\\_evaluationsbericht\\_apr\\_nrw.pdf?\\_\\_blob=publicationFile&v=2](https://www.innenministerkonferenz.de/IMK/DE/termine/to-beschluesse/2016-11-29_30/nummer%20zu%20anlage%201_evaluationsbericht_apr_nrw.pdf?__blob=publicationFile&v=2).

- Möller, Kurt/Lempp, Marion/Nolde, Kai/Schuhmacher, Nils (2017): Abschlussbericht zum Projekt ‚Rückgrat!‘ – Eine Wissenschaft-Praxis-Kooperation gegen Rechtsextremismus und gruppierungsbezogene Ablehnungen. Sachbericht an die Robert Bosch Stiftung und die Bundeszentrale für politische Bildung. Esslingen, Hochschule Esslingen (unv. Mscr.).
- Möller, Kurt/Neuscheler, Florian (2016): Bericht zur Evaluation von Konzeption und Struktur des Modellprojekts EXTRA – Ausstiegshilfe aus dem Rechtsextremismus – des Landes Sachsen-Anhalt. Esslingen (unv. Mscr.).
- Möller, Kurt/Neuscheler, Florian (2017): Bericht zur Prozess- und Ergebnisevaluation der Ausstiegshilfe aus dem Rechtsextremismus EXTRA (EXTRemismus Ausstieg) des Landes Sachsen-Anhalt. Esslingen (unv. Mscr.).
- Möller, Kurt/Neuscheler, Florian (2018): Abschlussbericht zur Evaluation der Beratungsstelle Hessen – Religiöse Toleranz statt Extremismus. Esslingen (unv. Mscr.).
- Möller, Kurt/Neuscheler, Florian (2019): Islamismus und Rechtsextremismus - Was wissen wir über Radikalisierungsprozesse, was kann dagegen unternommen werden? In: Zeitschrift für Jugendkriminalrecht und Jugendhilfe – ZJJ, H. 1, S. 12-19.
- Nationales Zentrum für Kriminalprävention, [https://www.nzkrim.de/typo3conf/ext/nzk/Resources/Public/pdf/evidence\\_index.pdf](https://www.nzkrim.de/typo3conf/ext/nzk/Resources/Public/pdf/evidence_index.pdf).
- Patton, Michael Quinn (2003): Qualitative Evaluation Checklist. <https://wmich.edu/sites/default/files/attachments/u350/2014/qualitativeevalchecklist.pdf>.
- Pawson, Ray/Tilley, Nick (1997): Realistic Evaluation. London, Sage.
- Pawson, Ray/Tilley, Nick (1997): Realist Evaluation. British Cabinet Office. [http://www.community-matters.com.au/RE\\_chapter.pdf](http://www.community-matters.com.au/RE_chapter.pdf).
- Romaniuk, Peter/Fink, Naureen C. (2012): From Input to Impact. Evaluating Terrorism Prevention Programs. Center on Global Counterterrorism Cooperation. Online: [http://globalcenter.org/wp-content/uploads/2012/10/CGCC\\_EvaluatingTerrorismPrevention.pdf](http://globalcenter.org/wp-content/uploads/2012/10/CGCC_EvaluatingTerrorismPrevention.pdf).
- Schuhmacher, Nils (o.J.; 2018): Evaluation der Fach- und Beratungsstelle für religiös begründete Radikalisierung – Legato. Abschlussbericht. Hamburg. Online: [www.wiso.uni-hamburg.de/fachbereich-sowi/professuren/hentschel/forschung/legato.html](http://www.wiso.uni-hamburg.de/fachbereich-sowi/professuren/hentschel/forschung/legato.html), 02.04.2019.
- 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.
- Spiegel, Hiltrud v. (2013): Methodisches Handeln in der Sozialen Arbeit. München, Reinhardt UTB (5. Aufl.).
- Steinke, Ines (2013): Gütekriterien qualitativer Forschung. In: Flick, Uwe, Kardorff, Ernst von & Steinke, Ines (Hrsg.): Qualitative Forschung. Ein Handbuch. 10. Aufl. Reinbek bei Hamburg, Rowohlt Taschenbuch Verlag, S. 319-331.
- Stockmann, Reinhard/Meyer, Wolfgang (2014): Evaluation. Eine Einführung. Zweite überarbeitete und aktualisierte Auflage. Opladen, Budrich (UTB).
- Stockmann, Reinhard (2006): Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement. Münster/New York/München/Berlin, Waxmann.
- Uhlmann, Milena (2017): Evaluation der Beratungsstelle „Radikalisierung“. Abschlussbericht. Nürnberg: Bundesamt für Migration und Flüchtlinge. Forschungsbericht 31.
- Ullrich, Simone/Moussa Nabo, Mitra/Nehlsen, Inga/de la Chau, Marlen (2018): Entwicklung von Evaluationskriterien in der Extremismusprävention. In: forum kriminalprävention, 2, S. 19-22.
- Walsh, Maria/Pniewski, Benjamin/Kober, Marcus/Armborst, Andreas (Hrsg.): Evidenzorientierte Kriminalprävention. Ein Leitfaden für Politik und Praxis. Wiesbaden, VS Springer.
- W. K. Kellogg Foundation (2001): Logic model development guide. Using logic models to bring together planning, evaluation, and action. Michigan, WKKF.

Williams Michael J./Kleinman, Steven M. (2013): A utilization-focused guide for conducting terrorism risk reduction program evaluations. In: Behavioral Sciences of Terrorism and Political Aggression <http://dx.doi.org/10.1080/19434472.2013.860183>.

Yarbrough, Donald B. (2011): The program evaluation standards. A guide for evaluators and evaluation users (3. ed.). Thousand Oaks, CA, Sage.

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 & Armbrorst, Andreas (Hrsg.): Evidenzorientierte Kriminalprävention in Deutschland: Ein Leitfaden für Politik und Praxis. Wiesbaden, VS Springer, S. 349-368.

## Quellen

ISO: [www.iso.org/standards.html](http://www.iso.org/standards.html).

Präventionstag: Quelle: <http://www.praeventionstag.de/nano.cms/news/details/2951>.

Präventionstag: Quelle: [www.praeventionstag.de/dokumentation.cms/4094](http://www.praeventionstag.de/dokumentation.cms/4094).

Univation, Institut für Evaluation: [www.univation.org/programmbaum](http://www.univation.org/programmbaum)