

Qualitätssicherung in der Gewaltprävention

EINE HANDREICHUNG ZUR PROJEKTPLANUNG UND EVALUATION

ULRICH WAGNER, MARCUS KOBER, SUSANNE GIEL, JOHANNES MASSER, MANFRED
ROLFES & EGON WACHTER

Qualitätssicherung in der Gewaltprävention – Eine Handreichung zur Projektplanung und Evaluation

Ulrich Wagner (Philipps-Universität Marburg), Marcus Kober (Deutsches Forum Kriminalprävention), Susanne Giel (Univation Institut für Evaluation), Johannes Maaser (Stadt Marburg), Manfred Rolfes (Universität Potsdam), Egon Wachter (Kriminologische Forschung Baden-Württemberg)

Zusammenfassung

Die Überprüfung der Wirkung von Programmen zur Gewaltprävention ist aus ethischen und ökonomischen Gründen geboten. Die Durchführung von Evaluationen wird außerdem zunehmend zum Kriterium für die Unterstützung und finanzielle Förderung von Präventionsmaßnahmen. Dabei ist nicht immer klar, was unter Evaluation verstanden wird. Hier wird ein Modell dargestellt, das Evaluation systematisch in die Projektplanung einbindet. Außerdem werden Wirkungsevaluationsdesigns diskutiert, die unterschiedlich aufwändig sind und in Abhängigkeit davon zu Aussagen unterschiedlicher Sicherheit führen. Auch Wirkungsevaluationen, die auf Grund nachvollziehbarer praktischer Restriktionen in ihrer Aussagekraft eingeschränkt sind, können einen wertvollen Beitrag zur Qualitätssicherung von Präventionsmaßnahmen liefern. Dazu müssen die jeweiligen Einschränkungen allerdings klar herausgearbeitet und selbstkritisch diskutiert werden. Manchmal ist es möglich, dass Projektanbieter*innen und -teams selbst Wirkungsevaluationen durchführen. Zur Unterstützung wäre dann eine niederschwellige beratende Begleitung aus der Wissenschaft hilfreich. Unsere Hoffnung ist, dass der nachfolgende Text hilft, Projektplanung und Evaluation besser nachzuvollziehen und selbst einzusetzen, externe Unterstützungsangebote einzuschätzen und gegenüber Geldgebern deutlicher vermitteln zu können, dass Qualitätssicherung und Evaluation eine hinreichende finanzielle Ausstattung benötigen.

Es besteht weitgehender Konsens, dass die Eskalation von Konflikten hin zu Gewalt unakzeptabel ist und vermieden werden soll. Dementsprechend gibt es vielfältige Bemühungen, Interventionen zu entwickeln, die darauf abzielen, Konflikteskalationen und Gewalt vorzubeugen. Personenbezogene Maßnahmen des Konfliktmanagements richten sich an Personen, beispielsweise in Form von Schultrainings zur Steigerung der Konfliktbearbeitungskompetenz aller Schülerinnen und Schüler oder als Antiaggressionstrainings für bereits auffällige Personen. Situationsbezogene Maßnahmen versuchen, Veränderungen in gewaltfördernden materiellen Lebensumständen in den Blick zu nehmen, wie fehlende Bildungs- und Beschäftigungsmöglichkeiten oder ein ungünstiges Wohnumfeld. Zu den situationsbezogenen Maßnahmen gehören auch Interventionen zur Beeinflussung von Tatgelegenheitsbedingungen, wie die Schaffung von Blickachsen im öffentlichen Raum.

Professionelle systematische Interventionen zur Gewaltprävention sollen Menschen beeinflussen und verändern – ein Vorgehen, das ethische Fragen aufwirft. Zumindest muss versucht werden auszuschließen, dass eine Maßnahme mehr Schaden als Nutzen anrichtet. Darüber hinaus setzen Präventionsmaßnahmen ein zum Teil

erhebliches Engagement von Durchführenden und Teilnehmenden voraus, und sie erzeugen – zumindest kurzfristig gesehen - Kosten. All das macht eine systematische Planung, Durchführung und Evaluation erforderlich. Dazu im Folgenden einige Empfehlungen.

Projektplanung und -durchführung

Auf der linken Seite von Abbildung 1 sind die wichtigsten Schritte für die Planung und Umsetzung von Konflikteskalations- und Gewaltpräventionsmaßnahmen zusammengefasst. Der Ablauf ähnelt den Beccaria-Standards des Landespräventionsrates Niedersachsen (vgl. Schindler et al., 2005). Allerdings ist mit Schritt 3 die Notwendigkeit einer Begründung des geplanten Vorgehens besonders hervorgehoben: Präventionsmaßnahmen sollen auf nachvollziehbaren Wirkungsannahmen basieren, also Annahmen darüber, warum das, was man als Intervention einsetzt, tatsächlich zu dem angezielten Effekt führen soll.

Abbildung 1: Qualitätssicherung in der Prävention

Planung und Umsetzung	Evaluation
1. Problemanalyse und positive Zieldefinition: Was läuft falsch, was wollen wir erreichen?	
2. Interventionsplanung: Was soll konkret getan werden, mit welcher Intensität, wann, mit welcher Zielgruppe? Gibt es anderswo schon Ähnliches und was können wir daraus lernen?	
3. Entwicklung einer Veränderungstheorie: Warum soll die Intervention hilfreich sein?	
4. Einbezug aller relevanten Akteure: Wer muss bei der Intervention mitmachen, und wie kann man die Akteure gewinnen? Wer bezahlt?	
5. Erfolgsindikatoren festlegen: Woran kann man erkennen, dass die Intervention(en) zur Zielerreichung der Maßnahme beiträgt? Wie kann man unbeabsichtigte Nebeneffekte messen?	
6. Durchführung der Maßnahme	<i>Prozessevaluation: Wird die Intervention sachgerecht umgesetzt?</i>
7. Wirkdatenauswertung • unzureichende Zielerreichung • erfolgreiche Zielerreichung: Bericht und Veröffentlichung des Berichts	<i>Wirkungsevaluation: Vorher-Nachher-Vergleich, ggf. mit Kontrollgruppe</i>

vgl. auch <https://veranstaltungen.muenchen.de/soz/veranstaltungen/netzwerk-cafe-kommunales-konfliktmanagement-im-oeffentlichen-raum/>

Was ist die Ursache des vorliegenden Problems? Wie kann meine Maßnahme zur Lösung, oder zumindest zur Reduktion der negativen Auswirkungen des Problems beitragen? In vielen wissenschaftlichen Studien ist die lerntheoretische Annahme gut belegt, dass Menschen aggressives Verhalten zeigen, wenn das für sie mit Anerkennung einhergeht und keine negativen Konsequenzen nach sich zieht (Bandura, 1983). Mit Bezug auf diese Theorie gibt es guten Grund anzunehmen, dass negative Konsequenzen, wie Platzverweise für aggressives Auftreten, präventiv auf aggressives Verhalten wirken.

Überlegungen über Wirkungszusammenhänge – im folgenden Veränderungstheorien genannt - erhöhen die Wahrscheinlichkeit des Erfolgs einer Maßnahme, weil mit ihrer Entwicklung die zugrundegelegten angenommenen Wirkungsmechanismen offen gelegt werden. Außerdem werden Brüche in zugrundegelegten Annahmen deutlich, die Aufmerksamkeit auf mögliche un intendierte Nebeneffekte gerichtet und Randbedingungen für die Wirksamkeit einer Maßnahme klar (vgl. auch Mayne, 2012).

Konfliktbearbeitungstrainings in Kindergärten und Schulen setzen oft vorausgehende Schulungen für das Lehrpersonal voraus. Eine Veränderungstheorie sollte plausibel erklären, warum, erstens, die eingesetzte Schulungsmaßnahme die Kompetenz des Personals zur Durchführung der Maßnahme sicherstellt und, zweitens, das (von entsprechend geschultem Personal) eingesetzte Training die Konfliktbearbeitungskompetenz der teilnehmenden Kinder erhöht.

In Einzelfällen mag man als Begründung für die Durchführung einer Maßnahme auf praktische Erfahrungen zurückgreifen, wenn Projektplaner*innen bestimmte Interventionsformen bereits wiederholt durchgeführt haben und selbstkritisch und glaubwürdig vom Erfolg dieser Interventionen berichten können. Der Rückgriff auf gut bewährte wissenschaftliche Theorien als Begründung für die Entwicklung einer Maßnahme ist allerdings besonders sinnvoll, weil solche Theorien in der Regel durch vielfältige und unabhängige wissenschaftliche Studien und Meta-Analysen in ihrer Wirkung überprüft sind, womit sich die Wahrscheinlichkeit der Wirksamkeit der Maßnahme erhöht. In jedem Fall soll die wissenschaftliche oder erfahrungsbasierte Begründung für die Ausgestaltung einer Maßnahme explizit gemacht, d. h. die Veränderungstheorie schriftlich festgehalten werden.

Die Veränderungstheorie ist nicht nur für die Planung der Intervention zentral, sondern auch für die Festlegung einer angemessenen Evaluation. Sie hilft zu entscheiden, welche Wirkungsannahmen überprüft werden sollen und wo u.U. auf aufwändige Datenerhebungen verzichtet werden kann. Bei begrenzten Ressourcen können so Evaluationen auf jene Aspekte konzentriert werden, die für die Weiterentwicklung der Maßnahme oder für zentrale Entscheidungsfragen tatsächlich relevant sind.

Der auf der linken Seite von Abbildung 1 dargestellte Ablauf zielt auf die Planung und Umsetzung einer neuen Intervention. Dabei kann es sich um eine Präventionsmaßnahme handeln, die auf spezifische Bedürfnisse einzelner Personen abzielt, wie bei einer individuellen Beratung zum Umgang mit eskalierenden Konflikten im familiären Umfeld. Das Ablaufschema passt gleichermaßen für die Planung und Durchführung von Konfliktbearbeitungstrainings in Schulen oder situationsbezogene Präventionsmaßnahmen, wie eine verstärkte Beleuchtung eines öffentlichen Platzes. Das Ablaufschema kann auch angewandt werden für komplexe Maßnahmenpakete, wie die Neueinführung einer Präventionsstruktur für eine Kommune (Weißmann et al., 2025).

Insbesondere bei solchen komplexen Interventionen, die Kombinationen von verschiedenen Maßnahmen zu u. U. unterschiedlichen Zeitpunkten vorsehen, ist die Entwicklung und Ausformulierung einer angemessenen Veränderungstheorie erforderlich, die beschreibt, wie die einzelnen Komponenten der Intervention aufeinander aufbauen sollen und was die zu erwartenden unmittelbaren und langfristigen Konsequenzen sind (vgl. z. B. Röding et al., 2021, die ein Modell für Kommunen vorstellen, die am Communities that Care Programm teilnehmen).

Planung und Umsetzung von Präventionsmaßnahmen, wie sie auf der linken Seite von Abbildung 1 dargestellt sind, und ihre Evaluation hängen eng zusammen. Zu den Evaluationsmaßnahmen gehören Prozess- und Wirkungsevaluation.

Evaluation

Evaluationen sind mit Aufwand verbunden. Sie erfordern Zeit, fachliche Kompetenzen sowie finanzielle und personelle Ressourcen, die im Praxisalltag von Präventionsakteuren oft nur begrenzt verfügbar sind. Daher sollte bereits frühzeitig geklärt werden, welche Frage eine Evaluation beantworten soll – und welche nicht. Evaluationen können unterschiedliche Funktionen erfüllen: Neben der Lern- und Reflexionsfunktion, die auf die Weiterentwicklung von Maßnahmen zielt, übernehmen sie häufig auch Steuerungs- und Legitimationsfunktionen gegenüber Fördermittelgebenden oder politischen Entscheidungsträgern. Darüber hinaus können Evaluationen eine symbolische Funktion haben, indem sie nach außen Reflexivität, Professionalität oder Verantwortungsbewusstsein signalisieren. Diese unterschiedlichen Funktionen stellen unterschiedliche Anforderungen an Design und Umfang der Evaluation.

Prozessevaluation versucht, die Qualität der Umsetzung einer Maßnahme zu erfassen: Wichtig sind dabei Fragen wie „Sind die in der Interventionsplanung (vgl. Abbildung 1, linke Seite Nr. 2) und der Begründung (3) beschriebenen Details in der Intervention angemessen umgesetzt“ und (4) „konnten alle für die Intervention wichtigen Akteure angemessen einbezogen werden“? (zu weiteren Aspekten vgl. DFK, 2013). Dazu gehört auch die Antwort auf die Frage, ob die beantragten Mittel sachgerecht eingesetzt wurden. Negative Antworten auf diese Fragen sollten zum Nachsteuern und zur Optimierung des Projektverlaufs veranlassen. Werden auf Basis der Prozessevaluation Veränderungen vorgenommen, kann auch von formativer Evaluation gesprochen werden.

Wirkungsevaluation soll überprüfen, ob eine Maßnahme tatsächlich dazu beiträgt, das angestrebte Ziel (vgl. 1) zu erreichen. Welche Form der Wirkungsevaluation angemessen ist, hängt von der Funktion der Evaluation, vom Entwicklungsstand der Maßnahme sowie von den verfügbaren Ressourcen ab. Nicht jede Intervention erfordert eine umfassende kausale Wirkungsprüfung; oft steht die Frage im Vordergrund, ob zentrale Wirkungsannahmen plausibel sind und für die Weiterentwicklung nutzbar gemacht werden können.

Wirkungsevaluation setzt voraus, Indikatoren zu benennen, die die Zielerreichung messbar machen (5).

Wenn man beispielsweise ein auf Schüler*innen abzielendes Konfliktbearbeitungstraining evaluieren will (vgl. z.B. Lemmer & Wagner, 2013), können zu den Indikatoren Kompetenztests gehören, die man den Schüler*innen vorlegt, oder auch die Entwicklung der Unfallstatistik. Richten sich Interventionen eher auf Veränderungen in spezifischen Raumeinheiten, wie dies beispielsweise bei Maßnahmen des kommunalen Konfliktmanagements der Fall ist (vgl. z.B. Meyer & Lephthin, 2025), können unabhängige Beobachtungen und kleinräumige Analysen von relevanten Indikatoren der Polizeilichen Kriminalitätsstatistik (PKS) angemessene Erfolgsindikatoren sein.

Antworten von Teilnehmer*innen oder von Durchführenden von Präventionsprogrammen auf Fragen wie „Hat das Programm geholfen?“ können im Rahmen von Selbstevaluationen hilfreiches Feedback einholen und Hinweise auf Verbesserungspotentiale liefern; sie sind jedoch in der Regel keine zuverlässigen Indikatoren für Wirkungsevaluationen, weil sie von Urteilsverzerrungen verfälscht sein können: Psychologische Forschung zeigt, dass man dazu neigt, das eigene Verhalten zu rechtfertigen - wenn ich schon an einer solchen Maßnahme teilgenommen habe und Zeit investiert habe, unterstelle ich, dass die Maßnahme auch nutzt.

Abbildung 2 stellt auf der linken Seite die wichtigsten Designs für die Durchführung einer Wirkungsevaluation zusammen. Dabei symbolisiert „i“ die jeweilige Intervention und „m“ die Messung.

Wirkungsevaluation zielt darauf ab festzustellen, ob es eine Veränderung im angestrebten Verhalten gibt, wie weniger Konflikteskalationen oder Gewalt, und ob diese Veränderung kausal auf die Präventionsmaßnahme zurückgeführt werden kann. Es geht also um die Beschreibung und den Nachweis von Veränderungen. Daher sind sogenannte Einmalmessungen (siehe erste Zeile in Abbildung 2), bei denen man die Erfolgsindikatoren nur nach Programmdurchführung erfasst, in der Regel nicht zielführend, eben weil ihnen Vergleichswerte fehlen.

Evaluator*innen sind nicht selten mit einer Situation konfrontiert, dass die Bewilligung eines Präventionsprojektes erst unmittelbar vor Projektstart erfolgt, das Bewilligungsschreiben aber gleichzeitig eine Evaluation einfordert. Dann bleibt oft nur die Erfassung von m_1 , eine Messung, die „nur“ noch zur Prozessevaluation herangezogen werden kann. Der Zugriff auf zwei miteinander vergleichbaren Messungen gehört zu den Mindestanforderungen von Wirkungsevaluationen.

Das einfachste inhaltlich interpretierbare Untersuchungsdesign in der Wirkungsevaluation ist ein Vortest-Nachtest-Vergleich der Messungen in der Interventionsgruppe, also ein Vergleich zwischen m_0 und m_1 . Hat sich die Kompetenz von Trainingsteilnehmer*innen zur Vermeidung eskalierender Konflikte von m_0 zu m_1 verbessert oder zeigt die lokale PKS von m_0 zu m_1 einen Rückgang an Straßenkriminalität, spricht einiges – aber nicht alles - dafür, dass dies auf den Einsatz der Interventionsmaßnahme i zurückgeht.

Abbildung 2: Möglichkeiten der Wirkungsevaluation: Designs

						Bewertung
Einmalmessung						Zur Wirkevaluation unbrauchbar.
	Interventions- gruppe		i	m ₁		
Vortest-Nachtest Design						Brauchbar, wenn andere Einflüsse (außer "i") unwahrscheinlich sind.
	Interventions- gruppe	m ₀	i	m ₁	(m ₂)	
Vortest-Nachtest Design mit Kontrollgruppe						Aussagekräftig, wenn Interventions- und Kontrollgruppe vergleichbar sind.
Evtl. Zufalls- zuweisung	Interventions- gruppe	m ₀	i	m ₁	(m ₂)	
	Kontroll-/ Vergleichsgruppe	m ₀		m ₁	(m ₂)	
i = Intervention / Maßnahme m = Messung (Wissen, Einstellung, Verhaltensabsicht, Verhalten, ...)						

Allerdings ergeben sich bei einem solchen Evaluationsdesign auch Einschränkungen: Zum Beispiel können sich die beteiligten Personen zeitgleich, aber unabhängig von der Maßnahme, also zwischen m₀ und m₁, verändert haben – z. B. reifer und friedlicher geworden sein, was insbesondere zu beachten ist, wenn zwischen m₀ und m₁ ein langes Zeitintervall liegt. Oder es ist etwas Wichtiges geschehen, das eigentlich die Veränderung in m₁ ausgelöst hat, wie z. B. Corona-Ausgangsbeschränkungen oder strenge Frosttage, die die Möglichkeiten für aggressive Auseinandersetzungen im öffentlichen Raum eingeschränkt haben. Um solche Reifungsprozesse und externen Einflüsse kontrollieren zu können, suchen Evaluator*innen nach Vergleichsmöglichkeiten, oft in Form einer möglichst ähnlichen Vergleichsgruppe, die keine Intervention durchläuft.

In Vortest-Nachtest Designs mit Kontrollgruppe werden m₀ und m₁ in der Interventions- und in der Vergleichsgruppe gemessen. Verlaufen die Veränderungen in Interventions- und Vergleichsgruppe unterschiedlich und in der Interventionsgruppe stellt sich ein Effekt ein, der über die Veränderung in der Kontrollgruppe hinausgeht, deutet das auf die Wirkung der Intervention hin – vorausgesetzt, Interventions- und Kontrollgruppe sind vergleichbar.

Eine Vergleichbarkeit von Interventions- und Kontrollgruppe wird idealerweise erreicht, wenn Untersuchungsteilnehmer*innen oder Untersuchungsregionen nach Zufall der Interventions- oder der Kontrollgruppe zugewiesen werden, ein sogenanntes

Pretest-Posttest-Kontrollgruppendesign mit Zufallszuweisung. Das ist in der Praxis oft kaum umsetzbar. Wenn eine solche Zufallszuweisung dennoch gelingt und Interventions- und Kontrollgruppe über eine hinreichend große Zahl von Personen oder Regionen verfügen, kann ein nachweisbarer Unterschied zwischen Interventions- und Kontrollgruppe in der Entwicklung von m_0 zu m_1 als überzeugender Beleg für die Wirksamkeit des Programms angesehen werden. Das Vortest-Nachtest Kontrollgruppen-Untersuchungsdesign mit Zufallszuweisung stellt das theoretische Ideal der Wirkungsevaluation dar (vgl. auch Shadish, Cook, & Campbell, 2002).

In der Abbildung 2 ist bei den verschiedenen Designs „ m_2 “ in Klammern angefügt. Das soll darauf hinweisen, dass häufig ein Interesse besteht, auch die langfristige Wirkung einer Maßnahme zu überprüfen. Das macht neben m_1 , die oft unmittelbar nach der Intervention erhoben wird, weitere verzögerte Messzeitpunkte m_2 , m_3 usw. wünschenswert – was neue Probleme aufwerfen kann: Wenn die Evaluation Befragungen von Projektteilnehmer*innen vorsieht, sind die mit zunehmendem Abstand von der Intervention oft nur noch schwer erreichbar. Ein solcher „Schwund“ lässt sich manchmal ein wenig reduzieren, wenn man Befragungen online durchführt und Beantwortungen über Handy anbietet. Kommt es dennoch zum Ausfall von Untersuchungsteilnehmer*innen, muss man die Auswertung mit den Angaben der Personen vornehmen, die zu den verschiedenen Messzeitpunkten erreichbar waren.¹

Oft streben Präventionsprogramme die Erreichung mehrerer Ziele an, für die Veränderungstheorien und Indikatoren entwickelt werden müssen. Und die verschiedenen Ziele sind in einer umfangreichen Ergebnisevaluation entsprechend zu berücksichtigen. Das „ m “ in Abbildung 1 symbolisierte dann mehrere Erfolgsindikatoren.

Beispielsweise zielt das Programm „Prävention im Team“ auf „die Entwicklung persönlicher Handlungsoptionen im Sinne der Fähigkeit der Vermeidung eines Konflikteinstieges und der Herbeiführung eines Konfliktausstieges sowie die Stärkung von Zivilcourage“ (unsere Hervorhebungen). Eine Evaluation sollte diese verschiedenen Kriterien einbeziehen (vgl. Lemmer & Wagner, 2013).

Was genau mit den verschiedenen Designs gemessen werden soll und wie die Messungen durchgeführt werden sollten, hängt von der Zielbeschreibung des Projekts ab.

Zielt eine Intervention darauf ab, Gewaltkriminalität zu reduzieren, wären PKS-Daten oder Unfallstatistiken angemessene Indikatoren. Geht es in einem Programm primär darum, das Sicherheitsgefühl der Bevölkerung zu verbessern, wären Daten aus Bevölkerungsbefragungen zu Kriminalitätsfurcht geeignete Messgrößen.

Schließlich sollte die Gesamtevaluation eines Projektes idealerweise auch die Erfassung unbeabsichtigter Nebeneffekte einbeziehen und diese in den Messungen abbilden. Wenn beispielsweise die Wirkung von verstärkter Beleuchtung öffentlicher Räume auf die Kriminalität überprüft werden soll, sind in einer umfangreichen Evaluation auch die Verdrängungseffekte zu berücksichtigen, die sich in

¹ Das kann allerdings die Sicherheit der Interpretation der Ergebnisse einschränken, wenn beispielsweise vor allem solche Personen ausfallen, die die Intervention als besonders unattraktiv erlebt haben. Komplexe statistische Auswertungen würden versuchen, solche Ausfälle zu kompensieren, indem für die fehlenden Messwerte zu m_1 oder m_2 der Ausgangswert von m_0 eingesetzt wird oder auf der Basis des gesamten Ergebnismusters Imputationen vorgenommen werden.

erhöhter Kriminalität in möglichen Ausweichräumen äußern kann. Das erfordert allerdings zusätzlichen Aufwand. Daher ist abzuwägen, welche möglichen Nebenfolgen besonders relevant sind und mit vertretbarem Aufwand erfasst werden können.

Was tun mit Evaluationsergebnissen?

Abbildung 1 veranschaulicht den Zusammenhang von Projektplanung und -umsetzung mit Evaluation. Die Pfeile in der Mitte der Abbildung sollen deutlich machen, was zu tun ist, wenn eine Wirkungsevaluation nicht zum gewünschten Ergebnis führt, d.h. die Ziele nicht erreicht werden oder zu viele negative Nebeneffekte auftauchen. Dann ist die Frage, auf welcher Stufe oder welchen Stufen des Prozesses etwas schiefgelaufen ist:

Bei der Interventionsplanung (2: Passt beispielsweise die Intervention nicht zur Zielgruppe?),
ist die zugrunde gelegte Begründung falsch (3: Spielen beispielsweise bislang unbeachtete Gruppenprozesse bei der Gewalteskalation eine Rolle?),
sind wichtige, für die Wirksamkeit der Intervention notwendige unterstützende Akteure bislang zu wenig einbezogen (4: die Eltern?),
sind die Effekte der Intervention richtig gemessen worden (5) und
wurde die Maßnahme wie geplant durchgeführt (6, was Gegenstand der Prozessevaluation ist).

Dieser Reflexionsprozess soll dazu beitragen, die Intervention, ihre Planung und Evaluation zu verbessern, statt weiter wie bislang zu verfahren und dabei Arbeitskraft und Geld für eine unwirksame und unter Umständen sogar schädliche Maßnahme zu verschwenden. Im Grunde bedeutet das, den Prozess von Planung, Umsetzung und Evaluation an der diagnostizierten Fehlerstelle erneut aufzugreifen und zu verbessern.

Zeigt die Evaluation, dass die Ziele der Maßnahme erreicht wurden und auch keine beachtenswerten negativen Nebeneffekte auftreten, sollten eine Beschreibung der Maßnahme, ihre Veränderungstheorie, das Einsatzfeld mit den beteiligten anonymisierten Trainingsteilnehmer*innen sowie die Evaluation und ihre Ergebnisse zusammengeschrieben (z.B. nach einer Gliederung wie in Abbildung 1) und der Öffentlichkeit zugänglich gemacht werden. Das ermöglicht, dass andere Interessente eine so qualitätsgesicherte Maßnahme aufgreifen können.

Zu bedenken ist allerdings, dass selbst sorgfältig geplante Wirkungsevaluationen nicht automatisch zu Lernen oder Veränderungen in der Praxis führen. Ob und wie Evaluationsergebnisse genutzt werden, hängt von spezifischen organisationalen Routinen, Entscheidungsstrukturen und vorhandenen Deutungsmustern ab. Ob in Organisationen aus Evaluationen gelernt wird, lässt sich nicht allein durch methodische Qualität sicherstellen.

Wirkungsevaluation in der Praxis – Annäherungen an das (theoretische) Ideal des Kontrollgruppendesigns mit Zufallszuweisung

Die Umsetzung eines Pretest-Posttest-Kontrollgruppendesigns mit Zufallszuweisung ist in der Praxis oft nur schwer oder überhaupt nicht möglich, weil zum Beispiel die

Zufallszuweisung nicht gelingt, oder weil etwa die Schulleitung entscheidet, welche Schulklasse am Training teilnehmen soll und welche nicht. Dann kann es hilfreich sein, auf eine möglichst ähnliche Vergleichsgruppe, wie eine Parallelklasse zur Interventionsklasse, zurückzugreifen. Je höher die Ähnlichkeit der Vergleichsgruppe, umso eher kann man annehmen, dass Reifungsprozesse und externe Einflüsse auch in der Kontrollgruppe auftreten und damit kontrolliert werden. Auch gibt es manchmal verfügbare Vergleichsdaten, die die Kontrollfunktion einer Vergleichsgruppe übernehmen können. Ein Beispiel wäre der Vergleich der Entwicklung der PKS-Daten zur Straßenkriminalität in einer Kommune, die ein kommunales Konfliktmanagement im öffentlichen Raum neu zum Einsatz gebracht hat, mit der Entwicklung der PKS-Straßenkriminalität im Land oder in einer möglichst ähnlichen Vergleichskommune. Vergleiche mit solchen nicht-randomisierten Kontrollgruppen können nicht alle Bedenken ausschließen: Wenn die Interventionsklasse nach Auffassung der Schulleitung deshalb am Training teilnehmen sollte, weil sie vorher als besonders gewalttätig eingeschätzt wurde, sind statistisch „Zufallsberuhigungen“, sogenannte Regressionen zur Mitte², zu erwarten. Oder in der Interventionskommune mag es weitere, unerkannte Einflussfaktoren gegeben haben, beispielsweise wenn die auffälligsten „Troublemakers“ verurteilt und inhaftiert wurden, was die eigentliche Ursache dafür ist, dass sich der öffentliche Raum in der ganzen Kommune beruhigt.

Erweist sich die Gewinnung einer Kontrollgruppe als nicht umsetzbar und sind auch andere Vergleichsdaten nicht verfügbar, bleibt nur die Umsetzung eines Vortest- und Nachtest-Vergleichs in der Interventionsgruppe. Ein solches Vorgehen ist immer noch von großem Wert und deutlich besser als der Verzicht auf jegliche Wirkungsevaluation. Man muss sich allerdings darüber im Klaren sein, dass damit Risiken bei der Interpretation der Evaluationsergebnisse einhergehen.³

Im Zusammenhang mit der Evaluation von Präventionsmaßnahmen zur Vermeidung von Konflikteskalation und Gewalt gibt es in der Regel keine ein für alle Mal eindeutigen Antworten, die über alle Zweifel erhaben sind. Bei allen Evaluationen und Evaluationsergebnissen ist die intensive und selbstkritische Suche nach möglichen Alternativerklärungen und eine Einschätzung der Wahrscheinlichkeit, mit der diese Alternativerklärungen die vorliegenden Evaluationsergebnisse verzerrt haben könnten, unabdingbar. Das gilt gleichermaßen für positive wie für negative Evaluationsergebnisse. So kann trotz sorgfältiger Umgestaltungen eines Angstortes keine Verbesserung der wahrgenommenen Qualität des Ortes feststellbar sein, weil gleichzeitig mit Umsetzung der Umgestaltungsmaßnahme die allgemeine Krisenwahrnehmung in der Bevölkerung zugenommen hat.

² Viele in der Evaluation verwendeten Messungen sind ungenau, wenn beispielsweise Befragte Fragebögen unkonzentriert beantworten. Manchmal führt diese Ungenauigkeit dazu, dass in einer Interventions- oder Kontrollgruppe in m_0 besonders extreme Werte auftauchen. Weil diese Extreme eher zufällig sind, ist zu erwarten, dass sie bei der nächsten Befragung weniger extrem ausfallen. Aufgrund einer solchen Regression zur Mitte können Interventionseffekte falsch eingeschätzt werden.

³ Hierzu gehört, neben den beschriebenen Reifungsprozessen und externen Einflüssen, die mangelnde Kontrollmöglichkeit der Wirkung von Messwiederholungseffekten: Wenn man Menschen in m_0 und m_1 wiederholt dasselbe fragt, kann allein die Wiederholung der Frage die Antworten verändern.

Wirkungsevaluation als Selbstevaluation

Die praktische Umsetzung von Evaluationen ist ein anspruchsvoller Prozess, der hohe Sachkompetenz, Arbeitszeit, finanzielle Mittel und ein zugängliches Praxisfeld voraussetzt. Häufig besteht keine Möglichkeit, externe Evaluationsfachleute zur Evaluation hinzuzuziehen, selbst bei einem so komplexen Prozess wie Wirkungsevaluation. Alternativ ist allerdings manchmal eine Wirkungsevaluation in Form einer Selbstevaluation umsetzbar, die dann eher den Zweck verfolgt, Verbesserungspotentiale zu identifizieren, Reflexionsprozesse zu unterstützen oder zur Qualitätssicherung beizutragen. Mit Blick auf Wirkungsevaluationen ist bei Selbstevaluationen allerdings oft mit Abstrichen in Bezug auf die beschriebenen Optimalstandards zu rechnen. Zudem stellt bereits eine einfache Selbstevaluation für Präventionsakteure eine zusätzliche Belastung dar. Umso wichtiger ist es sicherzustellen, dass Umfang und Tiefe der Evaluation in einem angemessenen Verhältnis zu den verfügbaren Ressourcen und zum erwartbaren Erkenntnisgewinn stehen.

Eine oft beobachtbare Einschränkung bei der Durchführung von Selbstevaluation betrifft die Umsetzung der Wirkungsmessung. Die Messung von Interventionswirkungen kann hoch-anspruchsvoll sein, wenn die Indikatoren beispielsweise aus Befragungen der Teilnehmer*innen einer Maßnahme gewonnen werden sollen: In der Psychologie z.B. betreiben das oft spezialisierte Diagnostiker*innen, die in aufwändigen Voruntersuchungen lange Skalen zur Erfassung von Gewaltneigung entwickeln. In der Praxis macht es in der Regel Sinn, auf solche etablierten Skalen zurückzugreifen, auch wenn es manchmal notwendig ist, nur einzelne Teile solcher Skalen einzusetzen, weil nicht genug Zeit für eine ausführliche Befragung zur Verfügung steht. Solche Einschränkungen in Kauf zu nehmen, ist besser, als auf eine Wirkungsevaluation ganz zu verzichten. Teils gibt es auch gut etablierte Kurzindikatoren zur Messung bestimmter Problemausprägungen. So hat sich in der internationalen Sicherheitsforschung die Frage „Wie sicher fühlen Sie sich, wenn Sie nachts allein in Ihrer Wohngegend unterwegs sind?“ als brauchbare Messmethode zur Erfassung von Kriminalitätsfurcht etabliert (vgl. Reuband, 2008).

Wenn eine Wirkungsevaluation nicht nur aus dem Vergleich zweier Messzeitpunkte eines Indikators besteht, wie einer Veränderung der PKS-Daten in einer Interventionskommune (möglicherweise noch ergänzt um die Daten aus einer Vergleichskommune), haben Selbstevaluationen oft mit Problemen bei der statistischen Auswertung ihrer Daten zu kämpfen. Die in Abbildung 2 eingeführten Buchstaben „m“ stehen für Messungen, die häufig aus Befragungen mehrerer Personen bestehen. Diese Einzelantworten dieser Personen müssen zu Mittelwerten zusammengezogen werden. Aus dem Vergleich dieser Mittelwerte vor (m_0) und nach der Intervention (m_1) kann man dann oft schon etwas über die Wirkung einer Intervention ablesen. Wirklich aussagekräftig werden solche Mittelwertvergleiche allerdings erst dann, wenn die Unterschiede auf Signifikanz abgetestet werden, eine statistische Konvention, die hilft, zu entscheiden, ob Unterschiede als Zufallseffekte oder als bedeutsam anzusehen sind. An dieser Stelle kommen viele Selbstevaluator*innen an ihre Grenzen und brauchen externe Unterstützung. Dennoch, Wirkungsevaluation in Form einer Selbstevaluation ist möglich und sollte zur Qualitätssicherung des eigenen Vorgehens eingesetzt werden.

Wissenschafts-Praxis Kooperation

Wirkungsevaluationen durch Programmentwickler*innen und -durchführenden kann in erster Linie der Qualitätssicherung des eigenen Vorgehens dienen: Hat die Intervention, so wie sie umgesetzt wurde, wirklich dazu beigetragen, die beschriebenen Ziele zu erreichen oder muss die Intervention weiter optimiert werden? Fragen der Übertragbarkeit einer Intervention auf andere Kontexte und andere Zielgruppen fallen dann eher in das Feld wissenschaftlich basierter Evaluationsforschung.

In einer idealen Präventions- und Evaluationswelt würden die Anbieter*innen von Interventionsprogrammen Wirkungsevaluationen zur Optimierung ihres Vorgehens selbst durchführen und an Stellen, an denen sie Nachfragen z.B. zur Messung oder zur Auswertung haben, die Beratung von Wissenschaftler*innen suchen. Gleichzeitig würde man sich wünschen, dass aus der Wissenschaft heraus in Kooperation mit Praktiker*innen groß angelegt hoch-sophistische Evaluationen über lange Zeiträume durchgeführt werden und deren Ergebnisse sowie die Ergebnisse von Selbstevaluationen in Metaanalysen noch einmal zusammengefasst werden. Das alles bedeutet den Einsatz von Arbeitskraft und eine entsprechende Finanzierung. Beides ist gegenwärtig weder von Projektfinanzierern noch von der Wissenschaftsförderung in hinreichendem Maße eingeplant. Das muss sich ändern.

Ganz schnell braucht es allerdings ein Beratungsnetzwerk, auf das Praktiker*innen bei der Selbstevaluation mit konkreten Einzelfragen oder bei Vermittlungen von Kontakten zu einschlägigen Expert*innen und Institutionen zurückgreifen können.

Die Forderung nach Evaluation ist berechtigt, muss aber die praktischen Möglichkeiten im Auge behalten

Die Vorlage von positiven Evaluationsergebnissen wird zunehmend zum Kriterium für die Unterstützung und die finanzielle Förderung von Präventionsmaßnahmen. Dabei hat man nicht immer den Eindruck, dass diejenigen, die einen Evaluationsnachweis erwarten, klare Vorstellungen davon haben, welche Funktion die Evaluation haben soll, welche Evaluationsform durchgeführt, was genau als Evaluationsergebnis vorgelegt werden soll und ob die gestellten Ansprüche nicht an den praktischen Möglichkeiten vorbeigehen.

Professionelle Intervention, so auch Maßnahmen zur Prävention von Konflikteskalationen und Gewalt, sollte einer Systematik folgen, wie sie in Abbildung 1 auf der linken Seite beispielhaft beschrieben ist, und eine Prozessevaluation vorlegen können. Die Sicherung dieser Qualitätsstandards allein erhöht die Erfolgswahrscheinlichkeit einer Intervention und ist in der Regel für Projektdurchführende auch leistbar. Begleitende Selbstevaluationen zur Wirksamkeit sind zusätzlich zu empfehlen, um Maßnahmen oder Projekte in der Praxis strukturiert verbessern zu können. Für neu entwickelte Präventions- und Interventionsmaßnahmen ist, nach einer Erprobungs- und Optimierungsphase, Wirkungsevaluation zwingend einzufordern, wenn es um die Fortsetzung, Weiterverbreitung und finanzielle Förderung geht. Welches Evaluationsdesign dabei zugrunde gelegt werden soll (Pretest-Posttest Design mit oder ohne Kontrollgruppe mit oder ohne Zufallszuweisung oder auch ein anderes, hier nicht diskutiertes Design, vgl. Shadish, Cook, & Campbell, 2002) ist im Einzelfall und u.U. nach Expert*innenberatung zu entscheiden.

Auch für den Einsatz von bereits etablierten und in ihrer Wirksamkeit nachgewiesenen effektiven Programmen (vgl. dazu z.B. die Grüne Liste Prävention, Communities that care, o.D.) ist eine angemessene Prozessevaluation notwendig. Wenn solche etablierten Interventionen in der Problem- und Zielbeschreibung, der Intervention und der Zielgruppe vom Original abweichen, ist darüber hinaus eine erneute Wirkungsevaluation sinnvoll. Deren Ergebnisse würden auch dabei helfen, das mögliche Einsatzfeld und die Breite der Wirksamkeit der Maßnahme besser einschätzen zu können.

Angesichts der in vielen Professionen und Handlungsfeldern lauter werdenden Forderung nach Evaluation von Maßnahmen als Voraussetzung für deren (Weiter)Förderung oder Verstetigung ist zwischen Entscheidungsträgern und Projektverantwortlichen zu klären, welche der genannten Evaluationsformen für sinnvoll und notwendig erachtet wird. Dabei gilt es Kompromisse zu finden zwischen Ansprüchen an die Aussagekraft von Evaluationsergebnissen und der praktischen Möglichkeit, auch komplexe Wirkungsevaluationsdesigns umzusetzen, die u.U. von externen Fachleuten unterstützt und durchgeführt werden müssten. In jedem Fall sollten zeitliche sowie finanzielle Ressourcen für die Durchführung von Projektevaluierungen eingeplant und von geldgebenden Institutionen auch bereitgestellt werden. Nur mit einer realistischen Ausstattung, mit Know-How und notwendigen Mitteln können aussagekräftige Evaluationen durchgeführt werden, die die Arbeit von Praktiker*innen und die Planung neuer Maßnahmen spürbar verbessern.

Literatur

Bandura, A. (1983). Psychological mechanisms of aggression. In R.G. Green & E.I. Donnerstein (eds.), *Aggression. Theoretical and empirical reviews* (pp. 1-40). Academic Press.

Communities that care (o.D.) Grüne Liste Prävention. <https://www.gruene-liste-praevention.de/nano.cms/datenbank/information>

Lemmer, G. & Wagner, U. (2013). PiT-Hessen. Ein standardisiertes Gewaltpräventionsprogramm zum Einsatz in Schulen und seine Evaluation. *Forum Kriminalprävention*. <https://www.forum-kriminalpraevention.de/files/1Forum-kriminalpraevention-webseite/pdf/2013-02/evaluation-2013-02.pdf>

Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18, 270-280. DOI: [10.1177/1356389012451663](https://doi.org/10.1177/1356389012451663)

Meyer, T. & Lephtin J. (2025), Mobile Jugendarbeit Stuttgart-Innenstadt. Institute for Science & Markets. https://www.mobile-jugendarbeit-stuttgart.de/site/assets/files/9972/abschlussbericht_mja_innenstadt_11_03_2025_endversion.pdf

Reuband, KH. (2008). Kriminalitätsfurcht. In H.J. Lange, H.P. Ohly, & J. Reichertz. (Hrsg.), *Auf der Suche nach neuer Sicherheit* (S. 233-251). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91212-7_14

Röding, D., Soellner, R., Reder, M., ... & Walter, U. (2021). Study protocol: a non-randomised community trial to evaluate the effectiveness of the communities that care

prevention system in Germany. BMC Public Health, <https://doi.org/10.1186/s12889-021-11935-x>

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inferences*. Houghton Mifflin.

Schindler, V., Bässmann, J., Meyer, A., Marks, E., & Linssen, R. (2005). *Beccaria-Standards zur Qualitätssicherung kriminalpräventiver Projekte*. www.beccaria.de/nano.cms/de/Beccaria_Standards1/

Stiftung Deutsches Forum für Kriminalprävention (Hrsg.) (2013). Entwicklungsförderung und Gewaltprävention für junge Menschen. Qualitätskriterienkatalog für die Auswahl und Durchführung wirksamer Programme. https://www.kriminalpraevention.de/files/DFK/dfk-publikationen/2013_dfk_kriterien.pdf

Weißmann, K., Wagner, U., & Maaser, J. (2025). Kommunale Kriminalprävention in Hessen. <https://www.forum-kriminalpraevention.de/files/1Forum-kriminalpraevention-webseite/pdf/2024-03/0324%2009%20Wei%C3%9Fmann.pdf>