

PRIF REPORT

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



MONA KLÖCKNER // MELANIE
VERHOVNIK-HEINZE // RAPHAELA
SCHLICHT-SCHMÄLZLE // REIKO
NAKAMURA // JULIUS STRUNK

**MULTIMETHODISCHE EVALUATIONS-
DESIGNS: EINE ERHEBUNG DER
SOZIALWISSENSCHAFTLICHEN PRAXIS**

REPORT-REIHE

PrEval


PRIF Report 1/2022

MULTIMETHODISCHE EVALUATIONSDESIGNS: EINE ERHEBUNG DER SOZIALWISSEN- SCHAFTLICHEN PRAXIS

MONA KLÖCKNER // MELANIE VERHOVNIK-HEINZE //
RAPHAELA SCHLICHT-SCHMÄLZLE // REIKO NAKAMURA //
JULIUS STRUNK

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

Coverbild:

Lizenzinformationen: User 652234: <https://pixabay.com/photos/architecture-modern-building-facade-1048092/> (Bearbeitung: Anja Feix).

Textlizenz:

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



Adresse:

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

DOI: 10.48809/prifrep2201

ISBN: 978-3-946459-69-9

DIE AUTOR:INNEN

Mona Klöckner ist Psychologin und wissenschaftliche Mitarbeiterin an der HSFK. Ihre aktuellen Arbeitsschwerpunkte liegen in Evaluation und wissenschaftlicher Methodik sowie in der Erforschung des gesellschaftlichen Umgangs mit Extremismen und Verschwörungsideologien in Deutschland.

Dr. Melanie Verhovnik-Heinze ist Kommunikationswissenschaftlerin und wissenschaftliche Mitarbeiterin am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Frankfurt. Ihre Forschungsschwerpunkte sind Medienberichterstattung, Medienwirkung und Wissenschaftskommunikation, sowie Radikalisierungs- und Extremismusprävention.

Dr. Raphaela Schlicht-Schmälzle ist Politik- und Verwaltungswissenschaftlerin an der HSFK. Im Projekt PrEval arbeitet sie an der Entwicklung von Evaluationsdesigns für verschiedene Bereiche der politischen Bildung. Ihr Forschungsinteresse gilt insbesondere Global Citizenship Education und sozialer Bildungsungleichheit.

Reiko Nakamura war zunächst als Praktikantin, dann als studentische Hilfskraft am DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation beschäftigt. Sie studierte Education (B.A.) und Educational Policy and Planning (M.Ed.) in Japan und England und schließt derzeit ihr Studium der Erziehungswissenschaften (M.A.) an der Goethe-Universität Frankfurt ab. Ihre Forschungsinteressen liegen in der international vergleichenden Bildungsforschung.

Julius Strunk ist studentische Hilfskraft im PrEval-Projekt an der HSFK. Er studiert Internationale Studien/Friedens- und Konfliktforschung an der Goethe-Universität Frankfurt und der TU Darmstadt. Seine Arbeitsschwerpunkte liegen in der Extremismus- und sozialen Bewegungsforschung mit Bezug zur MENA-Region, sowie in wissenschaftlicher Methodik.

HSFK-Reportreihe *PrEval*

Über die Evaluation und die wissenschaftliche Begleitung von Maßnahmen der Extremismusprävention und von Angeboten der politischen Bildung wird aktuell in Fachpraxis und Wissenschaft vielfältig diskutiert. Das liegt nicht zuletzt daran, dass über verschiedene Programme auf Ebene des Bundes, der Länder und der Kommunen seit einigen Jahren immer mehr Fördermittel für solche Maßnahmen bereit stehen, die Fachpraxis sich immer vielfältiger, innovativer und professioneller aufstellt und auch die Forschung zu den Phänomenfeldern, aber auch zu Evaluationsmethoden stetig Fortschritte macht. Folgerichtig fordert beispielsweise auch der Kabinettsausschuss zur Bekämpfung von Rechtsextremismus und Rassismus in seinem Maßnahmenpaket eine langfristige Stärkung der Qualität und Wirksamkeit von Präventionsmaßnahmen und knüpft damit an zahlreiche Initiativen von Bundes- und Landesregierungen, von Fachpraxis und Wissenschaft an.

Das Forschungs- und Transfervorhaben *PrEval* (Evaluationsdesigns für Präventionsmaßnahmen – Multimethodische Ansätze zur Wirkungsermittlung und Qualitätssicherung in der Extremismusprävention sowie den Schnittstellen zur Gewaltprävention und politischen Bildung) setzt hier an. Ziel des Projekts ist es, im Dialog aller an Evaluation in diesem Feld beteiligten Akteursgruppen den aktuellen Wissensstand zu Qualitätssicherung und Evaluation in der Extremismusprävention, der Gewaltprävention und der politischen Bildung zu erheben und gemeinsam Evaluationsdesigns, gerade an den Schnittstellen zwischen verschiedenen Präventionsbereichen, Zugängen, Phänomenfeldern und Akteuren, zu erarbeiten. Dabei befasst sich *PrEval* mit Maßnahmen aus der politischen Bildung, der primären, sekundären und tertiären Prävention sowie Schnittstellen zur Gewalt- und Kriminalprävention. *PrEval* nutzt Analyse-, Monitoring- und Mapping-Formate und entwickelt multimethodische Evaluationsdesigns im Rahmen von ausgewählten Pilotstudien. Ein zentraler Mehrwert des Vorhabens ist die Einbindung verschiedener Präventionsakteure – aus der Fachpraxis, den Sicherheitsbehörden, aus der Verwaltung und der Wissenschaft.

Die *PrEval*-Reportreihe fasst die zentralen Ergebnisse des Projekts zusammen. Dabei wird zum einen ein systematischer Überblick über die Evaluationsbedarfe der deutschen Trägerlandschaft in den Präventionsfeldern Islamistischer Extremismus und Rechtsextremismus geboten, die bestehenden Kapazitäten für Evaluation und wissenschaftliche Begleitung von Präventionsmaßnahmen erhoben sowie Evaluationsstrukturen international vergleichend analysiert. Zum anderen werden multimethodische Evaluationsdesigns vorgestellt, die im Rahmen der *PrEval*-Pilotstudien für spezifische Präventionsmaßnahmen in enger Zusammenarbeit mit den Trägern (und teilweise auch Mittelgebern) dieser Maßnahmen entwickelt wurden. Möglichkeiten und Grenzen von wissenschaftlicher Begleitung werden diskutiert. Die zehn Beiträge der Reihe eint ein breites Verständnis von Evaluation, um den Komplexitäten des Felds und der darin verorteten Maßnahmen gerecht zu werden. Gleichwohl setzt jeder Report eigene, dem jeweiligen Thema angepasste Akzente in der Begriffsverwendung.

Das *PrEval*-Projekt wird vom Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK) koordiniert und von einem Verbund aus Forschungsinstituten und Trägern aus der Fachpraxis gestützt. Mehr Informationen zum Projekt: <https://preval.hsfk.de>

Übersicht über alle Titel der *PrEval*-Report-Reihe

Erfolgsbedingungen und Herausforderungen für die Zusammenarbeit von Fachpraxis und Wissenschaft bei der Evaluation von Angeboten politischer Bildung von Raphaela Schlicht-Schmälzle // Désirée Theis // Sina Tultschinetski // Melanie Verhovnik-Heinze // Christian Kautz // Maryam Kirchmann. PRIF Report 5/2021.

Erfahrungen aus der Evaluationsplanung eines Aussteigerprogramms. Voraussetzungen für Wirksamkeitserfassung in der tertiären Extremismusprävention von Mona Klöckner // Svetla Koynova // Johanna Liebich // Lisa Neef. PRIF Report 6/2021.

Multimethodische Evaluationsdesigns: eine Erhebung der sozialwissenschaftlichen Praxis von Mona Klöckner // Melanie Verhovnik-Heinze // Raphaela Schlicht-Schmälzle // Reiko Nakamura // Julius Strunk. PRIF Report 1/2022.

Weitere geplante Titel in dieser Reihe

Webvideos und Livestream zur Aufklärung über Desinformation. Evaluationsansätze zu digitalen Formaten in der politischen Bildung von Désirée Theis // Melanie Verhovnik-Heinze // Stefan Kroll // Keshia Kierek.

Monitoring, Evaluation und Lernen: Erfahrungen und Bedarfe der Fachpraxis in der Prävention von Rechtsextremismus und Islamismus von Svetla Koynova // Alina Mönig // Matthias Quent // Vivienne Ohlenforst.

Evaluationskapazitäten im Bereich der Extremismusprävention und der politischen Bildung in Deutschland von Andreas Uhl // Benjamin Zeibig // Manuela Freiheit // Andreas Zick.

Evaluation der politischen Bildung im Jugendstrafvollzug – Ansätze, Chancen, Herausforderungen von Sina Tultschinetski // Alisa Rieth // Désirée Theis.

Developmental Evaluation zur Begleitung von Entwicklungsprozessen in Projekten der Radikalisierungsprävention – Ein Anwendungsbeispiel von Olga Schmidt // Johanna Liebich // Mona Klöckner // Dennis Walkenhorst // Bacem Dziri.

Klientenzentrierte Evaluation in Multi-Agency Settings der Extremismusprävention – Möglichkeiten und Grenzen eines wirkungsorientierten Vorgehens von Susanne Johansson // Julian Junk // Johanna Liebich // Dennis Walkenhorst.

Evaluation und Qualitätssicherung in der Kooperation sicherheitsbehördlicher und zivilgesellschaftlicher Akteure in der sekundären und tertiären Extremismusprävention von Julian Junk // Svetla Koynova // Vivienne Ohlenforst // Maximilian Ruf // Lea Scheu.

Außerdem im Rahmen des *PrEval*-Projekts als GPPi-Study erschienen

Extremismusprävention evaluieren: Institutionelle Strukturen im internationalen Vergleich von Ase-na Baykal // Sarah Bressan // Julia Friedrich // Giulia Pasquali // Philipp Rotmann // Marie Wagner.

Wie werden multimethodische Evaluationen umgesetzt, inwiefern wird darin auch multiperspektivisch gearbeitet und was können wir daraus für das Feld der Extremismusprävention schließen? Der vorliegende Report liefert einen Überblick darüber, was Evaluierende von sozialen Programmen und Projekten meinen, wenn sie ihre Arbeit als multimethodisch bezeichnen. Er gibt damit eine Orientierung für all diejenigen, die an der Planung und Umsetzung von Evaluationen beteiligt sind – seien es Auftraggebende, Evaluierte oder Evaluierende.

Soziale Programme und Projekte wie das Bundesprogramm *Demokratie leben!* arbeiten mit komplexen Themen, auf die sie oft mit multidimensionalen Angeboten reagieren, die von multidisziplinären Teams umgesetzt werden. Ebenso sind die Problemlagen, auf welche die Programme reagieren, wie auch ihre Angebote selbst, oft nicht statisch, sondern entwickeln sich über die Zeit, in Reaktion auf neue Herausforderungen und Bedarfe. Darüber hinaus sind viele verschiedene Akteursgruppen an der Umsetzung von Maßnahmen beteiligt, deren Perspektiven und Akzeptanz für die Entwicklung und Umsetzung von Evaluationen wichtig sind. Auch die Extremismusprävention in Deutschland hat ein vielfältiges Erscheinungsbild, mit unterschiedlichen Maßnahmen, die hinsichtlich ihrer Wirkungsziele, Zielgruppen, Formate und pädagogischen Ansätze unterschiedlichen Logiken folgen.

Evaluationen können in diesem Kontext verschiedene Fokusse setzen und unterschiedliche Zwecke erfüllen. Sie müssen sich jedoch immer der Individualität und Komplexität ihres Gegenstandes annehmen. In der Evaluationsforschung hat sich daher der Standard der Gegenstandsangemessenheit etabliert. Um auf die Anforderungen reagieren zu können und ein möglichst angemessenes Bild der evaluierten Arbeit zu zeichnen, kann es für Evaluierende sinnvoll sein, mehrere Forschungsmethoden zu kombinieren (Multimethodik). Damit können nicht nur unterschiedliche Datenquellen nutzbar gemacht und verschiedene Evaluationsfragen beantwortet werden, sondern durch die unterschiedlichen methodischen Brillen kann auch ein reicheres Verständnis des Gegenstandes gewonnen werden. Die Erhebung der Perspektiven verschiedener Personen, die an der Umsetzung des Programms oder Projektes beteiligt sind (Multiperspektivität), kann ebenfalls zur Gegenstandsangemessenheit beitragen. Zusätzlich kann dadurch der praktische Nutzen der Evaluation erhöht werden.

Ob und wie Multimethodik und Multiperspektivität in Evaluationsvorhaben jedoch umgesetzt werden, hängt unter anderem von methodischen sowie pragmatischen Überlegungen ab. Für die Nutzung unterschiedlicher Methoden braucht es ein Konzept für den jeweiligen Erkenntnisnutzen einzelner Methoden, ihre Reihenfolge, Gewichtung und spätere Integration in ein Gesamtbild. Zusätzlich binden die Umsetzung, Vor- und Nachbereitung der Multimethodik sowie der Multiperspektivität viele zeitliche und personelle Ressourcen. Die Umsetzung multimethodischer und -perspektivischer Evaluationsdesigns kann deshalb sehr unterschiedlich aussehen.

Der vorliegende Report beschreibt, wie Multimethodik und Multiperspektivität über eine Vielzahl an Fachdisziplinen hinweg in der Breite der sozialwissenschaftlichen Evaluationsforschung umgesetzt werden. Die Extremismusprävention stellt dabei nur einen kleinen Teilbereich dar, der jedoch in den letzten Jahrzehnten stetig angewachsen ist und in dem aktuell, unter anderem im Projekt *PrEval*, aktiv exploriert wird, wie (multimethodische) Evaluationen gegenstandsangemessen und bedarfsgerecht umgesetzt werden können. Andere Themenfelder, wie etwa die Bereiche Bildung und

Gesundheit, sind teils evaluationserprobter als die Extremismusprävention. Wie Multimethodik und Multiperspektivität dort umgesetzt werden, kann für Evaluationsaufträge in der Extremismusprävention, in denen ein solches Vorgehen angebracht ist, daher als Inspiration dienen.

Um die Umsetzung von Multimethodik und Multiperspektivität in der empirischen Evaluationsforschung zu erfassen, führen wir eine Analyse der zentralen methodischen Merkmale in einer umfangreichen Stichprobe ($n = 317$) von wissenschaftlich publizierten Evaluationsstudien durch. Die Stichprobe stammt aus einer Grundgesamtheit von $N = 4.668$ Studien. Diese wurden in fünf Themenfelder ((1) Bildung, Schule und Erziehung, (2) Gesundheit und Soziales, (3) Politik, Verwaltung und Organisation, (4) Kommunikation und Medien, (5) Kriminalität, Sicherheit und Justiz) kategorisiert und aus ihnen wurde proportional eine kleinere Stichprobe zur Analyse der Methodik gezogen. Aufgrund von eingeschränkter Vergleichbarkeit in der Berichtslegung und der Zugänglichkeit schließen wir sogenannte ‚graue Literatur‘ aus. Mit dem Report schaffen wir einen ersten, deskriptiven Eindruck der verwendeten Methoden und ihrer Kombinationen in als multimethodisch bezeichneten Evaluationen und vergleichen diese über verschiedene Fachdisziplinen hinweg. Wir nehmen dabei weder eine qualitative Bewertung der Evaluationen in Abhängigkeit ihrer Methodik vor noch treffen wir Aussagen dazu, wie weit Multimethodik generell in der Evaluationsforschung verbreitet ist.

Die Ergebnisse zeigen, dass Multimethodik in der vorliegenden Stichprobe am häufigsten durch die Nutzung zweier Erhebungsmethoden umgesetzt wurde. Dabei beschränkte man sich meistens auf die Kombination unterschiedlicher Befragungsmethoden, seltener wurden diese mit Dokumentenanalysen oder Feldbeobachtungen kombiniert. In den Bereichen (1) Bildung und (2) Gesundheit spielten darüber hinaus fachspezifische Erhebungsmethoden (Leistungstests und Diagnosebögen) eine wichtige Rolle. Knapp die Hälfte aller Evaluationsstudien erhob Daten aus unterschiedlichen Perspektiven (ca. 57%). Wenn dies der Fall war, wurden am häufigsten zwei Perspektiven einbezogen. Zwischen den verschiedenen Themenfeldern können wir keine starken Unterschiede bei der Umsetzung von Multiperspektivität und Multimethodik feststellen.

Was bedeuten diese Ergebnisse für die künftige Evaluationsforschung? Der Report zeigt, dass sowohl Multimethodik als auch Multiperspektivität mit eher wenigen Variationen angewendet werden. Unklar bleibt, aus welchen Gründen in den analysierten Evaluationen diese Umsetzungen gewählt wurden. Künftige Studien können den vorliegenden Report als Basis nutzen, um die Beweggründe für und gegen den Einsatz von Multimethodik und Multiperspektivität unter Evaluierenden genauer zu beleuchten. Neben methodischen spielen hier auch pragmatische Überlegungen eine Rolle und die Umsetzung beider Elemente ist zeitlich und personell ressourcenintensiv. Aufgrund dieser Ressourcenintensität müssten Evaluationsaufträge voraussichtlich genügend Kapazitäten schaffen, um Evaluierenden die Möglichkeit zu eröffnen, alle Wege zur Gegenstandsangemessenheit gleichermaßen abwägen zu können.

1. Einleitung	1
2. Begriffshintergründe – Multimethodik und Multiperspektivität	3
2.1 Die praktische Umsetzung von Multimethodik	3
2.2 Multiperspektivität: Verschiedene Akteure und Blickwinkel	5
2.3 Die Verbreitung von Multimethodik in der Evaluationspraxis	6
3. Methodisches Vorgehen	7
3.1 Grundgesamtheit, Auswahl und Kategorisierung der Untersuchungseinheiten	7
3.2 Stichproben und Kodierung der Analyse Kriterien	9
4. Ergebnisse	11
5. Diskussion	17
5.1 Ausblick auf weitere Aspekte der Evaluationsforschung	18
6. Fazit	20
Literatur	22
Anhang	25

1. EINLEITUNG

Evaluationen von sozialen Programmen wie beispielsweise vom Demokratieförderprogramm *Demokratie leben!* sind ein wichtiger Baustein für die planvolle Nutzung öffentlicher Gelder. Sie können die Programme in ihrer inhaltlichen und operativen Weiterentwicklung unterstützen. Insbesondere im Feld der Extremismusprävention hofft die Öffentlichkeit verständlicherweise darauf, überprüfen zu können, ob die präventiven Maßnahmen wie geplant wirken. Wie Evaluationsdesigns in diesem und anderen Feldern organisiert und umgesetzt werden, ist daher von hoher gesellschaftspolitischer Relevanz. Es ist allerdings voraussetzungsvoll, der Dynamik und Komplexität sozialer und pädagogischer Tätigkeiten mit wissenschaftlichen Methoden gerecht zu werden. Mehrere Methoden zu nutzen, kann dabei helfen, aktuellen Standards entsprechend gegenstandsangemessen zu evaluieren (DeGEval 2016; Möller et al. 2020: 419; Milbradt 2018), etwa, um ein möglichst facettenreiches Gesamtbild des Evaluationsgegenstandes zu gewinnen und so die Aussagekraft und die Nützlichkeit der Ergebnisse zu erhöhen. Im Feld der Extremismusprävention in Deutschland wird die Multimethodik daher als wertvolle Möglichkeit in der Durchführung von Evaluationen – unter geeigneten Umständen – behandelt, wenngleich dort (noch) keine einheitliche „Evaluationskultur“ besteht (Milbradt 2019). Die konkrete Umsetzung solcher multimethodischen Evaluationsdesigns kann jedoch sehr unterschiedlich aussehen.

Was meinen Evaluierende daher genau, wenn sie ihre Arbeit multimethodisch nennen? Der vorliegende Report präsentiert eine Übersicht der genutzten wissenschaftlichen Methoden aus 317 Evaluationen, die sich als multimethodisch bezeichnen. Die Studien wurden aus einer Vollerhebung aller Evaluationen geschöpft, die zwischen den Jahren 2010 und 2020 in wissenschaftlichen Zeitschriften und als Dissertationen in deutscher oder englischer Sprache publiziert wurden, und decken verschiedene sozialwissenschaftliche Fachbereiche und Themengebiete ab. Der Report gibt damit eine Orientierung für Auftraggebende, Evaluierende und Evaluierte im Feld der Extremismusprävention in Deutschland – einem verhältnismäßig jungen Bereich der Evaluation, in dem rege diskutiert und ausgehandelt wird, wie bedarfsgerechte und insbesondere multimethodische Evaluationen umgesetzt werden können (Milbradt 2019). Er geht aus dem einschlägigen Forschungsprojekt *PrEval – Evaluationsdesigns für Präventionsmaßnahmen* hervor und bildet die Basis für weitere, vertiefende Analysen.¹

In der Evaluationsforschung verwenden Wissenschaftler:innen Forschungsmethoden, um Erkenntnisse über die Ergebnisse oder Prozesse von verschiedensten Gegenständen zu gewinnen – von Gesetzen und politischen Entscheidungen über Bildungs- und Präventionsprogramme bis hin zu medizinischen Behandlungen und Personalentwicklungsmaßnahmen einzelner Unternehmen. Dabei können Evaluierende nicht mit vorgefertigten oder allgemeingültigen Evaluationsdesigns arbeiten, sondern müssen sich den individuellen Rahmenbedingungen des Evaluationsgegenstands sowie dem Erkenntnisinteresse und Zweck des jeweiligen Auftrags annehmen. In der Folge müs-

1 Erste Befragungen des Feldes, etwa zu Evaluationserfahrungen und -bedarfen, sind bereits erfolgt; Einblicke in erste Ergebnisse bieten die PRIF Spotlights 07/2021 (Quent 2021) und 08/2021 (Koyanova 2021).

sen sie ihre Forschungsmethoden entsprechend auswählen. Dieses Prinzip hat sich als ‚Gegenstandsangemessenheit‘ etabliert.

Soziale Programme und Projekte sind oft mit mehrdimensionalen Problemstellungen konfrontiert und begegnen diesen mit ebenso vielschichtigen Maßnahmen, die mehrere Ziele gleichzeitig verfolgen. Dies gilt auch im Feld der Extremismusprävention. Für entsprechende gegenstandsangemessene Evaluationen bietet sich daher der Einsatz mehrerer unterschiedlicher Methoden an (Multimethodik), was daher auch in vielen Evaluationsansätzen vorgesehen ist (Döring/Bortz 2016a: 991). Durch den Einsatz multipler Methoden wird es möglich, verschiedene Datenquellen zu erschließen, mehrere Fragestellungen parallel oder nacheinander zu behandeln oder unterschiedliche Bereiche des Gegenstands zu beleuchten. Damit können die Flexibilität, Glaubwürdigkeit und Praxisrelevanz von Evaluationen erhöht werden. Aus wissenschaftlicher Sicht können sich die Methoden außerdem gegenseitig ergänzen, falsifizieren oder in Kombination neue Fragestellungen und Schlussfolgerungen provozieren, die ansonsten nicht möglich gewesen wären (Creswell/Piano Clark 2018: 12).² Multimethodik ist somit auch ein Weg zur Triangulation – der Beleuchtung des Gegenstands aus unterschiedlichen Blickwinkeln. Doch nicht nur unterschiedliche Methoden können Perspektiven auf den Gegenstand eröffnen. Es kann auch zweckdienlich sein, die unterschiedlichen an der Planung und Umsetzung der Programme und Projekte beteiligten Personen zur Beantwortung der Evaluationsfragen (Multiperspektivität³) zu befragen (Möller 2019).

Multimethodik und Multiperspektivität können also sowohl wegen der Eigenschaften des Evaluationsgegenstands selbst (beispielsweise aufgrund seiner Komplexität) gegenstandsangemessen sein als auch wegen entsprechender Anforderungen des Evaluationsauftrags (beispielsweise, wenn ein möglichst facettenreiches und vollständiges Bild der Arbeit gewonnen werden soll). Aus diesem Grund betonen einige Wissenschaftler:innen seit mehreren Jahren die Gründe für multimethodische Evaluationsdesigns – so etwa in der Entwicklungszusammenarbeit (OECD 2012; Chigas et al. 2014), im Peacebuilding (Abu-Nimer 2020) und auch im Bereich der Extremismusprävention (Milbradt 2018; Möller 2019). Um zu überprüfen, ob der Einsatz mehrerer Methoden im jeweiligen Fall tatsächlich angemessen ist und einen Mehrwert bringen würde, müssen der Evaluationsgegenstand (Wie genau sieht die evaluierte Arbeit aus? Auf welchen Wirkungsannahmen und Prozessen baut sie auf?) sowie die Erkenntnisziele und Zwecke der Evaluation (Welche Fragen sollen beantwortet werden? Welchem Zweck sollen die gewonnen Erkenntnisse dienen?) betrachtet werden. Der Einsatz mehrerer Methoden bringt maßgeblichen Mehraufwand mit sich, weil methodische Vorüberlegungen und zusätzlicher Zeitaufwand für Umsetzung, Vor- und Nachbereitung notwendig sind. Insbesondere verschiedene (quantitative und qualitative) Methoden einzusetzen, bedarf Überlegungen dazu, welche Erkenntnisse durch welche Methode gewonnen werden sollen und inwiefern ihre Ergebnisse zusammenfließen können (Döring/Bortz 2016b). Darüber hinaus müssen durch den Mehraufwand

2 Zu der Frage, ob und wie quantitative und qualitative Methoden kombiniert und integriert werden können, bestehen langjährige Debatten. Diese umfassen auch die Frage, ob ein solcher Methodenmix neben dem qualitativen und quantitativen Forschungsparadigma als drittes Paradigma anerkannt werden kann und sollte (vgl. Döring/Bortz 2016b: 71).

3 Multiperspektivität kann ebenso die Einbindung mehrerer beteiligter Personen in die Planung der Evaluation meinen. Im vorliegenden Report wird sie verstanden als ihre Einbindung zur Beantwortung der Evaluationsfragen.

in der Regel auch mehr Personen an der Umsetzung beteiligt sein, was Koordination und intensive Austausche erfordert. Evaluationen müssen jedoch meist in einem begrenzten zeitlichen und finanziellen Budget umgesetzt werden. Außerdem sind die Möglichkeiten zur Umsetzung praktisch begrenzt durch das methodische Repertoire der Evaluierenden und den Zugang, der ihnen zu Daten des evaluierten Programms oder Projekts gewährt werden kann oder soll.

Bei der Organisation multimethodischer Evaluationsdesigns müssen diese Aspekte abgewogen werden, weshalb die Ausgestaltung von Multimethodik sehr unterschiedlich ausfallen kann. Der vorliegende Report gibt einen breiten empirischen Überblick darüber, was in der sozialwissenschaftlichen Evaluationspraxis als multimethodisch umgesetzt wird. Er gibt damit erste Hinweise, in welcher Form ein theoretischer Anspruch der Multimethodik in der Praxis tragbar ist. Dabei ermöglicht er außerdem einen Vergleich der Evaluationspraxis über Fachbereiche hinweg. Der Report präsentiert eine auf einer Datenbanksuche basierende Vollerhebung von wissenschaftlich publizierten Evaluationsstudien ($N = 4.668$) und eine Analyse des methodischen Vorgehens einer Teilstichprobe ($n = 317$) dieser Studien. Die Ergebnisse richten sich vor allem an Beteiligte von Evaluationen und bieten eine erste Orientierung dazu, wie multimethodische Evaluationen in der Praxis bisher umgesetzt werden. Für die Extremismusprävention kann der Report damit in aktuelle Debatten hineinspielen, die sich damit befassen, wie multimethodische Designs praktisch umgesetzt werden und zu möglichst gegenstandsangemessener und bedarfsgerechter Evaluation beitragen können. Zwar professionalisiert sich das Feld stetig und kann mittlerweile auf mannigfaltige Evaluationsexpertise zurückgreifen, die Umsetzung von multimethodischen Designs ist bisher jedoch nicht weit verbreitet (Feddes/Gallucci 2015). Inspirationen zu ihrer Umsetzung können im Feld daher (noch) nicht einzig aus sich selbst heraus geschöpft werden, wohl aber aus anderen, hier stärker ausgebauten Evaluationsefeldern. Der verhältnismäßig junge und dennoch bereits professionalisierte Fachbereich kann sich somit von den Erfahrungen und Praktiken aus anderen sozialwissenschaftlichen Evaluationsefeldern inspirieren lassen und von ihnen lernen.

Im Folgenden geht Kapitel 2 auf die Kernbegriffe des Reports – Multimethodik und Multiperspektivität – ein. Kapitel 3 beschreibt die methodische Vorgehensweise bei der Erstellung der Übersicht, Kapitel 4 ihre Ergebnisse. Kapitel 5 schließlich diskutiert die Schlussfolgerungen, die zur Umsetzung und Umsetzbarkeit von multimethodischen Evaluationen aus den Ergebnissen gezogen werden können.

2. BEGRIFFSHINTERGRÜNDE – MULTIMETHODIK UND MULTIPERSPEKTIVITÄT

2.1 DIE PRAKTISCHE UMSETZUNG VON MULTIMETHODIK

Es besteht keine Einigkeit darüber, auf was sich der Begriff der Multimethodik bezieht. Manche denken hierbei nur an die Ebene von Erhebungs- oder Analysemethoden, andere an größer gefasste methodologische Herangehensweisen. Als Minimaldefinition beschreibt multimethodisches Vor-

gehen das Kombinieren verschiedener Erhebungs- und Analysemethoden aus derselben oder unterschiedlichen Methodenfamilien (Palinkas et al. 2019). Es können also mehrere qualitative oder mehrere quantitative Methoden unter sich oder untereinander kombiniert werden (Hunter/Brewer 2015: 187). Um diesen unterschiedlichen Verständnissen Rechnung zu tragen, verwenden wir die niedrigste Definitionsebene der Erhebungs- und Analysemethoden. In diesem Sinne ist eine Forschungsarbeit in unserem Verständnis multimethodisch, wenn sie mindestens zwei Erhebungsmethoden miteinander kombiniert.

Mixed Method-Designs, die explizit quantitative und qualitative Methoden kombinieren, sind eine Variante multimethodischen Vorgehens (Hesse-Biber 2015: xxxix). Aufgrund der starken Verbreitung von *Mixed Method-Designs* werden die beiden Begriffe – Multimethodik und *Mixed Methods* – mitunter synonym verwendet (Johnson et al. 2007). Für manche Autor:innen ist eine Abgrenzung jedoch entscheidend, da *Mixed Methods*-Ansätze aus ihrer Sicht über das bloße methodische Kombinieren hinausgehen. Stattdessen hätten sie das Potenzial, eine vertiefte Betrachtung des Forschungsgegenstandes zu eröffnen und Erkenntnisse zu generieren, die mit nur einer Methodenfamilie nicht gewonnen werden könnten (Greene 2015: 607). Debatten um die Kombinierbarkeit von quantitativer und qualitativer Forschung haben eine lange Tradition. In der Evaluation, die als Anwendungsforschung der evaluierten Praxis nutzen soll, ist die Unterscheidung zwischen quantitativem und qualitativem Arbeiten jedoch weniger relevant. Eine Methodenvielfalt wird hier in einigen Evaluationsansätzen bereits vorausgesetzt (etwa bei der *Realist Evaluation* nach Pawson/Tilley (1997) oder der *Developmental Evaluation* nach Patton (2011)). Ebenso kommt ein Methodenmix durch das Ziel der Gegenstandsangemessenheit häufig von alleine zustande. Dennoch weisen einige Autor:innen darauf hin, dass ein Methodenmix nicht unsystematisch und ohne Überlegungen zur Integration der Ergebnisse und dem daraus entstehenden Erkenntnisnutzen zulässig sei (Greene 2015: 607; Döring/Bortz 2016a: 991).

Solche Überlegungen sollten beispielsweise die Reihenfolge und Gewichtung der verschiedenen Methoden informieren. Sollten sie parallel oder aufeinander aufbauend genutzt werden? Ist die Gewichtung gleichberechtigt oder wiegen einige Erkenntnisinteressen – und damit auch ihre Umsetzung in der Ressourcenplanung – schwerer? Für *Mixed Method-Designs* existieren hierzu mehrere Typologien, die als Orientierung dienen könnten. Creswell und Plano Clark (2018) etwa unterscheiden Designs zum einen dahingehend, ob sie vorab fixiert sind oder sich erst im Forschungsprozess entwickeln, zum anderen anhand der Sequenzierung der Methoden, deren jeweiliger Gewichtung und des Integrationszeitpunkts. Daraus ergeben sich die drei Idealtypen konvergierender, erklärend sequentieller und explorierend sequentieller Designs. Bei konvergierenden Designs erfolgt die qualitative und quantitative Datenerhebung parallel und die erhobenen Daten werden integrativ analysiert. In sequentiellen Designs geht eine Methodenfamilie der anderen voraus: Erklärend sequentielle Designs setzen zuerst quantitative, dann qualitative Methoden ein, explorierend sequentielle nutzen qualitative Methoden als Grundlage für den Einsatz von quantitativen Methoden.

2.2 MULTIPERSPEKTIVITÄT: VERSCHIEDENE AKTEURE UND BLICKWINKEL

An vielen sozialen Maßnahmen und Programmen, auch in der Extremismusprävention, ist eine große Bandbreite an Akteuren mit teils unterschiedlichen Wissens- und Erfahrungsbeständen sowie Interessen beteiligt. Um die Evaluationsfragen möglichst unverzerrt zu beantworten, kann der Einbezug dieser Perspektiven entscheidend sein. Dies ist beispielsweise der Fall, wenn Arbeitsprozesse beleuchtet werden sollen oder die Evaluierenden vorhandenes (Erfahrungs-)Wissen erheben möchten. Obwohl die meisten sozialen Programme die grundsätzlichen Annahmen und Abläufe ihrer Arbeit dokumentarisch festhalten, besteht darüber hinaus viel relevantes implizites Erfahrungs- und Faktenwissen, das die Arbeit beeinflusst.

Zu den relevanten Akteuren zählen die Mitarbeitenden, das Personal auf der Führungs- bzw. Koordinationsebene, die Adressat:innen der Arbeit, Mittelgebende sowie (Mitarbeitende in) Partnerorganisationen und weitere Akteure, etwa bei Kooperationen. Zum Beispiel bauen Maßnahmen politischer Bildung im Strafvollzug wesentlich auf der Zusammenarbeit der Maßnahmendurchführenden mit der Justizvollzugsanstalt auf, bei Maßnahmen in Schulen sind die Schulmitarbeitenden relevante Partner:innen, und bei Maßnahmen der tertiären Extremismusprävention arbeiten oft Zivilgesellschaft und Sicherheitsbehörden zusammen. In konstruktivistischer Lesart müssen die jeweiligen Perspektiven dieser Akteure dabei nicht objektiv „richtig“ sein, sondern bilden eine – oder die – relevante Wahrnehmung der Wirklichkeit ab. Besonders die Perspektive der Adressat:innen hat in Wirkungsfragen zu sozialen Programmen und Projekten Gewicht, da sie den Erfolg einer Maßnahme maßgeblich mitbestimmen (Lüders/Haubrich 2003: 326).

Welche Perspektiven relevant sind, entscheidet sich grundsätzlich am Erkenntnisinteresse der Evaluation. Konzentriert es sich darauf, welche Ergebnisse die Arbeit erzielt (sogenannte ergebnis- oder, enger gefasst, wirkungsorientierte Evaluationen), sind andere Perspektiven relevant als bei Evaluationsfragen, die auf die Arbeitsprozesse im Team fokussieren (prozessorientierte Evaluationen) (Döring 2019: 180). Für die Frage, wie wirksam ein Programm der politischen Bildung in der Schule ist, sind beispielsweise die Lernfortschritte und Wahrnehmungen der teilnehmenden Schüler:innen relevant, aber auch die Wahrnehmungen und Urteile von Lehrkräften, Schulsozialarbeiter:innen und Eltern. Geht es eher um Prozesse der (Zusammen-)Arbeit, sind die Perspektiven der aktiv Mitwirkenden aller beteiligten Organisationen gefragt, weniger jedoch die der Adressat:innen. Multiperspektivität ist in Kooperationskontexten, in denen unterschiedliche Institutionen das Programm oder Projekt gemeinsam umsetzen, besonders relevant. Dies betrifft viele Maßnahmen der Extremismusprävention, die häufig in einem Spannungsfeld zwischen zivilgesellschaftlichen und behördlichen Akteuren sowie deren spezifischen Zielen und Annahmen stehen (Möller et al. 2020; Ruf/Walkenhorst 2021).

Solange Perspektiven als die Sichtpunkte unterschiedlicher Akteure verstanden werden, funktioniert Multiperspektivität unabhängig von Multimethodik. Beispielsweise können Evaluierende die Blickpunkte verschiedener Personen mit der gleichen Methode erheben. Der Begriff „Perspektiven“ muss sich jedoch nicht auf die Sichtweisen von Akteuren beschränken. Er kann ebenso im Sinne unterschiedlicher methodischer Blickwinkel verstanden werden. Multimethodisch zu arbeiten bedeutet in diesem Sinne auch, multiperspektivisch zu arbeiten (Koch/Fertsch-Röver 2010; Klawe 2006). Das

Vorgehen, den Forschungsgegenstand aus verschiedenen Richtungen zu beleuchten, sei es mithilfe unterschiedlicher Methoden, Daten, Theorien oder Sichtweisen von Akteuren, wird als Triangulation bezeichnet (Flick 2008: 12; Weine 2015; Flick 2017).

2.3 DIE VERBREITUNG VON MULTIMETHODIK IN DER EVALUATIONSPRAXIS

Zwar kann sich die Umsetzung multimethodischer und multiperspektivischer Evaluationsdesigns für soziale Programme und Projekte in vielen Situationen entsprechend der Gegenstandsangemessenheit empfehlen, doch ihr aufwändiger Einsatz ist auch mit den Erkenntnisinteressen sowie begrenzten zeitlichen und finanziellen Ressourcen und Datenzugängen abzuwägen.

Bisherige empirische Übersichtsarbeiten deuten entsprechend darauf hin, dass Multimethodik in der Evaluation in begrenzter Reichweite Verwendung findet. Galport und Galport (2015) zeigen, dass sich 11% der Artikel des *American Journal of Evaluation* mit Multimethodik beschäftigten. In einer Übersichtsarbeit von Haunberger und Baumgartner (2017) ist die Bereitschaft zur Methodenkombination stark abhängig von der gewählten Evaluationsform. Die Autor:innen legen dar, dass 69% von 33 *Realist Evaluations* im Bereich der sozialen Arbeit bis Mitte 2013 multimethodisch ausgerichtet waren. 42% kombinierten hierbei qualitative Methoden untereinander, während 27% *Mixed Methods* anwendeten.

Ähnlich sieht es im Feld der Extremismusprävention auf internationalem Level aus. So stellen Feddes und Gallucci (2015) in einer Übersichtsarbeit zu 135 zwischen 1990 und 2014 veröffentlichten Evaluationen von Präventions- und Deradikalisierungsprogrammen fest, dass in 21% der Studien mehrere Erhebungsinstrumente und in 4% der Arbeiten mehrere Methoden genutzt wurden. Auch schreiben Pistone und Kolleg:innen (2019), basierend auf einer Übersicht von 112 Evaluationen aus den Jahren 1989 bis 2017, dass nur vier von 38 Evaluationsstudien mehrere (qualitative und quantitative) Methoden kombinierten.

Der vorliegende Report geht über diese Übersichten hinaus. Er analysiert, mit welchen Methodenkombinationen Multimethodik in verschiedenen Fachdisziplinen umgesetzt wurde, und ob zusätzlich dazu die Perspektiven mehrerer relevanter Personengruppen erhoben wurden. Multimethodik verstehen wir dabei als die Verwendung von mindestens zwei Erhebungsmethoden; Multiperspektivität verstehen wir als den Einbezug von mindestens zwei verschiedenen Akteursgruppen in der Beantwortung der Evaluationsfragestellungen. In der Übersicht beziehen wir keine sogenannte ‚graue Literatur‘ mit ein, also Literatur, die nicht in wissenschaftlichen Zeitschriften oder als Dissertation von einer Universität, sondern über eigene Wege publiziert wurde, obwohl viele deutsche Evaluationsberichte auf diesem Wege veröffentlicht werden. Die Gründe sind, dass ‚graue Literatur‘ schwieriger mit systematischen Suchweisen online auffindbar ist und dass die Berichtslegung in Evaluationsberichten häufig keine klar abgegrenzte Methodensektion enthält. Der Schwerpunkt liegt darin mehr auf den praxisbezogenen Schlussfolgerungen aus der Evaluation. Diese ist für den anwendungsbezogenen Zweck der Berichte überaus sinnvoll, hätte eine verglei-

chende Analyse wie im vorliegenden Report aber deutlich erschwert. Künftige Analysen sollten (auch) die Methodik von in ‚grauer Literatur‘ publizierten Evaluationen in den Blick nehmen.

3. METHODISCHES VORGEHEN

Für den vorliegenden Überblick haben wir eine breit angelegte Recherche nach Evaluationsstudien in den Sozialwissenschaften durchgeführt, die multimethodisch vorgehen. Die identifizierte Gesamtheit an Studien haben wir nach Themenfeldern kategorisiert und anhand methodischer Kriterien analysiert: Kodiert wurden die Art und Menge der Erhebungsmethoden, ob quantitativ und/oder qualitativ gearbeitet wurde, ob Prozesse oder Wirkungen analysiert wurden und ob ein experimentelles Design verwendet wurde. Zusätzlich haben wir die Anzahl an erhobenen Akteursperspektiven kodiert, die bei der Datenerhebung Berücksichtigung finden.

3.1 GRUNDGESAMTHEIT, AUSWAHL UND KATEGORISIERUNG DER UNTERSUCHUNGSEINHEITEN

Um das multimethodische Vorgehen in sozialwissenschaftlichen Evaluationen zu analysieren, wurden aus wissenschaftlichen Literaturdatenbanken relevante Evaluationsstudien anhand verschiedener Evaluationskriterien erfasst. Die Grundgesamtheit bildeten alle in den letzten zehn Jahren (2010–2020) in sozialwissenschaftlichen Fachzeitschriften veröffentlichten Publikationen sowie Dissertationen in deutscher oder englischer Sprache, die eine Evaluation zum Ziel haben und multimethodisch vorgehen.

Wie oben beschrieben, wurde sogenannte ‚graue Literatur‘ in der Grundgesamtheit nicht berücksichtigt, auch wenn davon auszugehen ist, dass diese gerade im Bereich der Extremismusprävention eine erhebliche Rolle spielt. Wenn sich die Suche nach ‚grauer Literatur‘ nur auf ein sehr abgegrenztes Forschungsfeld bezogen hätte und dabei relevante Datenbanken, Organisationen oder Suchmaschinen identifizierbar gewesen wären, wäre ihr Einbezug durchaus praktikabel gewesen. Unsere Zielsetzung war jedoch, das multimethodische und multiperspektivische Vorgehen in der Evaluationsforschung auf breiter Basis zu erheben und zwischen verschiedenen Fachbereichen zu vergleichen. Die Einbeziehung ‚grauer Literatur‘ hätte zur Folge gehabt, aufgrund der notwendigen umfangreichen manuellen Einflussnahme von einer systematischen und komparativen Suche abzuweichen.

Für die Erhebung der Grundgesamtheit wurden drei gängige und umfassende Literaturdatenbanken herangezogen: Über die Metadatenbank *EBSCOhost* wurden die Datenbanken *Education Research Complete*, *PsychINFO* und *PsychArticles* erfasst. Über den Onlinekatalog der Bayerischen Staatsbibliothek (BSB) München wurde der Recherchedienst „Aufsätze und mehr“ erschlossen, der ca. eine Milliarde Nachweise von elektronischen Zeitschriften und E-Books aufweist. Ergänzend dazu wurde noch das *Social Science Open Access Repository (SSOAR)* der *GESIS* herangezogen. In diesen Datenbanken wurde nach multimethodischen Evaluationsstudien in englischer und deutscher Sprache gesucht (vgl. Suchwortkombinationen in Tabelle A, Anhang). Die erhobenen Studien wurden im

Literaturverwaltungsprogramm *Citavi* erfasst, die Suche ergab insgesamt $N = 8.803$ Studien, ohne Dopplungen (Studien, die in mehreren Datenbanken gelistet werden).

Im Anschluss an die Erfassung wurden die Studien anhand ihrer Abstracts auf ihre tatsächliche Relevanz für die Analysen geprüft. Um in die Analyse einbezogen zu werden, mussten drei Kriterien erfüllt sein: (1) Die Studie musste empirisch sein, d.h., es mussten Daten erhoben worden sein, (2) sie musste nach unserer Definition multimethodisch sein, indem mindestens zwei Datenerhebungsmethoden angewendet wurden, (3) es musste eine Evaluation stattgefunden haben, wobei der Evaluationsbegriff hier sehr breit definiert wurde und nur reine Entwicklungen von Evaluationsinstrumenten (Messskalen) ausgeschlossen wurden. Sobald Daten erhoben und erfasst wurden und die Bezeichnung Evaluation verwendet wurde, wurde die Studie als Evaluation anerkannt. Im Zuge der Prüfung dieser Relevanzkriterien wurden 4.135 Studien aus der Analyse ausgeschlossen, letztlich blieb eine Liste von $n_G = 4.668$ Studien als Untersuchungseinheiten für alle weiteren Analysen.

In einem nächsten Schritt wurden die Studien induktiv in fünf thematische Grobkategorien eingeordnet: (1) Bildung, Schule und Erziehung, (2) Gesundheit und Soziales, (3) Politik, Verwaltung und Organisation, (4) Kommunikation und Medien sowie (5) Kriminalität, Sicherheit und Justiz. Jede Studie wurde nur einem Themenfeld zugeordnet. Diese Zuordnung ermöglichte es, tendenzielle multimethodische Evaluationskulturen zwischen Themenfeldern zu vergleichen. Einzelne Studien mit Schnittmengen zu mehreren dieser Themenfelder wurden dem Themenfeld zugeordnet, zu dem die Studie nach unserer Meinung die größte Schnittmenge hatte. Zusätzlich wurden aufgrund des Themenfokus des *PrEval*-Projekts in den Kategorien (1) Bildung, Schule und Erziehung sowie (5) Kriminalität, Sicherheit und Justiz jeweils eine Unterkategorie gebildet: (1.1) Politische Bildung und (5.1) Extremismusprävention. Studien, die diesen beiden Unterkategorien zugeordnet wurden, blieben auch dem übergeordneten Themenfeld zugeordnet.

Die Verteilung der Studien (Tabelle 1) zeigt, dass die meisten Studien zu fast gleichen Anteilen an der Gesamtstichprobe aus den Themenfeldern (1) Bildung, Schule und Erziehung sowie (2) Gesundheit und Soziales stammten. Beide recht breiten Themenfelder konnten daher einen relativ umfassenden Überblick über das dortige multimethodische Vorgehen in Evaluationsstudien in ihren Feldern liefern. Nur wenige Studien konnten in den Feldern (4) Kommunikation und Medien sowie (5) Kriminalität, Sicherheit und Justiz erfasst werden.

Aufgrund der hohen Relevanz der Extremismusprävention für das *PrEval*-Projekt und der geringen Anzahl der erfassten Studien aus diesem Themenfeld ($n_e = 8$) wurde eine weitere, stärker auf die Extremismusprävention zugeschnittene Datenbanksuche vorgenommen, um sicherzustellen, dass keine relevanten Studien übersehen wurden.⁴ Mit Hilfe des Recherchedienstes „Aufsätze & mehr“ im Onlinekatalog der BSB München wurde daher mit der Suchkombination „extremism UND evalua“ (für den Zeitraum 2010 bis 2020) nach Studien gesucht, die den oben genannten Einschlusskriterien entspre-

⁴ Dass mit engeren Suchbegriffen im Bereich der Extremismusprävention mehr Suchtreffer erzielt wurden als zuvor, liegt an den Logiken der Datenbankrecherche, die die eingegebenen Suchworte nur in bestimmten Teilen eines Aufsatzes (z.B. im Abstract, nicht aber im Volltext) sucht.

chen, sich aber auch explizit der Evaluation von Extremismusprävention widmen. Diese Suche ergab weitere 4.512 Suchtreffer, wobei davon nur 22 den genannten Relevanzkriterien entsprachen. Insgesamt ergab sich also im Themenfeld (5.1) eine Grundgesamtheit von 30 relevanten Studien.

THEMENFELD	N (ANZAHL DER ERHOBENEN STUDIEN)
(1) Bildung, Schule und Erziehung	$n_b = 2.031$
(1.1) Politische Bildung	$n_{pb} = 40$
(2) Gesundheit und Soziales	$n_{gs} = 2.165$
(3) Politik, Verwaltung und Organisation	$n_o = 251$
(4) Kommunikation und Medien	$n_k = 77$
(5) Kriminalität, Sicherheit und Justiz	$n_j = 142$
(5.1) Extremismusprävention	$n_e = 8$

Tabelle 1: Anzahl der erhobenen multimethodischen Evaluationsstudien (2010–2020) in den einzelnen Themenfeldern. Die Abweichung von $n = 2$ der Summe der Kategorien von der Gesamtzahl von 4.668 entsteht, da zwei der Studien aus der Vollerhebung keiner dieser Kategorien zugeordnet werden konnten.

3.2 STICHPROBEN UND KODIERUNG DER ANALYSEKRITERIEN

Im nächsten Schritt wurden zufällige Stichproben der Studien aus den Grobkategorien gezogen, wobei sich die Stichprobe nach der Anzahl der Studien in der jeweiligen Grobkategorie richtete (Tabelle 2). Insgesamt sollten etwa 5% aller erhobenen Studien (Annäherungswert 234 Studien) mit Blick auf ihr methodisches Vorgehen analysiert werden. Diese 5% sollten so auf die einzelnen Themenfelder verteilt werden, dass sie in etwa die Größe der Themenfelder in der Grundgesamtheit repräsentieren. Da der Anteil der Studien im Themenfeld (1) Bildung, Schule und Erziehung etwa 43% der Grundgesamtheit von 4.668 Studien ausmachte, sollte der Stichprobenanteil des Themenfeldes auch ca. 43% der Gesamtstichprobe von 234 Studien repräsentieren, was in etwa 100 Studien entsprach. Die Stichprobengröße sollte allerdings in jedem Themenfeld mindestens $n = 20$ groß sein, sodass bei den Themenfeldern (3) Politik, Verwaltung und Organisation, (4) Kommunikation und Medien sowie (5) Kriminalität, Sicherheit und Justiz jeweils 20 Studien zufällig gezogen wurden. Die Unterkategorien (1.1) Politische Bildung und (5.1) Extremismusprävention wurden aufgrund ihrer Relevanz für *PrEval* in ihrer Gänze analysiert. Die zufälligen Stichproben wurden innerhalb der einzelnen Themenfelder anhand eines Zufallszahlengenerators⁵ gezogen.

Im Anschluss daran wurden für alle Stichproben die Volltexte erfasst. Falls für eine Studie kein Volltext vorhanden war, wurde eine weitere Studie nachgezogen. Es zeigte sich, dass dies gehäuft bei Dissertationsprojekten vorkam und seltener bei in Zeitschriften veröffentlichten Texten. In den Kate-

5 Online unter: <https://www.matheretter.de/rechner/zufallszahlengenerator> (letzter Zugriff am 14.10.2021).

gorien (1.1) Politische Bildung und (5.1) Extremismusprävention war keine Nachziehung möglich, da es sich bei diesen bereits um Vollerhebungen handelte. Im Zuge der Volltextanalyse stellte sich heraus, dass einzelne Texte den Relevanzkriterien für die Studie nicht genügten und zum Beispiel nicht tatsächlich multimethodisch durchgeführt wurden. In diesen Fällen wurden keine Artikel nachgezogen, da sich durch den Wegfall auch die Grundgesamtheit der Studien verringerte. Die tatsächliche Stichprobengröße ist daher geringer als die ursprünglich angestrebte (siehe Tabelle 2).

THEMENFELD	N (GRUND- GESAMTHEIT)	ANZAHL DER ZUFÄLLIG GEZOGENEN TEXTE	STICHPROBEN- GRÖSSE*
(1) Bildung, Schule und Erziehung	$n_{b(ohne pb)} = 1.991$	100	96
(1.1) Politische Bildung	$n_{pb} = 40$	39 (Vollerhebung)**	28
(2) Gesundheit und Soziales	$n_{gs} = 2.165$	134	110
(3) Politik, Verwaltung und Organisation	$n_o = 251$	22	20
(4) Kommunikation und Medien	$n_k = 77$	22	20
(5) Kriminalität, Sicherheit und Justiz	$n_{j(ohne e)} = 134$	22	20
(5.1) Extremismusprävention	$n_e = 30$	30 (Vollerhebung)	23

Tabelle 2: Stichprobengrößen der einzelnen Themenfelder.

*Die Stichprobengröße weicht von der Anzahl der zufällig gezogenen Texte ab, da für Volltexte, die den Relevanzkriterien nicht entsprachen, keine Texte nachgezogen wurden.

**Die Vollerhebung weicht minimal von N (Grundgesamtheit) ab, da einmal kein Volltext erfasst werden konnte.

Die Volltexte der Studien in diesen Stichproben ($n = 317$) wurden auf das Evaluationsvorgehen und insbesondere das multimethodische Vorgehen der Studien analysiert. Die Kodierregeln für die kodierten Merkmale (Jahr der Publikation, Anzahl und Art der verwendeten Erhebungsmethoden, Methodendesign (quantitativ und/oder qualitativ), Multiperspektivität, Art der Evaluation, experimentelles Design) sowie ihre Ausprägungen finden sich im Anhang (Tabelle B).

4. ERGEBNISSE

Im Folgenden werden die Ergebnisse der Analyse vorgestellt und das methodische Vorgehen der Studien in den jeweiligen Themenfeldern beschrieben. Als ein zentrales Ergebnis lässt sich vorweg ein Aufwärtstrend in Bezug auf die Anzahl an Studien über den untersuchten Zeitraum hinweg (2010 bis 2020) nennen. Hier ist in fast allen Themenbereichen eine Zunahme an Artikeln festzustellen, die den Suchkriterien entsprechen. Eine Ausnahme davon bildet die politische Bildung. Des Weiteren ist ein starker Fokus auf *Mixed Method*-Designs mit zwei Erhebungsmethoden auszumachen, wobei der Schwerpunkt auf Formen der Befragung liegt.

Zur Untersuchung dessen, was in den vorliegenden Studien unter dem Begriff multimethodisch umgesetzt wurde, wurde in einem ersten Schritt erfasst, ob die betreffende Evaluation mit zwei, drei, vier oder mehr als vier Methoden durchgeführt wurde. Wie Abbildung 1 zeigt, wurde am häufigsten mit einer Kombination von zwei Methoden gearbeitet. Dies ist in allen untersuchten Themenbereichen der Fall. In etwa jeweils der Hälfte der untersuchten Artikel wurden zwei Methoden angewandt, mehr als vier Methoden wurden selten kombiniert. Feldunterschiede sind hierbei kaum vorhanden. Erfasst wurde ebenfalls, ob ein rein quantitatives Methodendesign, ein rein qualitatives oder eine Kombination aus beidem, also ein *Mixed Method*-Design, verwendet wurde. Überwiegend und über alle Themenfelder hinweg ist Letzteres der Fall – im Bereich (1) Bildung, Schule und Erziehung beispielsweise zu knapp 89%, im Bereich (1.1) Politische Bildung zu 96% und im Bereich (5) Kriminalität, Sicherheit und Justiz zu 90%. Betrachtet man die Anzahl der verwendeten Methoden in Kombination mit der Frage nach der Art des Designs, so zeigt sich, dass die häufigste Variante in allen Themenbereichen ein quantitativ-qualitatives Zwei-Methoden-Design ist.

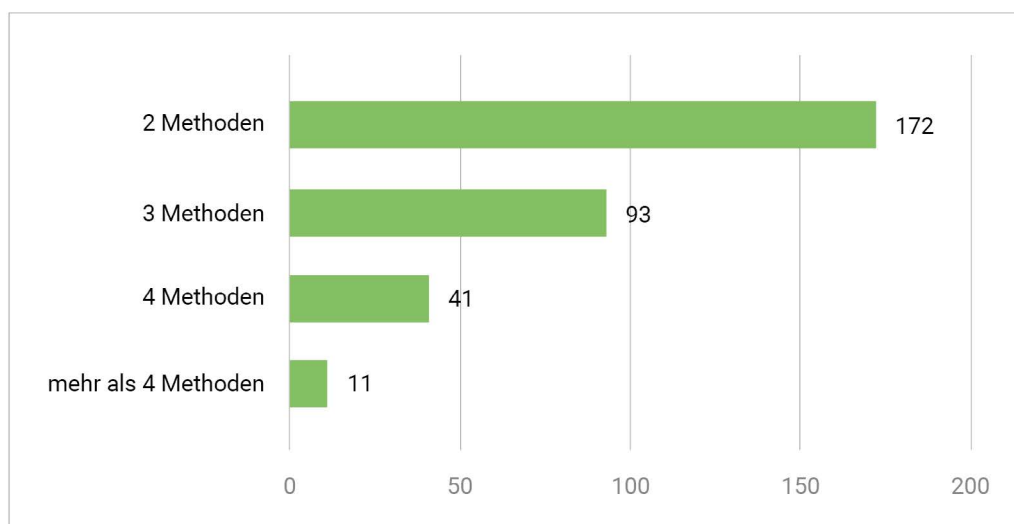


Abb. 1: Anzahl der verwendeten Methoden in den analysierten Artikeln ($n = 317$), Quelle: eigene Darstellung.

Betrachtet man die Erhebungsmethoden genauer, lässt sich ein Schwerpunkt über alle analysierten Themenfelder hinweg erkennen (Abbildung 2). Formen der Befragung (Fragebogen/Survey, Interview und Fokusgruppe) dominieren, sie machen in allen Themenbereichen die oder weit über die Hälfte der genutzten Methoden aus. Andere sozialwissenschaftliche Erhebungsmethoden wie Dokumentenanalyse oder Beobachtung kommen sehr viel seltener zum Einsatz. Feldunterschiede sind zwar vorhanden, allerdings in geringem Umfang.

Während die Befragung als Form der Erhebungsmethode in allen Themenfeldern klar an erster Stelle der gewählten Methoden steht, unterscheidet sich die am zweithäufigsten verwendete Methode von Themenfeld zu Themenfeld deutlich. Abbildung 3 zeigt, dass beispielsweise in den Bereichen (3) Politik, Verwaltung und Organisation, (5) Kriminalität, Sicherheit und Justiz sowie (5.1) Extremismusprävention die am zweithäufigsten angewendete Methode eine Dokumentenanalyse (z.B. quantitative oder qualitative Inhaltsanalyse) ist. Im Bereich (1.1) Politische Bildung ist es die Beobachtung und im Bereich (4) Kommunikation und Medien sowie im Bereich (1) Bildung, Schule und Erziehung sind es „sonstige Methoden“. Bei den „sonstigen Methoden“ lassen sich starke Fachspezifika identifizieren. Gemeint sind zum Beispiel Leistungstests im Themenfeld (1) Bildung, Schule und Erziehung oder diagnostische Tests im Feld (2) Gesundheit und Soziales. Betrachtet man, wie die verwendeten Methoden kombiniert werden, so lassen sich besonders häufig Kombinationen verschiedener Befragungsformen ausmachen. Im Weiteren werden Befragungen vor allem mit Dokumentenanalysen oder mit sonstigen Methoden kombiniert. Im Themenbereich (1) Bildung, Schule und Erziehung wurden drei Studien identifiziert, die ausschließlich sonstige – meist fachspezifische – Methoden anwenden. Die Methode der Beobachtung kommt außerdem ausschließlich in Kombination mit anderen Methoden und nie singular vor.

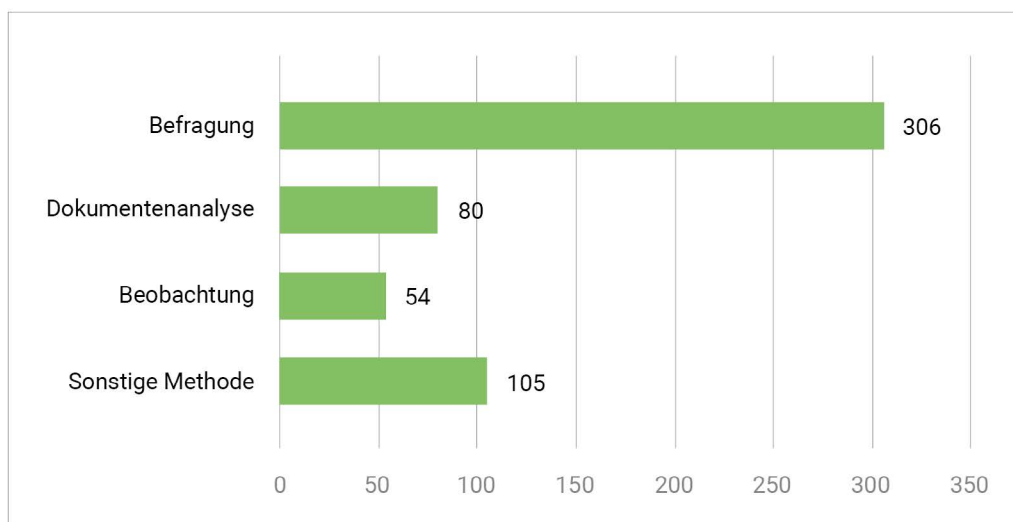


Abb. 2: Anzahl der Erhebungsmethoden über alle Themenfelder hinweg ($n = 317$), Quelle: eigene Darstellung.

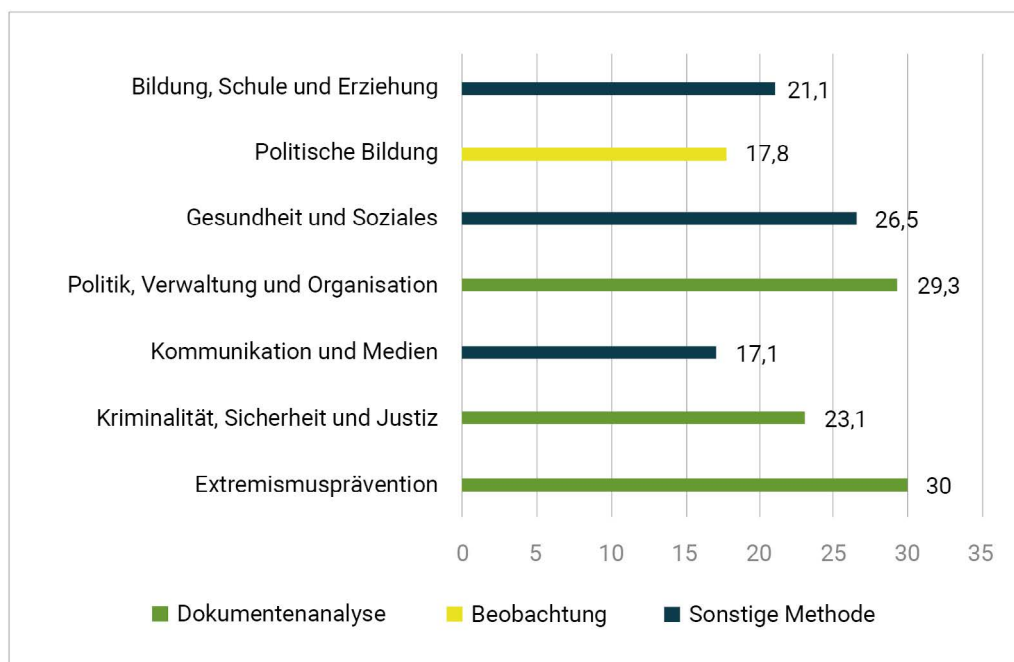


Abb. 3: Zweithäufigste Erhebungsmethode nach Themenfeld in Prozent ($n = 317$), Quelle: eigene Darstellung.

Ein detaillierterer Blick in die Themenfelder verdeutlicht einige Unterschiede in der Art, wie welche Methoden kombiniert und wie häufig diese Kombinationen verwendet werden. So zeigt Abbildung 4, dass der prozentuale Anteil an reinen Befragungen oder der Kombination mit Dokumentenanalysen von Themenfeld zu Themenfeld variiert. Aufgrund der unterschiedlichen Stichprobengrößen sind die Anteile allerdings nicht zwischen den Themenbereichen vergleichbar, sondern nur innerhalb eines Themenbereichs und zwischen den Bereichen mit großen bzw. kleinen Stichproben. Es zeigt sich vor allem, dass die zahlenmäßig vergleichbaren Bereiche (1) Bildung, Schule und Erziehung sowie (2) Gesundheit und Soziales in Bezug auf die Methodenkombination ähnlich aufgestellt sind – jeweils über 30% aller Artikel, die ausschließlich Befragungsvarianten aufweisen, sind hier zu verorten. Hingegen ist es in den anderen Themenbereichen, außer im Bereich (1.1) Politische Bildung, genau umgekehrt – auf diese Bereiche entfallen jeweils größere Anteile an Artikeln, die die Methodenkombination Befragung und Dokumentenanalyse aufweisen.

Dass in den multimethodischen Evaluationsstudien im Allgemeinen eher wenig Variation an Erhebungsmethoden vorhanden ist, setzt sich als Trend auch im Speziellen fort. In der Analyse haben wir zwischen den Befragungsformen Surveys, Interviews und Fokusgruppen unterschieden, wobei wir unter Surveys Befragungen mittels (stark) standardisierter Fragebögen verstanden haben. In Abgrenzung dazu wurden unter Interviews nicht- bzw. maximal teilstandardisierte (mittels Leitfäden durchgeführte) Befragungen gefasst. Deutlich wird, dass innerhalb der Befragung als Methode vor allem standardisierte Fragebögen zum Einsatz kommen (Abbildung 5). Innerhalb der

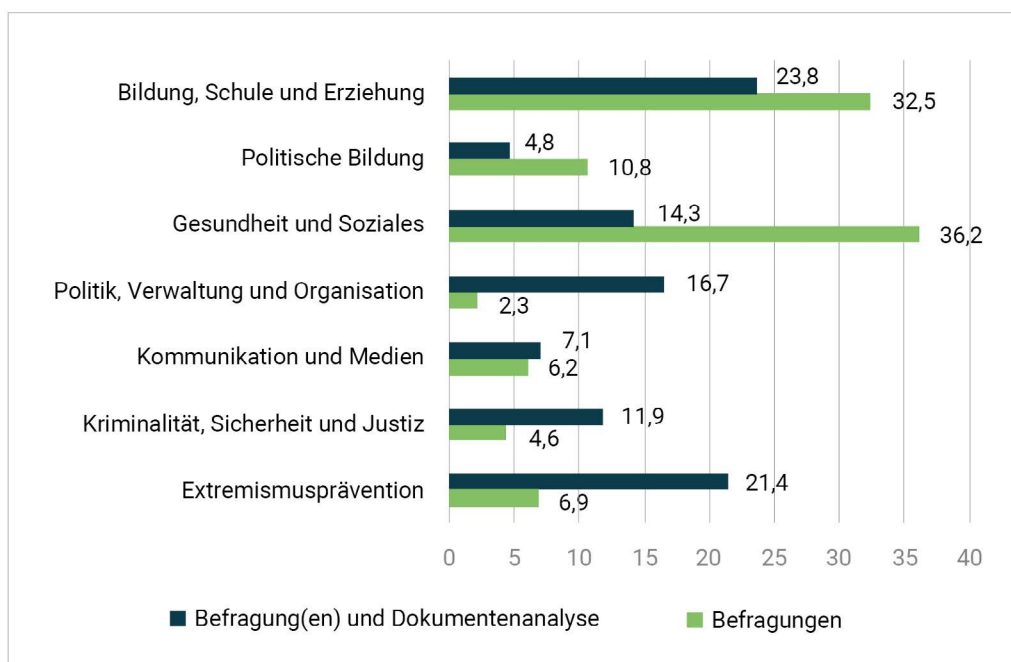


Abb. 4: Prozentualer Anteil der Methodenkombinationen „Befragungen“ und „Befragung(en) + Dokumentenanalyse“ je Themenbereich. Vergleich aufgrund unterschiedlicher Stichprobengrößen nur innerhalb eines Themenbereichs (n = 317), Quelle: eigene Darstellung.

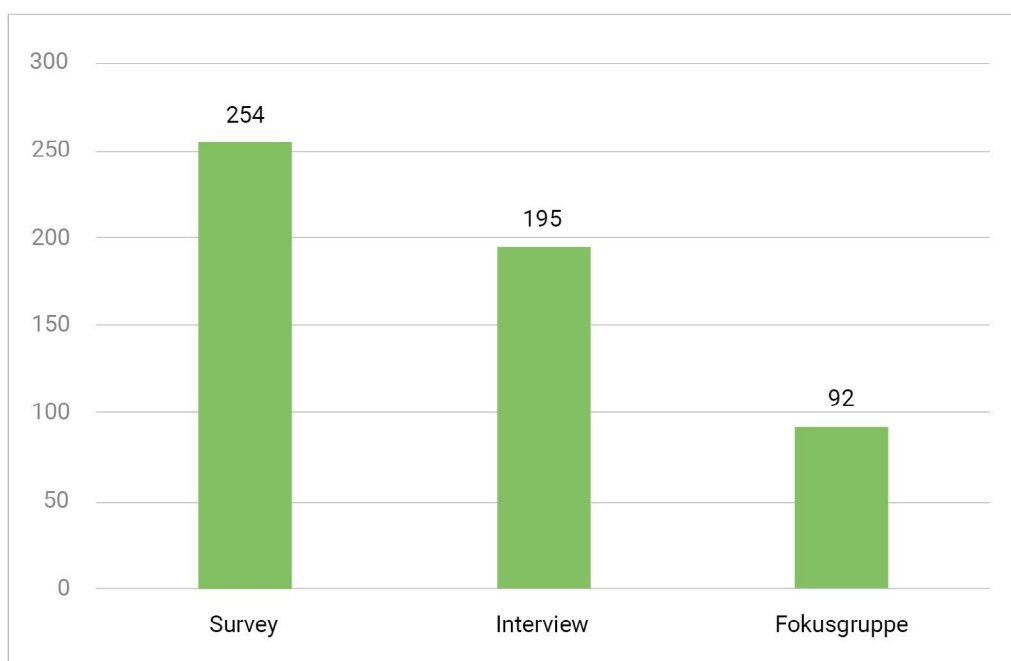


Abb. 5: Anzahl der unterschiedlichen Befragungsformen (n = 317, Mehrfachcodierung möglich), Quelle: eigene Darstellung.

Themenbereiche sind hier keine auffälligen Unterschiede zu erkennen – der starke Fokus auf standardisierte Befragungen zeigt sich überall.

Analog zur genauen Betrachtung der Methodenkombinationen wurde auch analysiert, in welchen Kombinationen Befragungsvarianten auftraten. Reine Survey-Studien entfallen zu einem großen Teil auf den Themenbereich Bildung; in allen anderen Themenbereichen ist die Kombination „Surveys und Interviews“ die häufigste.

Neben dem Fokus auf Multimethodik lag ein zweiter großer Analyseschwerpunkt auf der Multiperspektivität und der Frage, ob eine solche Teil der untersuchten Evaluationsstudien war. Multiperspektivität könnte beispielsweise in der Form abgebildet sein, dass die Erhebungen unterschiedliche Zielgruppen berücksichtigten. In 41% der untersuchten Stichprobe (130 von 317 Fällen) konnte ein multiperspektivisches Vorgehen festgestellt werden. Wie Abbildung 6 zeigt, wiesen davon 74 Studien (58%) zwei Perspektiven, 38 Studien (29%) drei Perspektiven und 12 Studien (9%) vier Perspektiven auf. Mehr als vier Perspektiven gab es in sechs Studien (knapp 5%). Konnte Multiperspektivität festgestellt werden, beschränkte sich also mehr als die Hälfte der Beiträge auf zwei Perspektiven, immerhin 43% verwendeten jedoch drei oder mehr Perspektiven; bezogen auf die Gesamtanzahl der untersuchten Artikel ($n = 317$) sind es knapp 18% der Artikel, die drei oder mehr Perspektiven einholen. Der prozentuale Anteil an Artikeln, die drei oder mehr Perspektiven einholen, variiert nur in Nuancen zwischen den Themenbereichen. Am stärksten ausgeprägt ist er im Bereich Extremismusprävention, wo in einem knappen Drittel aller analysierten Artikel drei und mehr Perspektiven kombiniert werden (30%, $n = 7$).

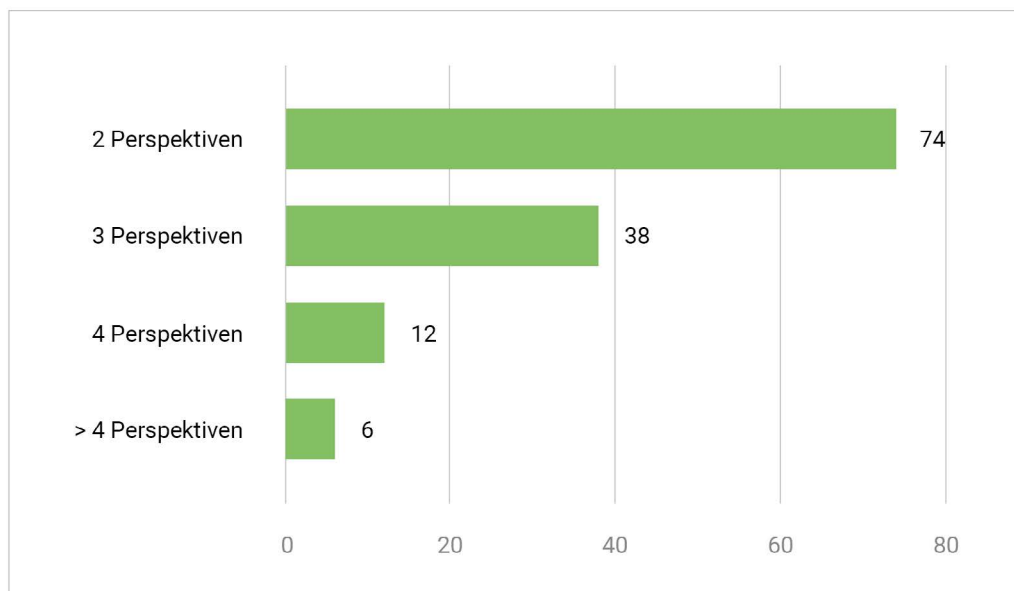


Abb. 6: Anzahl der erhobenen Perspektiven in multiperspektivischen Studien ($n = 130$), Quelle: eigene Darstellung.

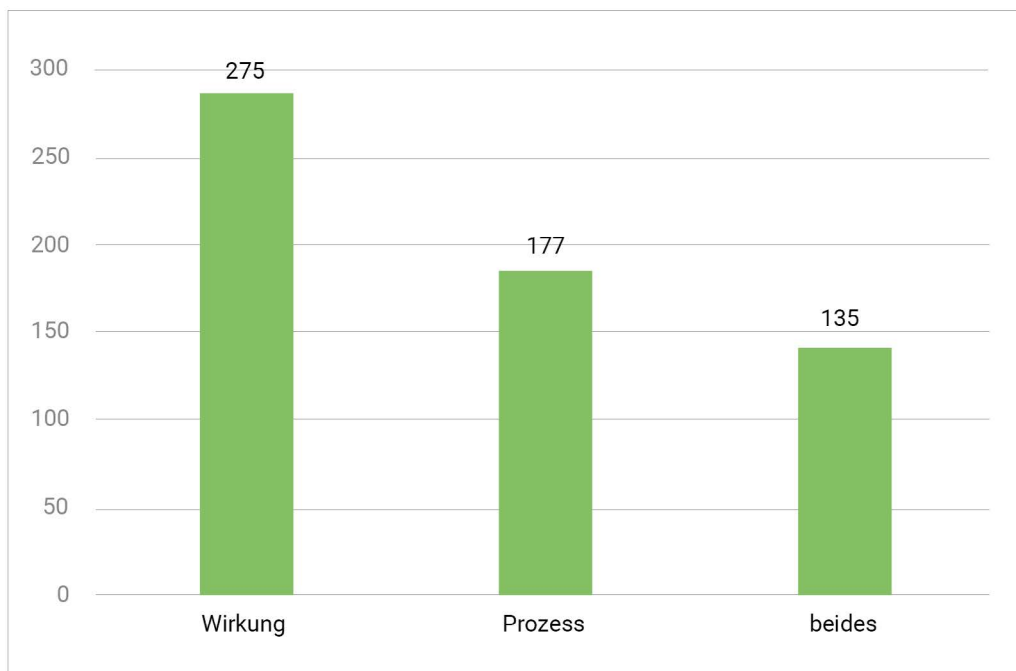


Abb. 7: Anzahl der Artikel, in denen eine Wirkungs- oder Prozessevaluation bzw. eine Kombination aus beidem durchgeführt wird ($n = 317$), Quelle: eigene Darstellung.

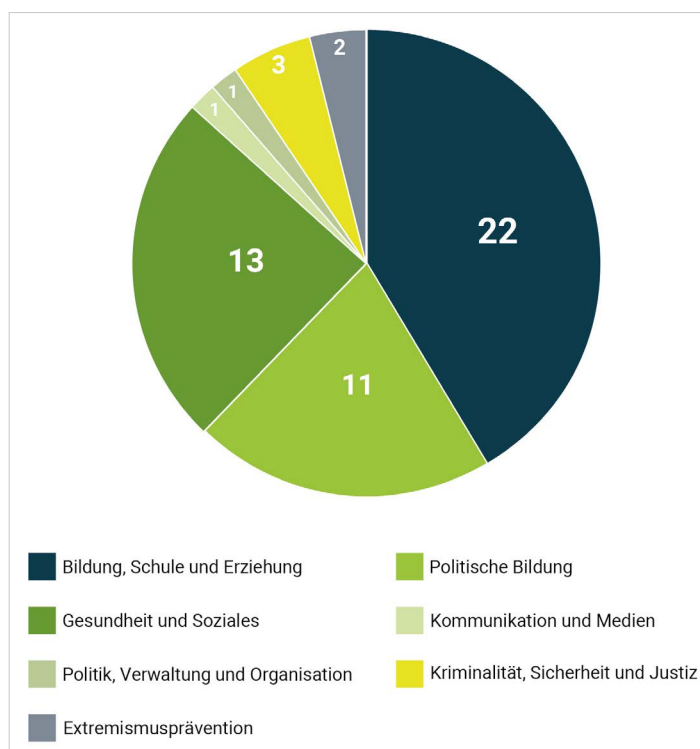


Abb. 8: Experimentelles Design nach Themenbereichen ($n = 53$), Quelle: eigene Darstellung.

Durch die Feinanalyse der Evaluationsstudien konnten weitere Ergebnisse in Bezug auf die Form der Evaluation herausgearbeitet werden. Hier ist ein deutlicher Schwerpunkt bei der Wirkungsevaluation auszumachen. Kaum eine der Studien kam ohne eine Form der Wirkungsevaluation aus. Die Prozessevaluation wurde in knapp der Hälfte aller Studien zusätzlich zur Wirkungsanalyse durchgeführt (Abbildung 7).

Im letzten Schritt der Untersuchung wurde außerdem erfasst, ob das Design experimentell angelegt war – unserer Definition nach musste hierzu eine Kontrollgruppe vorhanden sein. Überwiegend ist dies nicht der Fall, insgesamt konnte nur in 53 Studien ein experimentelles Design festgestellt werden (siehe Abbildung 8). Davon sind elf Artikel (39%) dem Bereich (1.1) Politische Bildung zuzuordnen und 22 Artikel (23%) dem Bereich (1) Bildung, Schule und Erziehung. In den anderen Themenbereichen weisen nur vereinzelte Artikel dieses Merkmal auf, im Bereich (5.1) Extremismusprävention sind es beispielsweise lediglich zwei Artikel, also knapp 9%.

5. DISKUSSION

Der vorliegende Report erfasst das methodische Vorgehen in wissenschaftlich publizierten sozialwissenschaftlichen Evaluationsstudien, die sich selbst als multimethodisch bezeichnen. Ein besonderes Augenmerk liegt dabei auf der Verwendung und Kombination unterschiedlicher Datenerhebungsmethoden (Multimethodik) und der Erhebung von Daten aus unterschiedlichen Akteursperspektiven (Multiperspektivität). Zu diesem Zweck wurde eine Vollerhebung von multimethodischen Evaluationsstudien durchgeführt, die im Zeitraum von 2010 bis 2020 in wissenschaftlichen Publikationen veröffentlicht wurden. Diese Vollerhebung wurde in verschiedene thematische Kategorien aufgeteilt: (1) Bildung, Schule und Erziehung (2) Gesundheit und Soziales, (3) Politik, Verwaltung und Organisation, (4) Kommunikation und Medien, (5) Kriminalität, Sicherheit und Justiz. Eine repräsentative Stichprobe von Studien aus allen Themenfeldern wurde auf Merkmale der Multimethodik und der Multiperspektivität analysiert. Ausgehend davon sollen im Folgenden Erkenntnisse diskutiert werden, die für Evaluationen in den Sozialwissenschaften genutzt werden können, sowohl im Sinne einer Bestandsaufnahme als auch im Sinne der Umsetzbarkeit von Multimethodik und -perspektivität in der Evaluationspraxis.

Insgesamt wuchs die Anzahl multimethodischer Evaluationsstudien pro Jahr über alle thematischen Felder hinweg von 2010 bis 2020 an. Zwei Felder ((1) Bildung, Schule und Erziehung sowie (2) Gesundheit und Soziales) machten einen großen Teil der Gesamtstichprobe aus. Die Analysen zeigten, dass sich die Umsetzung von Multimethodik bei der großen Mehrzahl der Studien auf die Kombination unterschiedlicher Befragungsmethoden, meist Interviews und Surveys, beschränkte. Viel seltener wurden Dokumentenanalysen, teilnehmende Beobachtungen oder andere Datenerhebungsmethoden angewendet. Ferner wurden zumeist zwei Methoden (und selten mehr) miteinander kombiniert. Themenfeldspezifische Unterschiede existierten in unserer Stichprobe bezüglich der Umsetzung von multimethodischer Datenerhebung kaum. In den Bereichen (1) Bildung, Schule und Erziehung sowie (2) Gesundheit und Soziales waren allerdings fachspezifische Erhebungsmethoden wie Leistungs- oder Diagnostetests neben den Befragungen sehr üblich und wurden meistens mit Sur-

veys und/oder Interviews kombiniert. Unsere Analysen zur Multiperspektivität zeigten weiter, dass die Mehrzahl der Studien bei der Beantwortung der Evaluationsfragen nicht mehrere unterschiedliche Akteursperspektiven berücksichtigt. Wenn multiperspektivisch vorgegangen wird, dann in geringem Ausmaß mit lediglich zwei Akteursperspektiven.

Multimethodik wurde in unserer Untersuchung in den meisten Fällen im Sinne des *Mixed Method*-Ansatzes umgesetzt, in dem qualitative mit quantitativen Datenerhebungsmethoden (im vorliegenden Fall meistens Surveys mit Interviews) kombiniert wurden. Dies entspricht bisherigen Diskussionen und Beobachtungen in Forschung und Feld, in denen *Mixed Methods* häufig als Synonym von Multimethodik verstanden werden (Johnson et al. 2007). *Mixed Method*-Designs können je nach Ergebnissynthese und Interpretation eine besondere Möglichkeit für den Erkenntnisgewinn über den Gegenstand schaffen. Darüber hinaus können quantitative und qualitative Methoden unterschiedliche Mehrwerte für die Weiterentwicklung der praktischen Arbeit bieten: Während quantitative Erhebungen sich dazu eignen, ein breites, überblicksartiges Bild zu zeichnen oder Prozesse auf einfachem Wege längerfristig und standardisiert zu dokumentieren, können qualitative Methoden Umstände in ihrer Komplexität ausleuchten und somit beispielsweise Findungsprozesse im Team unterstützen. Auch im Sinne des Erkenntnisinteresses können sie unterschiedliche Rollen erfüllen, etwa wenn theorienbildende (meist qualitative Erhebungsmethoden) und darauf aufbauend hypothesentestende Methoden (oft quantitative Methoden) verwendet werden.

Als weiterer relevanter Faktor wurde der Fokus der Studien auf Wirkungsanalysen oder Prozessanalysen erfasst. Über die Themenfelder hinweg zeigte sich hier, dass fast alle Studien eine Wirkungsanalyse vornahmen und nur sehr wenige Studien ausschließlich Prozessanalysen durchführten. Jedoch wurden in beinahe allen Themenfeldern in fast der Hälfte der Studien Wirkungs- und Prozessanalysen miteinander kombiniert.

5.1 AUSBLICK AUF WEITERE ASPEKTE DER EVALUATIONSFORSCHUNG

Unsere Analyse ist als erste Orientierung zu den tatsächlich möglichen Umsetzungen von Multimethodik und Multiperspektivität in der Evaluationsforschung zu werten und gibt Auskunft über die Nutzung von Methoden und ihrer Kombination. Künftige Arbeiten können diese Analyse etwa im Bereich der Extremismusprävention um ‚graue Literatur‘ ergänzen und einen noch detaillierteren Blick auf die Konzeptionsphase der Evaluationen werfen, um die Abwägungen für und gegen Multimethodik, die in der praktischen Umsetzung getroffen werden, zu beleuchten. Ebenso könnte über die deskriptive Ebene hinausgegangen und sowohl die Qualität der Umsetzung der genutzten Methoden(-kombinationen) als auch die Konsequenzen von Multimethodik und Multiperspektivität für die Güte der Evaluationen fokussiert werden (s.u.). Nicht zuletzt könnte beleuchtet werden, inwiefern Evaluierende (die Möglichkeit haben) Multimethodik als methodisches Gesamtkonzept (zu) handhaben, in dem jede Methode und ihre Kombination einen Sinn erfüllt und die Form ihrer Synthese vorab geplant und gewinnbringend eingebettet ist. Schließlich wäre zu untersuchen, in welchem Zusammenhang der Einsatz von Multimethodik mit verschiedenen (etwa summativen oder formativen) Evaluationszwecken steht.

Der vorliegende Report bietet zwar eine Übersicht über die verwendeten Methoden in multimethodischen Evaluationen, er trifft jedoch keine Aussage darüber, wie die Qualität dieser Evaluationen bewertet werden könnte. Künftige Analysen könnten diese Verknüpfung vertieft aufgreifen (Jugl et al. 2021). Als Gütekriterien für Evaluation können Evaluationsstandards herangezogen werden; prominente Beispiele sind hierfür die Standards der Deutschen Gesellschaft für Evaluation (DeGEval 2016). Diese 25 Standards sind vier Qualitätsdimensionen zugeordnet: Nützlichkeit (N), Durchführbarkeit (D), Fairness (F), Genauigkeit (G). Multimethodik zählt selbst nicht zu den Standards, könnte aber in einer gegebenen Anforderungskonstellation zu einigen Standards einen Beitrag leisten, etwa zur genauen und umfassenden Beschreibung des Evaluationsgegenstandes (G1) oder zur angemessenen Auswahl und zum Umfang der erfassten Informationen über den Gegenstand (N4). Multiperspektivität könnte ebenfalls zu Letzterem beitragen. Sie findet sich außerdem recht explizit im Standard F4, der unparteiischen Durchführung und Berichterstattung wieder: „Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation beachten. [...]“. Wie bereits diskutiert wurde, können sowohl Multimethodik als auch Multiperspektivität ebenso zur Gegenstandsangemessenheit beitragen.

Im Sinne der Gegenstandsangemessenheit muss sich die Auswahl der Erhebungsmethoden an den Evaluationszwecken, -fokussen und -fragestellungen orientieren. Wenn, wie im Feld der Extremismusprävention, die Grundlagenforschung sowie die Fachpraxis im Themenfeld (noch) über Konzepte und Begrifflichkeiten uneins sind (Walkenhorst 2019), kann dabei selten auf vorgefertigte Erhebungsinstrumente oder gar Hypothesen zurückgegriffen werden. Aufgrund der Komplexität und Einzigartigkeit von sozialen Programmen und Projekten muss sich die Konzeption einer jeden Evaluation außerdem den jeweils individuellen Ansprüchen annehmen. Dabei trifft das „Erwartungsdreieck“ aus Politik, Fachpraxis und Wissenschaft (im Sinne von externen Evaluierenden) aufeinander (Walkenhorst 2019). Die drei Seiten bringen oft unterschiedliche Erwartungen an Evaluation mit sich, die jeweils ihre eigene Berechtigung haben. Während die Politik beispielsweise auf Überprüfbarkeit der Wirksamkeit der Programme hofft, benötigt die Fachpraxis häufig eher formative Evaluationsformate, die sie in der Weiterentwicklung ihrer Arbeit unterstützt. Multiperspektivität – hier im Sinne der Einbindung verschiedener Akteure in der Konzeptionsphase der Evaluation – kann dabei helfen, ein gemeinsames Evaluationsziel zu formulieren. Wird die Perspektive der Fachpraxis einbezogen, werden die späteren Evaluationsergebnisse für die praktische Arbeit relevanter und sind damit einfacher in Weiterentwicklungen zu übersetzen (Möller et al. 2020). Ein breit aufgestellter Methodenkoffer erlaubt es Evaluierenden darüber hinaus, flexibel auf die Ansprüche der Akteure zu reagieren und Entwicklungen der Fachpraxis auch über veränderliche Bedarfe hinweg begleiten zu können.

Um all dies zu ermöglichen, sind jedoch ausreichend Ressourcen für Evaluation vonnöten, die auch Abstimmungen in der Planungsphase vorsehen. Künftige Analysen sollten einerseits beleuchten, welche Faktoren die Entscheidung von Evaluierenden für oder gegen Multimethodik und -perspektivität bedingen. Andererseits sollte untersucht werden, inwiefern diese Elemente Einfluss auf den Umgang mit dem „Erwartungsdreieck Evaluation“ (Walkenhorst 2019) und auf die Überführung der Evaluationsergebnisse in die Praxis haben. Die Möglichkeit zur Analyse all dieser Aspekte ist jedoch – solange sich einer Literaturübersicht bedient wird – vom Detailgrad der Beschreibung in den Evaluationsberichten abhängig. Andere meta-evaluative Ansätze könnten daher ebenfalls

zielführend sein, bei denen Evaluierende etwa über ihre Erfahrungen mit (multimethodischen) Evaluationen berichten. Entsprechende Ergebnisse aus dem *PrEval*-Projekt – zu Eindrücken der Fachpraxis – finden sich bei Koynova (2021), Quent (2021) sowie zusammengeführt bei Koynova et al. (2022). Reflexionen zu den *PrEval*-Pilotstudien, in denen mit Partner:innen aus der Präventionspraxis gemeinsam beispielhafte, multimethodische Evaluationsdesigns entwickelt wurden, werden in der vorliegenden Reportreihe zu finden sein.

6. FAZIT

Die Ergebnisse des vorliegenden Reports legen nahe, dass multimethodische Evaluationsforschung im Bereich sozialer Programme und Projekte den Kanon möglicher Erhebungsmethoden und deren Kombination tendenziell selten in seiner Breite nutzt. Multimethodische Evaluationen nutzen in der Regel zwei Methoden und dabei meist Befragungen. Die jeweilige Auswahl an Methoden muss an den konkreten Bedarfen der Evaluationsfragen orientiert sein; künftig wird also zu untersuchen sein, ob diese Umsetzung von Multimethodik (genügend) zur Gegenstandsangemessenheit der entsprechenden Evaluationen beiträgt. Ebenfalls wird meta-evaluativ zu untersuchen sein, inwiefern der Einsatz von Multimethodik zur Güte von Evaluationen (über die Gegenstandsangemessenheit hinaus) beiträgt und inwiefern sich ein Zusammenhang zwischen dem Einsatz multipler Methoden und der Nützlichkeit der Ergebnisse für die (Weiterentwicklung der) Praxis zeigt. Künftige Arbeiten sollten außerdem den Designprozess von Evaluationen empirisch beleuchten, denn der vorliegende Report erfasst nur die letztendlich sichtbare Evaluationsumsetzung. Diese ist jedoch die Konsequenz aus methodischen und pragmatischen Überlegungen, aus der Abwägung von idealen Evaluationsdesigns mit realen Umsetzungsbedingungen. Letztere zu verstehen ist elementar, um die Nutzung von Multimethodik in der Evaluation realistisch ermöglichen zu können, wo sie angebracht ist. Unter den hier untersuchten multimethodischen Evaluationen war der Einsatz von Multiperspektivität – im Sinne der Erhebung mehrerer Akteursperspektiven zur Beantwortung der Evaluationsfragen – begrenzt. Künftige Studien sollten auch hier analysieren, was die Gründe für und Konsequenzen von diesem Umstand sind. Es ist dabei nicht gesagt, dass Evaluationen, die nach der vorliegenden Definition nicht als multiperspektivisch galten, unweigerlich ein weniger vollständiges Bild ihrer Gegenstände zeichnen. Auch hier gilt der Grundsatz der Gegenstandsangemessenheit und die Erhebung mehrerer Perspektiven sollte sich nach dem jeweiligen Erkenntnisinteresse und den gegebenen Möglichkeiten richten. Dennoch sei an dieser Stelle die Relevanz der Einblicke aller beteiligten Akteure bei sozialen Programmen und Projekten betont. Relevante Akteursperspektiven nicht zu erheben, kann im schlimmsten Fall zu einer verzerrten Darstellung der Begebenheiten führen. Zum Beispiel werden Sozialarbeiter:innen in einer Justizvollzugsanstalt eine pädagogische Intervention unter Umständen anders bewerten als Vollzugsbeamte:innen, weshalb idealerweise beide Perspektiven einbezogen werden sollten.

Der vorliegende Report zeigt, in welchen Formen Multimethodik in der Praxis von Evaluierenden genutzt wird. Er gibt damit Hinweise auf die tatsächliche Anwendbarkeit von Multimethodik in der Evaluation. Für das Feld der Extremismusprävention, in dem eine sich stetig professionalisierende Evaluationspraxis zwar Multimethodik fordert (Milbradt 2018; Möller et al. 2020), aber (noch) nicht breit einsetzt (Feddes/Gallucci 2015; Möller et al. 2020), kann dies eine Orientierung dazu sein, wie

dem selbstgesteckten Anspruch begegnet werden kann. Evaluation ist hier wegen einer Kombination aus einem hohen öffentlichen und politischen Erwartungsdruck, erschwerten Datenzugängen und einer hohen Komplexität der behandelten Phänomene (im extremistischen Spektrum) voraussetzungsreich. Auch die involvierten Parteien (Fachpraxis, Wissenschaft, Behörden und Politik) bringen unterschiedliche Erwartungen mit an das, was Evaluation leisten soll (und kann) (Walkenhorst 2019). Multidimensional zu evaluieren, kann in eben diesem Bereich zu einer bedarfsgerechteren Evaluationspraxis beitragen (Milbradt 2018; Milbradt 2019), indem eine Bandbreite an Perspektiven einbezogen und das Arsenal an methodisch beantwortbaren Fragen erweitert werden.

Sollen multimethodische und multiperspektivische Designs in der Praxis umsetzbar sein, müssen bereits mit der Entwicklung von Maßnahmen Überlegungen zu ihrer Evaluation angestellt werden. Es gilt, spätestens bei der Auftragsvergabe für eine Evaluation zu klären, welche Zwecke und Erkenntnisziele verfolgt werden sollen und welche Ressourcen und Datenzugänge dazu zur Verfügung stehen (können). Erst dann kann geprüft werden, ob und in welcher Form der Einsatz mehrerer Methoden und die Abfrage mehrerer Perspektiven möglich und angemessen wären und inwieweit sie einen erkenntnisbezogenen Mehrwert bringen könnten, der den notwendigen Aufwand wert wäre. Soll eine multimethodische Evaluation möglich sein, muss sichergestellt sein, dass Evaluierende das notwendige Know-how dazu mitbringen. Zur Entwicklung der Fragestellungen und des letztendlichen Designs gilt es, alle Seiten des „Erwartungsdreiecks“ (Walkenhorst 2019) und insbesondere die evaluierte Fachpraxis einzubinden. Solche bedarfsgerechten Evaluationen setzen Flexibilität aufseiten der Evaluierenden und der Evaluierten voraus, dafür bedarf es auch entsprechender zeitlicher und finanzieller Rahmenbedingungen.

- Abu-Nimer, Mohammed 2020: Challenges in Peacebuilding Evaluation. Voices from the Field, in: D'Éstrée, Tamara P. (Hrsg.): *New Directions in Peacebuilding Evaluations*, Lanham, Maryland, 43–67.
- Chigas, Diana/Church, Madeline/Corlazzoli, Vanessa 2014: *Evaluating Impacts of Peacebuilding Interventions. Approaches and Methods, Challenges and Considerations (A Conflict, Crime, and Violence Results Initiative (CCVRI))*, London.
- Creswell, John W./Piano Clark, Vicki L. 2018: *Designing and Conducting Mixed Methods Research*, Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne.
- DeGEval 2016: *Standards für Evaluation*, Mainz.
- Döring, Nicola 2019: Evaluationsforschung, in: Baur, Nina/Blasius, Jörg (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden, 173–189.
- Döring, Nicola/Bortz, Jürgen 2016a: Evaluationsforschung, in: Döring, Nicola/Bortz, Jürgen (Hrsg.): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, Berlin, Heidelberg, 975–1036.
- Döring, Nicola/Bortz, Jürgen 2016b: Wissenschaftstheoretische Grundlage der empirischen Sozialforschung, in: Döring, Nicola/Bortz, Jürgen (Hrsg.): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, Berlin, Heidelberg, 31–79.
- Feddes, Allard R./Gallucci, Marcello 2015: A Literature Review on Methodology used in Evaluating Effects of Preventive and De-radicalisation Interventions, in: *Journal for Deradicalization* 5, 1–27.
- Flick, Uwe 2008: *Triangulation. Eine Einführung*, Wiesbaden.
- Flick, Uwe 2017: Triangulation in der qualitativen Forschung in der Psychologie, in: Mey, Günter/Mruck, Katja (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie*, Wiesbaden, 1–15.
- Galport, Matt/Galport, Nicole 2015: Methodological Trends in Research on Evaluation, in: *New Directions for Evaluation* 2015: 148, 17–29.
- Greene, Jennifer C. 2015: Preserving Distinctions Within the Multimethod and Mixed Methods Research Merger, in: Hesse-Biber, Sharlene N./Johnson, R. B. (Hrsg.): *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, Oxford, 606–615.
- Hauemberger, Sigrid/Baumgartner, Edgar 2017: Wirkungsevaluation in der Sozialen Arbeit mittels Realistic Evaluation: empirische Anwendungen und methodische Herausforderungen. Eine systematische Literaturreview, in: *Zeitschrift für Evaluation* 16: 1, 121–145.
- Hesse-Biber, Sharlene 2015: Introduction: Navigating a Turbulent Research Landscape. Working the Boundaries, Tensions, Diversity, and Contradictions of Multimethod and Mixed Methods Inquiry, in: Hesse-Biber, Sharlene N./Johnson, R. B. (Hrsg.): *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, Oxford, xxxiii–liii.
- Hunter, Albert/Brewer, John 2015: Designing Multimethod Research, in: Hesse-Biber, Sharlene N./Johnson, R. B. (Hrsg.): *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, Oxford, 185–205.
- Johnson, R. B./Onwuegbuzie, Anthony J./Turner, Lisa A. 2007: Toward a Definition of Mixed Methods Research, in: *Journal of Mixed Methods Research* 1: 2, 112–133.

- Jugl, Irina/Lösel, Friedrich/Bender, Doris/King, Sonja 2021: Psychosocial Prevention Programs against Radicalization and Extremism. A Meta-Analysis of Outcome Evaluations, in: *The European Journal of Psychology Applied to Legal Context* 13: 1, 37–46.
- Klawe, Willy 2006: Multiperspektivische Evaluationsforschung als Prozess. Wirkungsrekonstruktion aus der Sicht der Beteiligten, in: *Projekt eXe* (Hrsg.): *Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis*, München, 125–141.
- Koch, Günther/Fertsch-Röver, Jörg 2010: Multiperspektivität als methodische Antwort auf die Komplexität Sozialer Arbeit?, in: *Maykus, Stephan* (Hrsg.): *Praxisforschung in der Kinder- und Jugendhilfe. Theorie, Beispiele und Entwicklungsoptionen eines Forschungsfeldes*, 123–135.
- Koynova, Svetla 2021: Evaluationsbedarfe der Praxis. Ergebnisse einer Bedarfsabfrage im Präventionsfeld Islamistischer Extremismus, *PRIF Spotlight* 8/2021, Frankfurt am Main.
- Koynova, Svetla/Mönig, Alina/Quent, Matthias/Ohlenforst, Vivienne 2022: Monitoring, Evaluation und Lernen. Erfahrungen und Bedarfe der Fachpraxis in der Prävention von Rechtsextremismus und Islamismus, *PRIF Report* im Erscheinen, Frankfurt am Main.
- Lüders, Christian/Haubrich, Karin 2003: Qualitative Evaluationsforschung, in: *Schweppe, Cornelia* (Hrsg.): *Qualitative Forschung in der Sozialpädagogik*, Opladen, 305–330.
- Milbradt, Björn 2018: Ein Plädoyer für gegenstandsangemessene Evaluationsforschung (Blogreihe „Gesellschaft Extrem“), Frankfurt am Main.
- Milbradt, Björn 2019: (Neue) Evaluationskultur in der Radikalisierungsprävention? Forschungsmethoden, Akteurskonstellationen und Logik(en) der Praxis (Serie: Evaluation), Bonn.
- Möller, Kurt 2019: Evaluation neu denken: Der Dritte Raum (Serie: Evaluation), Bonn.
- Möller, Kurt/Buschborn, Jan/Pfeiffer, Thomas 2020: Zur Evaluation von Praxisansätzen der Extremismusprävention, in: *Ben Slama, Brahim/Kemmesies, Uwe* (Hrsg.): *Handbuch Extremismusprävention. Gesamtgesellschaftlich. Phänomenübergreifend*, 389–424.
- OECD 2012: *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results* (DAC Guidelines and References Series).
- Palinkas, Lawrence A./Mendon, Sapna J./Hamilton, Alison B. 2019: Innovations in Mixed Methods Evaluations, in: *Annual Review of Public Health* 40, 423–442.
- Patton, Michael Q. 2011: *Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use*, New York.
- Pawson, Ray/Tilley, Nick 1997: *Realistic Evaluation*, London.
- Pistone, Isabella/Eriksson, Erik/Beckman, Ulrika/Mattson, Christer/Sager, Morten 2019: A Scoping Review of Interventions for Preventing and Countering Violent Extremism. Current Status and Implications for Future Research, in: *Journal for Deradicalization* 19, 1–84.
- Quent, Matthias 2021: Prävention von Rechtsextremismus. Erfahrungen und Herausforderungen für Evaluation und wissenschaftliche Begleitung, *PRIF Spotlight* 7/2021, Frankfurt am Main.
- Ruf, Maximilian/Walkenhorst, Dennis 2021: Evaluation im Kooperationskontext. Chancen zur Gestaltung der Zusammenarbeit von Sicherheitsbehörden und Zivilgesellschaft, *PRIF Spotlight* 3/2021, Frankfurt am Main.

Walkenhorst, Dennis 2019: Das „Erwartungsdreieck Evaluation“: Eine Praxisperspektive (Serie: Evaluation), Bonn.

Weine, Stevan 2015: Applying Multimethod and Mixed Methods to Prevention Research in Global Health, in: Hesse-Biber, Sharlene N./Johnson, R. B. (Hrsg.): The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry, Oxford, 447–465.

TABELLE A: SUCHWORTKOMBINATIONEN

BEGRIFFSBLOCK MULTIMETHODISCH	VERKNÜPFUNG	BEGRIFFSBLOCK EVALUATION
<ul style="list-style-type: none"> - multimethodisch OR - multimethod OR - „mixed method*“ OR - mixed-method* OR - Methodenkombination OR - Methodenmix OR - „method combination*“ OR - „combination of methods“ 	<p>UND</p>	<ul style="list-style-type: none"> - Evalu* (*ation, *ierung, *ating, *ieren, etc.)

TABELLE B: ANALYSEKRITERIEN ZUR BESCHREIBUNG DER MULTIMETHODIK

ANALYSEKRITERIUM	DEFINITION	KODIERUNG
Jahr der Publikation	Jahreszahl zwischen 2010 und 2020	<i>Numerisch:</i> vierstellige Jahreszahl
Anzahl der verwendeten Erhebungsmethoden	Anzahl der unterschiedlichen Instrumente, die zur Erhebung von Daten genutzt wurden. Dies können auch zwei unterschiedliche Fragebögen oder Interviewleitfäden sein oder zwei Fokusgruppen mit unterschiedlicher Zielgruppe.	<i>Numerisch/kategorial:</i> 2=2 Methoden, 3=3 Methoden, 4=4 Methoden, >4= mehr als vier Methoden
Methodendesign	Die Verwendung quantitativer und/oder qualitativer Datenerhebungsmethoden.	<i>Kategorial:</i> „quantitativ“ (Fragebögen mit geschlossenen Fragen, andere Messinstrumente zur Erhebung von Daten (z.B. Leistungstest in Bildung oder Diagnoseinstrumente in Gesundheit, quantitative Dokumentenanalysen); „qualitativ“ (teilnehmende Beobachtungen, qualitative Dokumentenanalysen, Interviews, Fokusgruppen, Fragebögen mit offenen Fragen); quantitativ-qualitativ: Studien, die sowohl quantitative als auch qualitative Erhebungsinstrumente verwendet haben.

ANALYSEKRITERIUM	DEFINITION	KODIERUNG
Erhebungsmethode Befragung	Es wurden quantitative und/oder qualitative Befragungsinstrumente zur Erhebung von Daten verwendet.	<i>Binär: ja-nein</i>
Erhebungsmethode Dokumentenanalyse	Es wurden quantitative und/oder qualitative Dokumentenanalysen zur Erhebung von Daten verwendet.	<i>Binär: ja-nein</i>
Erhebungsmethode teilnehmende Beobachtung	Es wurden teilnehmende Beobachtungen zur Erhebung von Daten durchgeführt.	<i>Binär: ja-nein</i>
Sonstige Erhebungsmethoden	Es wurden sonstige (oft fachspezifische) Datenerhebungsinstrumente angewendet (z.B. Leistungstests oder Diagnoseinstrumente).	<i>Binär: ja-nein</i>
Befragungsmethode Fragebogen/Survey	Es wurden gedruckte oder internetbasierte Fragebögen mit offenen und/oder geschlossenen Fragen zur Erhebung von Daten verwendet.	<i>Binär: ja-nein</i>
Befragungsmethode Interview	Es wurden standardisierte oder teilstandardisierte Interviews zur Erhebung von Daten durchgeführt.	<i>Binär: ja-nein</i>
Befragungsmethode Fokusgruppe	Es wurden Fokusgruppen zur Erhebung von Daten durchgeführt.	<i>Binär: ja-nein</i>
Triangulation-vorhanden	Bei der Datenerhebung wurden Perspektiven unterschiedlicher Akteure erfasst.	<i>Binär: ja-nein</i>
Triangulation-Intensität	Anzahl der unterschiedlichen Akteure, deren Perspektive erhoben wurde.	<i>Numerisch</i>
Wirkungsanalyse	Es wurde eine Analyse der Wirkung einer Maßnahme durchgeführt.	<i>Binär: ja-nein</i>
Prozessanalyse	Es wurden Arbeits- und Implementationsprozesse einer Maßnahme oder eines Programms analysiert.	<i>Binär: ja-nein</i>
Wirkungs- und Prozessanalyse	Es wurden sowohl eine Wirkungs- als auch eine Prozessanalyse durchgeführt.	<i>Binär: ja-nein</i>
Experimentelles Design	Die Datenerhebung fand im Rahmen eines experimentellen Designs statt. Ein experimentelles Design wurde als gegeben definiert, sobald das Design eine Versuchs- und eine Kontrollgruppe vorsah.	<i>Binär: ja-nein</i>

PRIF REPORT

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

Drews, Friederike/Mannitz, Sabine (2021): Probleme der Aufarbeitung kulturellen Genozids. Rechtliche Regelungslücken und politische Defizite am Beispiel Kanadas, PRIF Report 7/2021, Frankfurt/M.

Klößner, Mona/Koynova, Svetla/Liebich, Johanna/Neef, Lisa (2021): Erfahrungen aus der Evaluationsplanung eines Aussteigerprogramms. Voraussetzungen für Wirksamkeitserfassung in der tertiären Extremismusprävention, PRIF Report 6/2021, Frankfurt/M.



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

PRIF SPOTLIGHT

Die PRIF Spotlights diskutieren aktuelle politische und gesellschaftliche Themen.

Schörnig, Niklas (2022): AI for Arms Control. How Artificial Intelligence Can Foster Verification and Support Arms Control, PRIF Spotlight 1/2022, Frankfurt/M.

Peters, Dirk (2021): Eine militärisch autonome EU? Europäische Sicherheit und transatlantische Partnerschaft nach Afghanistan, PRIF Spotlight 17/2021, Frankfurt/M.

PRIF SPOTLIGHT

DIE RICHTIGE FRAU AM RICHTIGEN ORT

Die Wirkung des Frauenförderprogramms 2016 an der HAW Magdeburg

Bei der ersten Lesung des Buchs 'Die richtige Frau am richtigen Ort' von der Autorin Dr. Ingrid Schörnig, die im Rahmen des PRIF Spotlights am 14. März 2022 in der HAW Magdeburg stattfand, wurde die Wirkung des Frauenförderprogramms 2016 an der HAW Magdeburg diskutiert. Die Autorin hat die Ergebnisse ihrer Untersuchung über die Wirkung des Frauenförderprogramms 2016 an der HAW Magdeburg in dem Buch 'Die richtige Frau am richtigen Ort' veröffentlicht. Das Buch ist ein Beitrag zur Diskussion über die Wirkung von Frauenförderprogrammen in der Hochschullandschaft.



www.hsfk.de/PRIF-Spotlights

PRIF BLOG

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

PRIF BLOG

HEINRICH HEIMANN INSTITUT FÜR FRIEDENS- UND KONFLIKTFORSCHUNG

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

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

www.facebook.com/HSFK.PRIF

www.twitter.com/HSFK_PRIF

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

Gefördert durch:



Bundesministerium
des Innern, für Bau
und Heimat

aufgrund eines Beschlusses
des Deutschen Bundestages

PrEval

Evaluationsdesigns für Präventionsmaßnahmen –
Multimethodische Ansätze zur Wirkungsermittlung und
Qualitätssicherung in der Extremismusprävention sowie den
Schnittstellen zur Gewaltprävention und politischen Bildung

MONA KLÖCKNER // MELANIE VERHOVNIK-
HEINZE // RAPHAELA SCHLICHT-SCHMÄLZLE //
REIKO NAKAMURA // JULIUS STRUNK

**MULTIMETHODISCHE EVALUATIONSDESIGNS:
EINE ERHEBUNG DER SOZIALWISSENSCHAFT-
LICHEN PRAXIS**

Bei der Evaluation von sozialen Programmen, wie etwa im Bereich der Extremismusprävention, kann es förderlich sein, mehrere Forschungsmethoden (Multimethodik) und Akteursperspektiven (Multiperspektivität) einzubeziehen, um dem Evaluationsgegenstand gerecht zu werden. In der Umsetzung können multimethodische Evaluationen jedoch unterschiedlich aussehen. Der vorliegende Report beschreibt auf Basis einer umfangreichen Analyse publizierter Evaluationen, wie Multimethodik und Multiperspektivität in der Breite der sozialwissenschaftlichen Evaluationsforschung umgesetzt werden. Der Report vergleicht die Umsetzung von Multimethodik und Multiperspektivität über verschiedene Disziplinen wie Bildung, Gesundheit, Politik, Kommunikation und Justiz hinweg und bietet damit eine Orientierungshilfe für ein entsprechendes Vorgehen bei der Evaluation in der Extremismusprävention.

ISBN 978-3-946459-69-9