

Meta-Evaluation, Evaluationssynthese  
und Entwicklung von Bewertungskriterien

Berlin 2014

### **Impressum**

Camino – Werkstatt für Fortbildung, Praxisbegleitung und Forschung  
im sozialen Bereich gGmbH  
Scharnhorststraße 5  
10115 Berlin  
Telefon (030) 786 29 84  
Fax (030) 785 00 91  
mail@camino-werkstatt.de  
www.camino-werkstatt.de

## **Inhalt**

Einleitung	5
Meta-Evaluation und Evaluationssynthese	6
Meta-Evaluation	7
Methodisches Vorgehen	7
Empirische Umsetzung	11
Ergebnisse der Meta-Evaluation	12
Evaluationssynthese	13
Methodisches Vorgehen	14
Ergebnisse der Evaluationssynthese	14
Was wirkt?	15
Für wen wirkt es?	16
Wie wirkt es?	17
Warum wirken sie?	19
Zusammenfassung der Meta-Evaluation und Evaluationssynthese	20
Kriterienentwicklung im Rahmen von Qualitätszirkeln	22
Erarbeitung von inhaltlichen Bewertungskriterien	22
Zusammenfassung der Bewertungskriterien	25
Literatur	26



## Einleitung

Seit Juli 2013 ist Camino – Werkstatt für Fortbildung, Praxisbegleitung und Forschung im sozialen Bereich gGmbH – damit beauftragt, die Arbeitsstelle Jugendgewaltprävention in Berlin umzusetzen. Auftraggeber ist die Landeskommission Berlin gegen Gewalt. Ziel der Arbeitsstelle ist es, die Umsetzung und Weiterentwicklung des Gesamtkonzepts zur Reduzierung von Jugendgewalt zu begleiten, das das Land Berlin seit 2008 realisiert, um der Jugendgewalt entgegenzutreten.

Die Aufgaben der Arbeitsstelle Jugendgewaltprävention beziehen sich dabei auf die Bereiche „Entwicklung und Fortschreibung eines Evaluationskonzeptes“, „Monitoring zur Entwicklung von Jugendgewaltdelinquenz“, „Durchführung von Veranstaltungen und Workshops“, „Koordination und Kontrolle der Umsetzung des Gesamtkonzepts zur Reduzierung von Jugendgewalt“ sowie – auf Basis der Ergebnisse aus diesen Arbeitsfeldern – die Erarbeitung von Empfehlungen zur Optimierung des Gesamtkonzepts.

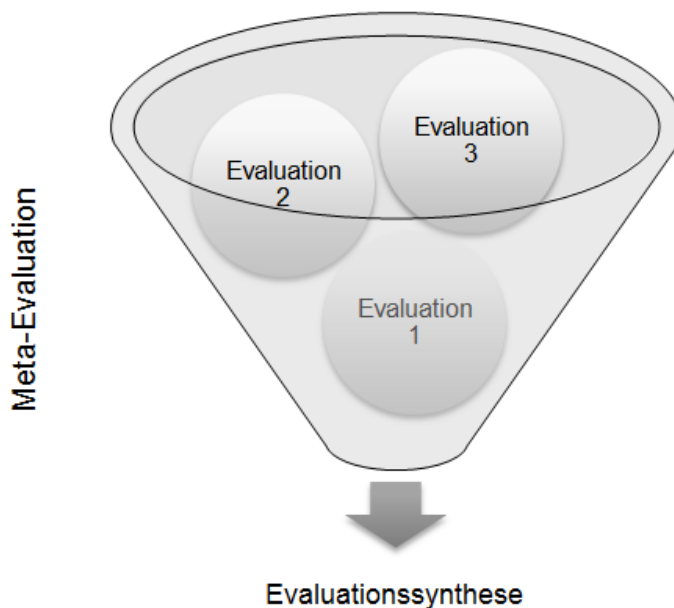
Im Rahmen der Entwicklung und Fortschreibung eines Evaluationskonzeptes wurden Kriterien entwickelt, um Evaluationen unter methodischen und inhaltlichen Gesichtspunkten zu bewerten, und vorhandene Evaluationen zu gewaltpräventiven Projekten gesichtet und bewertet. Dies wird im folgenden Text ausgeführt, der Bestandteil des (unveröffentlichten) Kurzberichtes 2013 der Arbeitsstelle Jugendgewaltprävention ist.

## Meta-Evaluation und Evaluationssynthese

Ein Ziel der Arbeitsstelle Jugendgewaltprävention ist es – neben der Generierung konzeptionellen Wissens und bewertender Kriterien –, belastbare empirische Befunde über die gewaltpräventive Wirkung von Programmen, Maßnahmen und Initiativen aus bereits evaluierten Projekten zusammenzuführen.

Die inhaltlichen Ergebnisse aus einzelnen Evaluationsstudien zu kombinieren, um allgemeine Aussagen über die Reichweite von Interventionen zu treffen, ist das Ziel von Evaluationssynthesen (Caspari 2010, 1; Cooksy/Caracelli 2005, 32). Die Qualität einer solchen Synthese hängt jedoch in hohem Maße von der Qualität der einbezogenen primären Evaluationsstudien ab: Methodisch schwache Evaluationen können die Ergebnisse der Synthese unzulässig verfälschen (Cooksy/Caracelli 2005, 32). Bevor die jeweiligen Evaluationsstudien inhaltlich zusammengeführt werden, muss also zunächst ihre methodische Zuverlässigkeit und wissenschaftliche Güte überprüft werden. Das geschieht im Rahmen einer Meta-Evaluation, die eine „Evaluation der Evaluation“ ist (Caspari 2010, 2; Cooksy/Caracelli 2005, 31; Stufflebeam 2001, 183). Sie ist – kurz gefasst – ein Trichter, mit dessen Hilfe ausschließlich methodisch überzeugende Evaluationen zur Generierung von Ergebnissen verwendet werden (siehe Abbildung 1).

**Abbildung 1: Meta-Evaluation und Evaluationssynthese**



Um die Ergebnisse einzelner Evaluationsstudien zu kombinieren, sind mehrere Schritte notwendig, die von der Arbeitsstelle Jugendgewaltprävention methodisch entwickelt und empirisch umgesetzt wurden.<sup>1</sup> Sie werden im Folgenden kurz skizziert:

- Entwicklung von Bewertungskriterien für die Bestimmung der methodischen Qualität und wissenschaftlichen Güte von Evaluationsstudien
- Systematische Recherche sowie Auswahl geeigneter Evaluationen
- Bewertung der ausgewählten Evaluationsstudien nach den vorab definierten Kriterien
- Qualitative Auswertung der Evaluationsstudien

Bevor die jeweiligen Evaluationsstudien im zweiten Abschnitt also inhaltlich synthetisiert werden, wird zunächst ihre methodische Zuverlässigkeit anhand einer Meta-Evaluation überprüft. Die wesentlichen Befunde der Meta-Evaluation und der Evaluationssynthese, die sich auf bereits vorhandene Evaluationen beziehen, werden im dritten Abschnitt des Kapitels zusammengefasst.

## Meta-Evaluation<sup>2</sup>

Meta-Evaluationen stellen eine „Bewertung von Evaluationsstudien“ (Widmer 1996, 4) dar: Sie bewerten systematisch sowohl den Prozess als auch die Ergebnisse von anderen Evaluationen (Cooksy/Caracelli 2005, 31; DeGEval 2008, 38). Dem Verfahren können verschiedene Typen der Evaluation, wie Programm-, Produkt-, Institutionen-, Personal oder Projektevaluationen unterzogen werden (Stufflebeam 2001, 183). Die Meta-Evaluation selbst kann sowohl formativ als auch summativ angelegt sein (Widmer 1996, 4), kann eine oder mehrere Evaluationsstudien zur Grundlage haben (Cooksy/Caracelli 2005, 32).

### Methodisches Vorgehen

Eine Meta-Evaluation ist zwar eine spezifische Form der Evaluation, aber im Prinzip können und sollen alle Instrumente und Verfahren, die aus der Evaluation bekannt sind, auch für eine Meta-Evaluation fruchtbar gemacht werden (DeGEval 2008, 37; Widmer 1996, 10).

Besondere Bedeutung hat die Entwicklung von Standards bzw. Kriterien, anhand derer die primären Evaluationen bewertet werden.<sup>3</sup> Für die Entwicklung dieser Kriterien kann auf verschiedene Evaluationsstandards, die in Form von Katalogen bei den entsprechenden Fach-

---

<sup>1</sup> Dieser Ablauf ist von der Arbeitsstelle Jugendgewaltprävention in Anlehnung an das von Stufflebeam (Stufflebeam 2001, 189ff) vorgeschlagene Vorgehen bei einer Meta-Evaluation konzipiert worden.

<sup>2</sup> Die Langfassung dieses Kapitels wird im Juni 2014 auf [www.jugendgewaltpraevention.de](http://www.jugendgewaltpraevention.de) veröffentlicht.

<sup>3</sup> Die für eine Meta-Evaluation verwendeten Bewertungskriterien sollten nach Widmer vier Anforderungen erfüllen: Sie sollten möglichst offen, umfassend, breit akzeptiert und detailliert angelegt sein (Widmer 1996, 9ff).

verbänden vorliegen, zurückgegriffen werden. Sie bieten Leitlinien, wie die Qualität von Evaluationen sicherzustellen ist (DeGEval 2008, 19; Widmer 1996, 10). In Deutschland regen die Standards der Gesellschaft für Evaluationen e.V. (DeGEval) an, dass Evaluationen vier grundlegenden Eigenschaften genügen sollten: Sie sollen nützlich, durchführbar, fair und genau sein (DeGEval 2008, 14). Diesen grundlegenden vier Eigenschaften sind jeweils weitere Einzelstandards zugeordnet (DeGEval 2008, 14), sodass sich am Ende 25 Standards für Evaluationen ergeben.<sup>4</sup>

Für die Entwicklung von Bewertungskriterien liegt es nahe, vorrangig solche Standards zu verwenden, die sich auf die methodische Zuverlässigkeit und die wissenschaftliche Güte der Primärstudien konzentrieren. Dementsprechend wichtig sind die Einzelstandards aus der Gruppe der *Genauigkeitsstandards* (G-Standards) (Beywl et al. 2004, 10).<sup>5</sup> Zwar birgt diese Fokussierung den Nachteil, dass tatsächlich nur ein Aspekt der Evaluation überprüft wird. Dennoch hat dies den Vorteil, dass die von Widmer beschriebene Verflechtung einzelner Kriterien untereinander (Widmer 1996, 40) reduziert werden kann. Denn entscheidend für die Qualitätsbewertung ist, dass eine möglichst einheitliche Zuordnung und Einstufung der Primärstudien gewährleistet werden kann.

Dafür bieten die Leitlinien der DeGEval eine gute Orientierung. Die übergeordneten Standards – wie beispielsweise „Beschreibung des Evaluationsgegenstands“ (Genauigkeitsstandard 1) oder „Beschreibung von Zwecken und Vorgehen“ (Genauigkeitsstandard 3) – können inhaltlich übernommen, jedoch mit Hilfe anderer Modelle oder Kataloge weiter ausformuliert und differenziert werden. Dafür werden die „Meta-Evaluation Checklist“ (Stufflebeam 1999) und die Checkliste des CIPP-Evaluationsmodells (Stufflebeam 2007) sowie die für den United Nations Development Fund for Women durchgeführte Meta-Evaluation (UNIFEM Evaluation Unit 2009) herangezogen. Die von der der Arbeitsstelle Jugendgewaltprävention auf dieser Basis entwickelten Kriterien, die zur Bewertung der primären Evaluationsstudien herangezogen werden, sind in der nachfolgenden Tabelle dargestellt.

### **Tabelle 1: Bewertungskriterien für die Meta-Evaluation<sup>6</sup>**

Quelle: (DeGEval 2008; Stufflebeam 1999, 2007; UNIFEM Evaluation Unit 2009)

---

<sup>4</sup> Für eine Übersicht siehe DeGEval 2008.

<sup>5</sup> Die Genauigkeitsstandards beziehen sich auf die Einhaltung wissenschaftlicher Standards und Methoden bei der Planung, Durchführung und Auswertung der Evaluation (DeGEval 2008, 12).

<sup>6</sup> Die Spalte „Kriterienerfüllung/Punktzahl“ wurde lediglich in der ersten Kriteriengruppe „Beschreibung des Evaluandums“ zur Veranschaulichung eingefügt. Aus Gründen der besseren Lesbarkeit wurde in den anderen Kriteriengruppen darauf verzichtet, auch wenn es dort genauso gehandhabt wird.



<b>Beschreibung des Evaluandums (G1)</b>	Ziele des Programms, des Projekts oder der Maßnahme werden expliziert
	Intendierte Outputs, Outcomes und Impacts werden erläutert
	Programmaktivitäten und Verfahren der Programmumsetzung werden beschrieben
<b>Kriterienerefüllung / Punktzahl</b>	Annahmen zu den Wirkweisen des Programms
	4=exzellent    3=gut            2=mittel            1=schwach            0=nicht erfüllt
<b>Analyse des Kontexts (G2)</b>	Soziale, politische, ökonomische, kulturelle, technische, rechtliche Umweltbedingungen werden thematisiert
	Interessen der Betroffenen/Beteiligten
	Externe und interne Projektressourcen werden dargestellt
	Übertragbarkeit der Resultate der Evaluation auf andere Kontexte
<b>Beschreibung von Zwecken und Vorgehen (G3)</b>	Fragestellung wird präzise formuliert (ggf. Neuformulierung während des Evaluationsprozesses)
	Zweck der Evaluation ist konzise erläutert (ggf. Neuformulierung während des Evaluationsprozesses)
	Bewertungskriterien werden nachvollziehbar festgelegt („Operationalisierung“)
	Vorgehen ist genau dokumentiert
<b>Angabe von Informationsquellen (G4)</b>	Deskription der genutzten Informationsquellen
	Unterschiedliche Datenquellen und Datentypen werden herangezogen
	Bericht über die Kriterien und Methoden, mit denen die Informationsquellen gewonnen und ausgewertet wurden
	Einschätzung und Bewertung der Quellen
<b>Valide und reliable Informationen (G5)</b>	Daten sind objektiv, reliabel sowie intern und extern valide
	Daten sind intersubjektiv nachvollziehbar, kommunikativ validiert und/oder trianguliert
	Unterschiedliche Methoden („mixed methods“), um die jeweiligen methodischen Schwächen einzelner Erhebungsverfahren auszugleichen
	Abschließende Validierung
<b>Analyse qualitativer und quantitativer Informationen (G7)</b>	Evaluationsziel und die gewählten Verfahren der Erhebung und Analyse passen zusammen
	Analyseergebnisse leiten sich logisch und konsistent aus den gesammelten und analysierten Daten ab
	Analyseergebnisse werden vor dem Hintergrund der Programmziele bzw. der Programmlogik reflektiert
	Reichweite der Analyseergebnisse wird diskutiert
<b>Begründete Schlussfolgerungen (G8)</b>	Wirkweise des Programms wird reflektiert
	Ergebnisse thematisieren die intendierten als auch die unintendierten Outputs, Outcomes und Impacts
	Ergebnisse thematisieren die geplanten als auch ungeplanten Programmaktivitäten
	Nachvollziehbare und begründete Empfehlungen werden generiert

Ein Kriterienkatalog weist große Vorteile für die Bewertung von Evaluationsstudien auf, dennoch stellen sich auch methodische Probleme, unter denen die Gewichtung der einzelnen Kriterien wohl die größte Herausforderung darstellt (Baumgartner et al. 2002, 65ff.).<sup>7</sup> Wenn gleich alle oben aufgelisteten Kriterien entscheidend für die Beurteilung der methodischen Zuverlässigkeit einer Evaluation sind, dürften einige Kriterien weniger bedeutsam sein als andere. Bevor also über die wissenschaftliche Güte der Evaluationsstudien befunden werden kann, sollte zunächst über eine Gewichtung der ermittelten Bewertungskriterien gemäß ihrer relativen Bedeutung für die Meta-Evaluation im Rahmen einer Nutzwertanalyse entschieden werden.

Die Nutzwertanalyse ist ein multikriterielles Bewertungsverfahren (Harth 2006, 3; Peters/Zelewski 2004, 296).<sup>8</sup> Sie wird vorrangig angewandt, wenn es um eine Bewertung von komplexen Alternativen bei der Entscheidungsfindung geht. Die Nutzwertanalyse stellt eine Weiterentwicklung von einfachen Punktebewertungs- oder Scoring-Modellen dar, wobei mehrere Zielkriterien im Hinblick auf ein Zielsystem gewichtet werden (Bechmann 1982, 799). Ihre Attraktivität für die vorliegende Meta-Evaluation liegt darin, dass sie als eine „qualitative Bewertungsmethode“ (Bundesministerium des Inneren 2013, 313) bezeichnet werden kann. Auch wenn es mittlerweile mehrere Generationen der Nutzwertanalyse gibt, ähneln sich die grundsätzlichen Verfahrensschritte (Bechmann 1982, 800; Bundesministerium des Inneren 2013, 314f.). Zunächst werden die Zielkriterien bestimmt.<sup>9</sup> Danach werden die Zielkriterien entsprechend ihrer Bedeutung gewichtet,<sup>10</sup> sodann wird der Grad der Zielerreichung<sup>11</sup> – meist in Form von Punktzahlen – ermittelt, abschließend der Teilnutzen sowie der

---

<sup>7</sup> Die von Stufflebeam entwickelte „Checkliste“ für Meta-Evaluation baut auf einem einfachen Rating bzw. Scoring-System auf. Mit seinem Ansatz „Numerisches Gewichten und Summieren“ (NGS) hat Michael Scriven ein Verfahren entwickelt, mit dem die einzelnen Kriterien gewichtet werden können (Baumgartner et al. 2002, 65). Allerdings wird u. E. der Frage, wie einigermaßen systematisch gewichtet werden kann, dabei zu wenig Bedeutung geschenkt. Im Unterschied dazu geht Stockmann davon aus, dass eine Gewichtung der Kriterien an und für sich in einer Evaluation kaum gewinnbringend ist (Stockmann 2006, 205).

<sup>8</sup> Zwar wird die Nutzwertanalyse bevorzugt in den Ingenieurwissenschaften sowie in der Raum- und Umweltplanung verwendet, sie ist aber durchaus auch an sozialwissenschaftliche Gegenstände zu adaptieren (Kirchner/Sing 2012, 2; Müller-Herbers 2007, 41). Sie stellt keine wie auch immer geartete „objektive“ Bewertung dar, die unabhängig von den Werturteilen des Bewerbers stattfindet. Dennoch kann die Nutzwertanalyse als ein Instrument oder Hilfsmittel verstanden werden, um Bewertungen transparent und konsistent durchzuführen (Jungbluth 1989, 49; Krems o.J.).

<sup>9</sup> Für die Bestimmung der Zielkriterien sind, wie bereits erläutert, die Evaluationsstandards der DeGEval bzw. anderer Kriterienkataloge benutzt worden.

<sup>10</sup> Für diesen zweiten Schritt existieren unterschiedliche Verfahren: der „paarweise Vergleich“ bzw. die „Präferenzmatrix“ (Bundesministerium des Inneren 2013, 318ff). Kerngedanke ist, dass auf einer Matrix alle Kriterien *paarweise* miteinander verglichen und in ihrer *relativen* Bedeutung bewertet werden, wobei es unerheblich ist, ob es sich dabei um Einzelkriterien oder Kriteriengruppen handelt (Jungbluth 1989, 50). Deshalb wurde eine Präferenzmatrix erstellt, die sich auf die übergeordneten Kriteriengruppen (G1–G8) bezog. Unabhängig voneinander bewerteten die Mitarbeiterinnen der Arbeitsstelle Jugendgewaltprävention die (relative) Wichtigkeit der einzelnen Kriteriengruppen anhand dieser Präferenzmatrix.

<sup>11</sup> Die vorliegenden Evaluationsstudien werden zunächst einzeln bewertet (vgl. das Vorgehen von Stufflebeam 2001, 202). Das geschieht auf der Grundlage der „Bewertungskriterien für die Meta-Evaluation“. Mit der Einstufung, in welchem Ausmaß die (Sub-)Kriterien in den jeweiligen Evaluationen erfüllt wurden, wird eine Punktzahl vergeben,

Gesamtnutzen errechnet. Als Ergebnis entsteht eine Wertsynthese, die einen Gesamtnutzwert – in unserem Fall die methodische Qualität und wissenschaftliche Güte einer Evaluation – übersichtlich ausweist.

### Empirische Umsetzung

Das im vorangegangenen Abschnitt entwickelte Vorgehen wird auf die Evaluationen zu Projekten, Maßnahmen und Programmen im Kontext von Jugendgewaltprävention angewendet. Um die empirische Umsetzung der Meta-Evaluation nachvollziehbar darzustellen, werden die dafür notwendigen Schritte im Folgenden genauer erläutert.

Der *erste Schritt* umfasste die Sichtung und Auswertung einschlägiger Datenbanken sowie eine freie Suche im Internet. Entscheidend für die Auswahl der Datenbanken und Suchmaschinen war, dass diese einschlägig waren. Dabei wurden nicht nur Datenbanken aus dem sozialwissenschaftlichen Bereich gesichtet, sondern auch solche aus dem psychologischen bzw. kriminologischen Bereich. Dementsprechend wurde im KVK, in Sowiport, in KrimDok sowie in OPUS, PsychSpider<sup>ZPID</sup> und JSTOR gesucht.<sup>12</sup> Zusätzlich wurden freie Suchen im Internet durchgeführt. Die dabei verwendeten Schlagworte waren grundsätzlich immer gleich, wurden jedoch handlungsfeldspezifisch angepasst. Die Suche wurde bewusst offen gehalten, um auch Evaluationen zu finden, die möglicherweise durch den Titel oder durch andere Umstände nicht sofort als solche zu erkennen waren.

Durch die Recherche konnten für den Zeitraum zwischen 2000 und 2013 rund 243 Berichte oder Artikel identifiziert werden.<sup>13</sup> Die wenigsten dieser Evaluationen beschäftigten sich jedoch mit in Berlin umgesetzten Maßnahmen, Initiativen und Projekten.

In einem *zweiten Schritt* wurden die Materialien beschafft – ein Arbeitsschritt, der sich als sehr zeitaufwändig herausstellte. Alle beschafften Dokumente wurden in einem *dritten Schritt* daraufhin geprüft, ob und inwieweit sie den Auswahlkriterien genügten. Diese sahen vor, dass es sich dabei um eine Evaluation a) im engeren Sinne<sup>14</sup>, b) von im deutschsprachigen Raum angesiedelten Projekten und c) von im genannten Zeitraum durchgeführten Projekten handelte. Alle Evaluationen, die diese Hürde überwunden hatten, wurden in einem *vierten Schritt* in eine Datenbank eingegeben, in der folgende Informationen gespeichert wurden: Code der Evaluation, Name des evaluierten Projekts, Form der Evaluation, Evalua-

---

wobei diese zwischen null („nicht erfüllt“) und vier („exzellent“) liegt. Jede Kriteriengruppe erhält so eine Endnote, die gewichtet dann in die abschließende Analyse einfließt.

<sup>12</sup> KVK (Karlsruher Virtueller Katalog), Sowiport (sozialwissenschaftliche Fachdatenbank), KrimDok (kriminologische Fachdatenbank), OPUS (Meta-Suchmaschine elektronischer Hochschulschriften), PsychSpider<sup>ZPID</sup> (psychologische Suchmaschine) und JSTOR (Online-Zeitschriftenarchiv).

<sup>13</sup> Die Zahl der tatsächlich evaluierten Projekte dürfte aber deutlich niedriger ausfallen, weil in dieser Liste ein Evaluationsprojekt durchaus auch mit mehreren Publikationen vertreten sein kann.

<sup>14</sup> Siehe die Definition von „Evaluation“ in Fußnote **Fehler! Textmarke nicht definiert.**

tionsort, Handlungsfeld, Zielgruppe, Alter der Zielgruppe, Art des Projekts und der Intervention sowie die Methode(n) der Evaluation.

Für den vorliegenden Bericht 2013 wurden 25 Evaluationen ausgewertet, von denen sich rund zwei Drittel auf überregionale Projekte, Programme und Maßnahmen beziehen. Eine vollständige Übersicht der ausgewerteten Evaluationen befindet sich im Anhang. Um auch erste inhaltliche Aussagen zur Wirksamkeit gewaltpräventiver Maßnahmen treffen zu können, wurden zwischen drei und vier Evaluationsstudien pro Handlungsfeld<sup>15</sup> untersucht, die *thematisch ähnliche Projekte und Maßnahmen* evaluierten. Die Auswahl erfolgte also nicht zufällig, sondern themenbezogen.

### Ergebnisse der Meta-Evaluation

Bei der Bewertung der methodischen Güte der Evaluationen zeigte sich, dass die allgemeine Qualität der untersuchten Evaluationen gut war: Von den 25 betrachteten Evaluationen wurden lediglich fünf aus der weiteren Analyse ausgeschlossen. Das entspricht einem Anteil von rund 20 %. Überwiegend wurden Studien ausgeschlossen, weil sie in der Kriteriengruppe G5 („valide und reliable Informationen“) weniger als zwei Kriterien erfüllten.<sup>16</sup> Dabei handelte es sich sowohl um qualitativ als auch um quantitativ ausgerichtete Evaluationen. In den von uns betrachteten Studien überwogen die summativen Evaluationen, formative Evaluationen stellten eine Minderheit dar.

Auch wenn die allgemeine Qualität der Evaluationen gut war, lassen sich ein paar wiederkehrende Desiderate benennen, welche die methodische Qualität und wissenschaftliche Güte einschränkten.

In den meisten Evaluationen fehlte eine Kontextanalyse. Das heißt, dass sich in den wenigsten Evaluationen Informationen zu den sozialen, politischen, ökonomischen und rechtlichen Umweltbedingungen des untersuchten Projekts oder Programms fanden. Ebenso wenig wurden die für das Projekt zur Verfügung stehenden internen oder externen Projektressourcen beschrieben. Dabei wurden in der vorliegenden Meta-Evaluation auch immaterielle Mittel, wie beispielsweise ehrenamtliches Engagement oder kostenlose Raumnutzung, berücksichtigt. Das Fehlen solcher Informationen erschwert auch die inhaltliche Einschätzung eines Projekts oder Programms. Denn es ist nicht unerheblich, mit welchem Ressourceneinsatz eine gewaltpräventive Wirkung in einem Projekt erzielt wird. Ähnliches lässt sich für die

---

<sup>15</sup> Die von der Arbeitsstelle Jugendgewaltprävention identifizierten präventionsrelevanten Handlungsfelder sind Kinder- und Jugendhilfe, Schule, Polizei, Justiz bzw. Arbeit mit Jugendlichen nach Straffälligkeit (siehe Fußnote 22), Stadtentwicklung und Sport sowie der Bereich Integration und Migration.

<sup>16</sup> Studien, die in der Meta-Evaluation eine zusammenfassende Qualitätsbewertung über „1“ erhielten, die also eine „mittlere“ bis „exzellente“ Qualität aufwiesen, wurden in die anschließende Synthese inkludiert. Evaluationen, die bei der Kriteriengruppe valide und reliable Informationen weniger als zwei Kriterien erfüllten, wurden methodisch als zu schwach bewertet, um in die inhaltliche Synthese aufgenommen zu werden. Sie wurden von einer weiteren Betrachtung ausgeschlossen, selbst wenn sie bei in den anderen Kriteriengruppen bessere Werte erreichten.

Übertragbarkeit der Evaluationsergebnisse auf andere Kontexte feststellen. Bei den allermeisten Evaluationen fanden sich keine Hinweise, ob und unter welchen Bedingungen ein Projekt oder Programm auch an anderen Orten oder für andere Zielgruppen anwendbar ist.

Als eine Schwäche der qualitativ orientierten Evaluationen stellte sich die fehlende Operationalisierung von Bewertungskriterien heraus: So blieb bei einigen Evaluationen ein wenig undeutlich, wie der „Erfolg“ oder auch „Misserfolg“ eines Programms oder eines Projekts gemessen wurde. Das galt zwar mitnichten für alle qualitativen Evaluationen, war aber bei diesen tendenziell häufiger ein Problem als bei den quantitativen Studien. Allerdings, auch das zeigte die Auswertung der Meta-Evaluation, erbrachten die qualitativ orientierten Evaluationen häufiger Hinweise zu den Wirkungen eines Programms als die quantitativ orientierten Arbeiten. In den qualitativen Studien fanden sich gut nachvollziehbare Überlegungen dazu, wie das jeweilige Programm oder Projekt bei der Zielgruppe wirkte. Anders bei den quantitativen Evaluationen: Bei diesen lag in fast allen Fällen eine gute Operationalisierung der Bewertungskriterien (bspw. in Form von psychologischen Tests) vor, aber es wurden keine oder kaum Überlegungen darüber angestellt, warum ein bestimmtes Programm wirkt. Überlegen waren – von wenigen Ausnahmen abgesehen – die Evaluationen, die sich eines multimethodischen Designs bedienten: Sie punkteten sowohl im Bereich der Operationalisierung als auch bei der Beschreibung der Wirkungsweise.

## Evaluationssynthese

Eine Evaluationssynthese zielt, wie eingangs bereits beschrieben, auf eine Kombination der Ergebnisse einzelner Evaluationsstudien, um allgemeine Aussagen über die Reichweite von Programmen, Projekten und Maßnahmen treffen zu können (Caspari 2010, 1; Cooksy/Caracelli 2005, 32).

Eine statistische (Neu-)Berechnung von Daten, mit Hilfe derer die Effekte eines Programms quantitativ abgeschätzt werden sollen, wird oftmals als „meta-analysis“ bezeichnet (Cooksy/Caracelli, 32; Stufflebeam 2001, 187; Widmer 1996, 3). Evaluationssynthesen, die sich mit Studien beschäftigen, die sich qualitativer Methoden bedienen, werden dagegen eher als „narrative Synthesen“ und/oder „research reviews“ bezeichnet (Cooksy/Caracelli 2005, 32; (Widmer 1996, 3).

Bei allen Unterschieden liegen die Vorteile einer Evaluationssynthese auf der Hand: Durch die Kombination unterschiedlicher Studien kann zeit- und ressourcensparend die Aussagekraft einzelner Studien erhöht werden, sollten sich ähnliche Ergebnisse in anderen Studien mit unterschiedlichen Zugängen und Methoden finden (U.S. General Accounting Office 1992, 65; Widmer 1996, 3). Darüber hinaus kann aufgezeigt werden, in welchen Bereichen noch Wissenslücken bestehen (U.S. General Accounting Office 1992, 9).

## Methodisches Vorgehen

Wie eine Evaluationssynthese methodisch umgesetzt wird, ist weitestgehend offen. Grundsätzlich können aber alle Arbeitsschritte oder Verfahren verwendet werden, die für die Synthese von primären wissenschaftlichen Studien entwickelt wurden.

In einem ersten Schritt wird die Fragestellung formuliert, die an das zu untersuchende Material herangetragen wird. Mit der vorliegenden Evaluationssynthese soll die Frage beantwortet werden, ob, und wenn ja, wie gewaltpräventive Projekte und Maßnahmen wirken. In einem zweiten Schritt werden die Befunde aus den unterschiedlichen Studien analysiert und integriert (Cooper/Hedges 2009, 9).

Für die vorliegende Evaluationssynthese sind die Ergebnisse der Evaluationsberichte qualitativ ausgewertet worden. Dafür wurden zunächst die Evaluationsberichte bzw. Teile der Evaluationsberichte in die Software MAXqda<sup>17</sup> importiert. Einzelne Textsegmente wurden den entwickelten Kategorien zugeordnet, somit eine Codierung des Materials durchgeführt. In einem weiteren Schritt wurden die Textsegmente entlang der relevanten Themenschwerpunkte zusammengefasst. Das Vorgehen entspricht einer qualitativen Inhaltsanalyse, bei der das Material kategorienbasiert ausgewertet wird (Mayring 2001, 7).

## Ergebnisse der Evaluationssynthese

Die Frage, die mittels der vorliegenden Evaluationen beantwortet werden sollte war, wie, und wenn ja, für wen und in welcher Form verschiedene Projekte oder Maßnahmen gewaltpräventiv wirken. Um das herauszufinden, wurden die 20 Evaluationen, die eine methodisch ausreichende Qualität aufwiesen, qualitativ ausgewertet. Die in die vorliegende Evaluationssynthese eingeflossenen primären Evaluationen konzentrierten sich auf folgende Themenbereiche:

- Ambulante Maßnahmen und Übergangsmanagement
- Anti-Aggressivitäts-Training und Coolness-Training
- Fußballprojekte gegen Gewalt
- Trainingskurse für Eltern mit Migrationshintergrund
- Soziale Trainingsprogramme an Schulen
- Städtebauliche/kommunale Kriminalitätsprävention
- Mehrfach- und Intensivtäterprogramme

---

<sup>17</sup> MaxQDA ist eine Software für eine (computergestützte) qualitative Datenanalyse.

Etwa zwölf der 20 Evaluationen analysierten in irgendeiner Form „Trainingskurse“<sup>18</sup>. Dass in jedem Handlungsfeld thematisch ähnliche Projekte untersucht wurden, war gewollt, schränkt die Generalisierbarkeit der Ergebnisse jedoch ein.

### Was wirkt?

Alle Evaluationen kommen – mit einer Ausnahme – zu dem Schluss, dass das untersuchte Projekt oder Programm auf verschiedene Weisen gewaltpräventiv wirkt.<sup>19</sup>

Gerade den *Trainingskursen* wird ein Effekt zugesprochen, wobei es dabei zunächst relativ unerheblich ist, ob sich der Trainingskurs primärpräventiv an Schulkinder richtet oder tertiärpräventiv an bereits straffällig gewordene Jugendliche. So kommt eine Evaluation eines sozialen Trainingsprogramms an Schulen zu dem Schluss, dass „die Evaluationsergebnisse für ein universelles primärpräventives Interventionsprogramm sehr zufriedenstellend [sind; d. Verf.]. Aggressives Verhalten wird reduziert und soziale Kompetenzen werden gefördert“ (Spiel et al. 2011, 80). Zu einem ähnlichen Ergebnis kommt eine Evaluation, welche die Wirkung von sozialen Trainingskursen im Rahmen ambulanter Maßnahmen nach richterlicher Weisung untersucht: „Der Vorher-Nachher-Vergleich der gewichteten Delinquenzbelastung ergab bei allen drei untersuchten ambulanten Maßnahmen eine hochsignifikante Reduktion der im Zentral- und Erziehungsregister erfassten Delinquenz“ (Körner 2006, 273).

Wenngleich weniger direkt wirkend, wird auch verschiedenen *Elterntrainingskursen* ein gewaltpräventiver Effekt zugerechnet. In diesen Kursen soll die Erziehungs- und Konfliktlösekompetenz der Eltern erhöht werden. „Grundsätzlich erscheint TESSYA family als ein sehr geeigneter und erfolgreicher Ansatz, um die Verhaltensauffälligkeiten der Jugendlichen nicht isoliert zu bearbeiten, sondern mit Unterstützung der Eltern“ (Schwenzer 2010, 22). Darüber hinaus zeigt sich anhand der untersuchten Evaluationen, dass *Mehrfach- oder Intensivtäterprogramme* der Polizei, bei denen ein täterorientierter Ermittlungsansatz eingesetzt wird, präventiv wirken können. Das gilt auch für Ansätze der *kommunalen Kriminalprävention* und der *städtebaulichen Kriminalitätsprävention*, die zwar weitaus vermittelter wirken, denen aber dennoch ein Effekt auf die Kriminalitätsfurcht oder aber auf die informelle soziale Kontrolle zugewiesen wird. Eine Evaluation zur kommunalen Kriminalitätsprävention fasst ihre Resultate folgendermaßen zusammen: „Die Ergebnisse der Analysen zeigten, dass sich die Kriminalitätsfurcht im Rhein-Neckar-Kreis in den letzten 10 Jahren erheblich verringert hat“ (Hermann 2008, 230).

---

<sup>18</sup> Unter Trainingskursen werden im Folgenden alle Kurse oder Maßnahmen verstanden, die eine gewaltpräventive Wirkung durch *Kompetenzerweiterung* erreichen wollen. Dabei ist es zunächst unerheblich, auf welche Zielgruppe sich diese Maßnahmen fokussieren (so gibt es bspw. Trainingskurse, die sich an bereits delinquente Jugendliche richten, aber es existieren auch primärpräventive Trainingskurse, die sich an alle Jugendlichen richten), entscheidend ist, dass die Gewaltprävention im Vordergrund steht.

<sup>19</sup> Teilweise bezogen sich die Evaluationen nicht nur auf Gewaltdelikte, sondern auch auf andere Delikte (wie z. B. Diebstahl).

Die Ergebnisse einiger Evaluationen zeigen zudem, dass nicht alle Bausteine oder Module eines Projekts oder Programms in gleicher Weise zur gewaltpräventiven Wirkung beitragen. Das zeigt sich beispielhaft an den Trainingskursen mit aggressiv auffälligen oder bereits straffällig gewordenen Jugendlichen, die sich an der Methode der konfrontativen Pädagogik orientieren. Gerade die Teile des Trainings, die einerseits auf eine am „Einzelfall orientierte Umgangsweise mit ‚Dauerproblemen‘ einzelner Teilnehmer/innen“ (Palloks 2004, 169) ausgerichtet sind oder andererseits bestimmte Trainingselemente eingeführt haben, die einen Transfer des Gelernten in den Alltag leisten können (Grieger/Schroer 2002, 75), sind in zwei Evaluationen als besonders gelungen bewertet worden. Als besonders wirksam haben sich in diesen Untersuchungen also gerade die Module des Anti-Aggressivitäts-Trainings herausgestellt, die am wenigsten von der konfrontativen Pädagogik beeinflusst sind.

### Für wen wirkt es?

Grundsätzlich zeichnen die meisten Evaluationen einen hohen Grad der Zielgruppenerreichung für die von ihnen untersuchten Projekte nach. Das ist für Maßnahmen, Projekte oder Programme, die ihre Teilnehmer/innen entweder verpflichtend rekrutieren (z. B. soziale Trainingskurse im Jugendstrafvollzug) oder universal gewinnen (z. B. soziale Trainingskurse in der Schule), allerdings auch nicht weiter verwunderlich.

Größere Relevanz im Hinblick auf die oben formulierte Frage haben die gewaltpräventiven Projekte oder Maßnahmen, die auf freiwilliger Basis beruhen. Sie müssen geeignete Strategien entwickeln, um die Kinder und Jugendlichen, aber auch deren Eltern zu erreichen. Einem Großteil der evaluierten Projekte ist es gelungen, die Zielgruppen zu erreichen, auch wenn es bei dem einen oder anderen zu Beginn einige Schwierigkeiten gab, wie die folgende Evaluation zeigt: „Die Projekte haben zur Hälfte ihre Zielgruppen zumindest teilweise verändert, was als Resultat und Ausdruck anfänglicher Zugangsschwierigkeiten zu bewerten ist“ (Schenkel/Schwenzer 2010, 38). Deshalb stellt das Präventionsprojekt „Straßenfußball für Toleranz“ in dieser Hinsicht eine Ausnahme dar: Hier wird in aller Offenheit formuliert, dass es mit diesem Projekt nicht gelungen ist, die anvisierte Zielgruppe der gewaltbereiten Jugendlichen einzubinden: „Die ‚Straßenfußballer‘ weichen kaum vom Durchschnitt der brandenburgischen Jugendlichen ab. Die Befunde der Fragebogen-Erhebung wie auch der Interviews zeichnen sie als eine Gruppe, die sich weit eher als friedfertig denn als gewaltbereit ausweist. Insofern kann sicherlich nicht davon die Rede sein, dass das Projekt ‚Straßenfußball für Toleranz‘ vornehmlich gewaltbereite Jugendliche angesprochen habe. Vielmehr wird man davon ausgehen dürfen, dass die gesamte Bandbreite der brandenburgischen Jugendlichen erreicht wurde“ (Borković 2010, 344).

Für die Einschätzung der Wirksamkeit ist es ebenfalls wichtig zu wissen, auf wen das Projekt oder die Maßnahme einen Effekt hat. Gibt es – anders formuliert – bestimmte Programme, die für bestimmte Situationen oder auch Zielgruppen besonders wirksam sind? Erste Hinweise geben die Evaluationen: So zeigt eine Untersuchung, die sich mit der Wirksamkeit von



schulischen Präventionsmaßnahmen im Bereich Mobbing beschäftigt, dass sich statistisch deutlich nachvollziehbare Effekte lediglich in Klassen mit einem problematischen Klassenklima zeigten, in allen anderen Klassen veränderte sich hingegen nichts (Landesinstitut für Lehrerbildung und Schulentwicklung der Behörde für Schule und Berufsbildung Hamburg 2009, 17). Auch die Evaluation des Faustlos-Curriculums kommt zu dem Ergebnis, dass „insbesondere bei den Jungen und den Schülerinnen und Schülern mit hohen Aggressionskennwerten ... sich ein statistisch bedeutsamer Rückgang im Vergleich zu Schülerinnen und Schülern mit niedrigeren Aggressionskennwerten zu Beginn des Curriculums und den Kontrollgruppenschülerinnen und -schülern“ (Bowi et al. 2008, 518) zeigt. Ähnlich, wenn auch im Bereich der tertiären Prävention verortet, kommt eine Untersuchung von Mehrfach- und Intensivtäterprogrammen zu dem Ergebnis, dass „die polizeilichen Konzepte zum Umgang mit jungen Mehrfach- und Intensivtätern tatsächlich bei denjenigen Gruppen erfolgreich sind, für die sie gedacht sind. Wenngleich die Datengrundlage nur vorsichtige Schlüsse zulässt, deutet sich doch an, dass besonders junge MIT [Mehrfachintensivtäter/innen; d. Verf.] und jMIT [junge Mehrfachintensivtäter/innen; d. Verf.] mit hoher Belastung an Gewaltdelikten und jMIT [junge Mehrfachintensivtäter/innen; d. Verf.] mit Migrationshintergrund von den Programmen erreicht werden konnten“ (Bliesener et al. 2010, 186).

Soziale Trainingskurse, so das übereinstimmende Ergebnis von zwei Evaluationen, nützen insbesondere Jugendlichen, die eine hohe Sozialkompetenz aufweisen und sich an Gleichaltrigen orientieren: „Jugendliche, die sich für Gleichaltrige interessieren, profitieren von sozialen Trainingskursen“ (Körner 2006, 274). Für andere Jugendliche wiederum kann ein Einzeltraining mit einem Pädagogen/einer Pädagogin hilfreicher sein. Der sogenannte „Heiße Stuhl“, ein besonderes Instrument im Rahmen der konfrontativen Pädagogik, ist – so das Ergebnis – auch nicht für alle Jugendlichen geeignet: „Die Erfahrungen mit dieser Gruppe sprechen dafür, das sich der ‚Heiße Stuhl‘ gut für Jugendliche eignet, die sich schnell provoziert fühlen, die auf Provokationen mit Gewalt reagieren und die bisher nicht in der Lage sind, Provokationen verbal oder durch Ignorieren abzuwehren“ (Grieger/Schroer 2002, 76). Unabhängig von Persönlichkeitsmerkmalen angewendet, birgt der „Heiße Stuhl“ die Gefahr, eher kontraproduktiv die Persönlichkeit eines Jugendlichen anzugreifen.

Übergreifend lässt sich feststellen, dass eine hohe Eigenmotivation eine entscheidende Voraussetzung dafür ist, dass Kinder und Jugendliche von gewaltpräventiven Maßnahmen profitieren.

### Wie wirkt es?

Gewaltpräventive Projekte und Maßnahmen sind nicht nur im Hinblick darauf interessant, ob, und wenn ja, wen sie erreichen, sondern auch bezüglich dessen, wie sie wirken. Die Wirkung wird jedoch, darauf hat die Meta-Evaluation bereits hingewiesen, oftmals nur am Rande thematisiert. Aus der Zusammenschau verschiedener Evaluationsergebnisse lassen sich jedoch einige Erkenntnisse ableiten.

Sieht man einmal von einigen Projekten und Programmen im Strafvollzug oder der Jugendgerichtshilfe ab, wird die Wirkung der Gewaltprävention im weitesten Sinne im Bereich des *sozialen Lernens* und der *sozialen Kompetenzerhöhung* verortet. Hierunter fallen nicht nur die Maßnahmen, die sich an Kinder und Jugendliche richten, sondern ebenso Trainings, die sich an Eltern und/oder Multiplikator/innen richten.

Dabei geht es im primärpräventiven Bereich zunächst darum, dass Kinder und Jugendliche mit den Projekten und Programmen *prosoziales Verhalten* erlernen, das es ihnen ermöglicht, mit schwierigen Situationen, z. B. im Schulalltag, angemessen umzugehen. Die Bereitschaft, sich auf andere einzulassen und mit ihnen zu kooperieren, wird als eine wichtige Erklärungsvariable gesehen. Allerdings, das macht beispielsweise die Evaluation des Faustlos-Curriculums im Bereich Schule deutlich, wirkt das Programm nicht nur durch veränderte Aggressions- und Empathiewerte bei den einzelnen Schüler/innen, sondern auch durch ein verändertes Schulklima: „Konfliktsituationen auf dem Schulhof, in den Pausen, ließen sich einfacher regeln. Weiterhin berichteten die Lehrkräfte eine Reduktion der Eskalationen aufgrund dieser gemeinsamen Regeln und der gemeinsamen Sprache“ (Bowi et al. 2008, 519). Zwar verweist die Evaluation zur Schulsozialarbeit darauf, dass Schulsozialarbeit auch ohne eine „objektive Verbesserung des Schulklimas“ gewaltpräventiv wirkt, folgert aber dennoch, dass „Schulsozialarbeit kriminalpräventiv noch effizienter werden [könnte; d. Verf.], wenn sie verstärkt Ursachen von Gewalt beeinflussen würde, insbesondere das Schulklima“ (Hermann et al. 2011, 145).

Bei gewaltpräventiven Maßnahmen, Projekten oder Programmen, die sich an aggressiv auffällige oder bereits straffällig gewordene Jugendliche wenden, wird die Wirkung etwas spezifischer im Bereich der *Konfliktlösungsfähigkeit* verortet. In den Fällen, in denen die Projekte oder Programme wirken, werden offensichtlich zunächst auf kognitiver bzw. motivationaler Ebene Veränderungsprozesse angestoßen. Jugendliche lernen dadurch auch, gewaltorientierte Handlungsrouninen zu durchbrechen und sich alternative Strategien anzueignen. Das lernen sie, gerade in den sozialen Trainingskursen, über die Einhaltung von Regeln und durch die Interaktion mit anderen: „Positiv ist das Erlernen bestimmter Techniken der Konfliktbewältigung zu werten“ (Palloks 2004, 167). Das Ausmaß, in welchem Lernprozesse nicht nur die kognitive Ebene, sondern auch das tatsächliche Verhalten verändern, wird aus den vorliegenden Evaluationen nicht eindeutig ersichtlich. Dennoch lässt sich festhalten, dass in den meisten Untersuchungen vermutet wird, dass Maßnahmen zur Erhöhung der Konfliktlösungsfähigkeit ihre Wirkung mittels beider Mechanismen entfalten. So kommt die Evaluation zu einem sozialen Trainingskurs zu folgendem Ergebnis: „Neben den dargestellten Entwicklungen auf kognitiver Ebene werden anhand konkreter Konfliktbeschreibungen der Jugendlichen auch Veränderungen im Handeln deutlich“ (Grieger/Schroer 2002, 71). Bei den sekundär- oder tertiärpräventiven Trainingskursen wird – zumindest anhand der meisten Evaluationen – deutlich, dass die Maßnahmen auch wirken, weil sich die Jugendlichen unter pädagogischer Begleitung intensiv mit den eigenen Problemlagen auseinandersetzen. Es sind aber auch die Erfolgserlebnisse, die Jugendliche im Rahmen verschiedener gewaltpräventiver

Maßnahmen erfahren, die ihr *Selbstwertgefühl* stärken – und damit auch ihre Konfliktlösungsfähigkeit positiv beeinflussen: „Für die meisten Jugendlichen ist das Durchhalten des mehrere Monate dauernden Trainings bereits ein nicht zu unterschätzender persönlicher Erfolg, der sich auf das Selbstwertgefühl niederschlagen dürfte“ (Palloks 2004, 168).

### Warum wirken sie?

Auch wenn in den meisten Evaluationen kaum systematisch nach förderlichen und hinderlichen Faktoren gefragt wird, lassen sich anhand der Auswertung einige zentrale Aspekte benennen.

Eine zentrale Bedingung, unter der Projekte und Maßnahmen wirken können, ist – so profan es zunächst klingen mag – die *Akzeptanz durch die Betroffenen und die Beteiligten*. Sie müssen das Gefühl haben, dass das Projekt für sie hilfreich ist oder unterstützend wirkt. Das gilt nicht nur für die Jugendlichen selbst, sondern auch für die Umsetzenden. Gewaltpräventive Maßnahmen an Schulen, wie beispielsweise das Faustlos-Curriculum oder auch Projekte gegen Mobbing, werden offensichtlich dann engagiert von den Lehrkräften implementiert, wenn sie selbst einen positiven Effekt der Programme wahrnehmen: „Die Resonanz auf das Training war bei den durchführenden Lehrerinnen und Lehrern, Kindern und deren Eltern positiv. Die am Projekt beteiligten Lehrkräfte setzten das Curriculum nach Durchlaufen der drei Trainingsjahre erneut mit neuen Eingangsklassen in ihrem Unterrichtsalltag ein“ (Bowi et al. 2008, 519). Gerade für Jugendliche, die nach einer Straftat, auf richterliche Weisung oder im Strafvollzug, an Projekten teilnehmen, muss deutlich werden, dass es sich für sie „lohnt“. Dabei kann sich der Mehrwert auf eine Steigerung der individuellen sozialen Kompetenzen beziehen, aber auch auf konkrete Verbesserungen der eigenen Lebenslage (z. B. eine bessere Integration in den Arbeitsmarkt nach der Haftentlassung). Das zeigt die Evaluation des BASIS-Projekts im Strafvollzug der JVA Adelsheim. „Nach eigenen Angaben fühlen sich die entlassenen Strafgefangenen nach der Teilnahme an BASIS gestärkter im Hinblick auf ihr Leben in Freiheit und es sind ihnen zumindest konkrete Perspektiven aufgezeigt worden, wie sie ihr Leben straffrei weiterführen können“ (Pruin 2013, 711).

Als weitere zentrale Bedingung lassen sich – zumindest für die Präventionsprojekte, die auf ein Mindestmaß an Kooperation angewiesen sind – gute *Netzwerkstrukturen* zwischen verschiedenen Akteuren benennen. Das gilt sowohl für Projekte im Bereich des Übergangsmagements als auch für Projekte der Jugendhilfe: „Zwischen den Mitarbeiter/innen von Polizei (Jugendsachbearbeiter/innen) und IPP kann ein guter Austausch und gegenseitige Akzeptanz der Arbeit festgestellt werden, wenn diese, befördert durch eine räumliche Nähe, wiederholt klientenbezogen zusammenarbeiten“ (Wustmann et al. 2007, 55). Bei Projekten oder Maßnahmen, deren Zielgruppe tendenziell schlechter zu erreichen ist, wie beispielsweise Familien mit Migrationshintergrund, spielt die Einbindung von zentralen Einrichtungen sowie Multiplikator/innen eine förderliche Rolle (Schenkel/Schwenzer 2010, 38).

Als ein eher hinderlicher Faktor haben sich in der Evaluation eines Sportprogramms die standardisierten Abläufe desselben herausgestellt, die mit den unterschiedlichen Erfordernissen vor Ort schlecht in Einklang zu bringen waren (Scheithauer/Hess 2010, 81).

## Zusammenfassung der Meta-Evaluation und Evaluationssynthese

Die Ergebnisse der Meta-Evaluation und der Evaluationssynthese werden abschließend noch einmal theseartig zusammengefasst und aufeinander bezogen. Die Resultate sind jedoch nicht einfach zu generalisieren, weil in jedem Handlungsfeld bewusst thematisch ähnliche Projekte, Programme oder Maßnahmen untersucht wurden.

- Die Qualität der Evaluationen ist gut: Lediglich fünf der untersuchten 25 Evaluationen waren methodisch zu schwach, um in die weitergehende inhaltliche Synthese aufgenommen zu werden.
- Auch wenn die Qualität der Evaluationen als gut bezeichnet werden kann, zeichneten sich bei der methodischen Bewertung der vorhandenen Evaluationen wiederkehrende Schwachpunkte ab: Diese lassen sich im Bereich der Kontextanalyse, der Frage nach der Reichweite bzw. der Übertragbarkeit der Evaluationsergebnisse, der Wirkung der Programme sowie der Operationalisierung verorten.
- Unterschiede zwischen den einzelnen Evaluationen lassen sich dahingehend zuspitzen, dass qualitativ ausgerichtete Studien es versäumen, die der Evaluation zugrunde liegenden Bewertungskriterien offen und nachvollziehbar darzulegen, während quantitativ ausgerichtete Studien wiederum zu wenig darlegen, wie die Programme oder Projekte tatsächlich wirken. Multimethodisch orientierte Studien erweisen sich als überlegen.
- In der inhaltlichen Analyse der Evaluationen hat sich gezeigt, dass die untersuchten Projekte, Programme und Maßnahmen auf verschiedene Weisen gewaltpräventiv wirken. Das gilt für alle untersuchten Handlungsfelder. Gerade den Trainingskursen wird ein gewaltpräventiver Effekt zugeschrieben, unabhängig vom Handlungsfeld. Das Ergebnis ist allerdings vor dem Hintergrund der ausgewählten Studien zu relativieren: Über die Hälfte der untersuchten Evaluationen beschäftigte sich mit Trainingskursen.
- Gewaltpräventive Maßnahmen sprechen bestimmte Zielgruppen tatsächlich an – ein Ergebnis, das beispielhaft an den sozialen Trainingskursen für Jugendliche verdeutlicht werden kann: Besonders profitieren sozial kompetente Jugendliche von den Gruppenkursen, während eher in sich gekehrten Jugendlichen ein Einzeltraining mehr nützen kann.
- Unabhängig davon scheint die Eigenmotivation der Teilnehmer/innen entscheidend zu sein.
- Wie ein Programm oder eine Maßnahme wirkt, bleibt aufgrund der bereits genannten Schwächen überwiegend unklar. Zeigen lässt sich jedoch, dass die Programmwirkung im

Bereich des sozialen Lernens oder der sozialen Kompetenzerhöhung verortet wird. Das gilt gerade für Maßnahmen im primärpräventiven Bereich.

- Als förderlich zeichnet sich eine hohe Akzeptanz des Programms oder der Maßnahme bei den Betroffenen und Beteiligten ab, ebenso wie gute Kooperationsstrukturen.

Für künftige Evaluationskonzepte kann daraus gefolgert werden, dass sie sich, wenn möglich, stärker an multi-methodischen Untersuchungsdesigns orientieren sollten. Zudem sollten der Kontext, die Übertragbarkeit sowie die Operationalisierung in weiteren Evaluationen berücksichtigt werden. Besonderes Augenmerk sollte in dem Konzept zur externen Evaluation von gewaltpräventiven Projekten darauf gelegt werden, dass künftige Evaluationen verstärkt die Annahmen zu den Wirkungen thematisieren und abschließend mit der tatsächlichen Wirkung abgleichen sollten.

Welche Folgerungen sich aus den Qualitätszirkeln für die Evaluationskonzepte ergeben, wird im nachfolgenden Kapitel thematisiert.

## Kriterienentwicklung im Rahmen von Qualitätszirkeln

Zur Sicherung der Qualität der Erhebungen und Ergebnisse der Arbeitsstelle Jugendgewaltprävention dienen Qualitätszirkel, in die Expert/innen aus Wissenschaft und Praxis eingebunden sind. Um den Erfordernissen der unterschiedlichen Handlungsfelder, in denen präventive Maßnahmen und Projekte umgesetzt werden, gerecht zu werden, wurden sieben Qualitätszirkel eingerichtet, die den Handlungsfeldern Kinder- und Jugendhilfe, Schule, Arbeit mit Jugendlichen nach Straffälligkeit,<sup>20</sup> Polizei, Sport, Stadtentwicklung sowie Integration und Migration zugeordnet wurden. Diese Qualitätszirkel haben sich (bis auf eine Ausnahme) im Jahr 2013 konstituiert und werden sich ab 2014 zweimal jährlich treffen.

Zentrale Aufgabe der Qualitätszirkel ist die Beteiligung an der Bewertung künftig vorliegender Evaluationsergebnisse im Rahmen eines expertengestützten Bewertungsverfahrens, das die wissenschaftliche Expertise und die impliziten Wissensbestände der Praxis in Bezug auf eine erfolgreiche Projektumsetzung systematisiert und zusammenführt.

In einem ersten Schritt bestand für die handlungsfeldbezogenen Qualitätszirkel die Aufgabe der Erarbeitung von Bewertungskriterien für die Evaluation von Projekten der Jugendgewaltprävention und -intervention. In einem zweiten Schritt werden sie ab 2014 die Arbeitsstelle darin unterstützen, die Ergebnisse der Evaluationen zu reflektieren, einzuschätzen und auf ihre Plausibilität hin zu überprüfen, sodass handlungsfeldbezogene Modifizierungen in Hinblick auf die Weiterentwicklung der Projektlandschaft diskutiert werden können.

Eine Reflexion der Ergebnisse unter Berücksichtigung des Gesamtkonzepts zur Reduzierung der Jugendgewalt wird in einem übergeordneten Qualitätszirkel stattfinden. Hier werden die Erkenntnisse aus den jeweiligen Handlungsfeldern, der Evaluationen, des Monitorings und der Dunkelfeldbefragung so aufbereitet, dass sie im Sinne einer integrierten Handlungsstrategie in die Weiterentwicklung des Gesamtkonzepts zur Reduzierung von Jugendgewalt einfließen können.

## Erarbeitung von inhaltlichen Bewertungskriterien

Der inhaltliche Schwerpunkt des ersten Treffens der handlungsfeldspezifischen Qualitätszirkel lag auf der Erarbeitung von inhaltlichen Bewertungskriterien für Evaluationen von Projekten der Jugendgewaltprävention und -intervention in den jeweiligen Handlungsfeldern. Dabei ging es nicht darum, konkret messbare Kriterien im Sinne von Indikatoren zu entwickeln,

---

<sup>20</sup> Dieses Handlungsfeld wurde in der ursprünglichen Konzeption und auch in der Umsetzung des Qualitätszirkels 2013 als Handlungsfeld Justiz bezeichnet. Da diese Bezeichnung als irreführend angesehen werden kann – schließlich wird ein großer Teil der Maßnahmen, die sich an Jugendliche nach Straffälligkeit richten, durch Institutionen der Kinder- und Jugendhilfe und nicht der Justiz umgesetzt –, wurde hier ein neuer Begriff eingeführt. In der Dokumentation der Qualitätszirkel (Anhang) wurde der alte Begriff beibehalten, da diese so bereits an die Teilnehmenden versendet wurde.

sondern darum, wichtige Themenbereiche zu identifizieren, die eine Wirkungsevaluation zur Gewaltprävention in den jeweiligen Handlungsfeldern berücksichtigen sollte. Die Qualitätszirkel leisten damit einen wichtigen Beitrag zur Entwicklung des Evaluationskonzepts, die identifizierten Themenbereiche können in Form von Fragestellungen bei der Ausschreibung externer Evaluationen gezielt berücksichtigt werden.

Die Ergebnisse der Arbeit der Qualitätszirkel umfassen zum einen evaluationsrelevante Fragestellungen, die für alle Handlungsfelder bedeutsam sind, und zum anderen Fragestellungen, die jeweils nur für ein spezifisches Handlungsfeld gelten.

- Die Frage nach den Wirkungsweisen eines Projekts wurde in allen Handlungsfeldern gleichermaßen thematisiert. Bereits bei der Evaluationssynthese fiel auf, dass nur wenige Evaluationen dezidierte Aussagen zur Wirkungsweise von Projekten treffen. Ein Verständnis davon, wie bzw. wodurch eine Wirkung erzielt wird, kann aber wichtige Hinweise für die Entwicklung neuer Projekte geben.
- Ein weiterer übergeordneter Themenbereich ist die Frage nach der Nachhaltigkeit von Wirkungen. Dabei geht es nicht um die Fortsetzung bzw. Verstetigung eines Projekts, sondern darum, ob die durch das Projekt oder die Maßnahme angestrebten Ergebnisse bzw. die intendierten Wirkungen Bestand haben können. In der Regel ist es auf Grund der oft kurzen Laufzeit einer Evaluation schwierig, hierzu Aussagen zu treffen, gleichwohl sollten durch eine Evaluation zumindest Anhaltspunkte identifiziert werden, die Aussagen zur Nachhaltigkeit erlauben.
- In nahezu allen Handlungsfeldern wurde die Frage thematisiert, inwieweit sich ein Projekt um einen Lebensweltbezug bemüht: In den Blick genommen werden sollte, inwieweit ein Projekt das gesamte Umfeld eines Jugendlichen berücksichtigt oder ob es nur einen Problemausschnitt behandelt. Zum anderen gehört hierzu aber auch die Berücksichtigung der Frage, inwieweit die inhaltliche Ausrichtung eines Projekts den Besonderheiten einer Zielgruppe gerecht wird.
- Als wesentlich wurde auch die Frage erachtet, ob Projekte, die sich direkt an Jugendliche richten, Effekte im Sinne einer positiven Selbstwirksamkeit erreichen können. In allen Handlungsfeldern ist es aus Sicht der Expert/innen relevant, dass Projekte über ihre eigentlichen Ziele hinaus Jugendliche dabei unterstützen, Handlungskompetenzen zu entwickeln, sich Handlungsressourcen zu erschließen und über Handlungsalternativen zu verfügen.
- In Bezug auf die Projektumsetzung sollen Evaluationen der Frage nachgehen, ob die gewählten Mittel und Methoden mit der Zielsetzung und der Zielgruppe harmonisieren. Auch wenn ein bestimmtes Projekt oder Arbeitskonzept bei einer bestimmten Problemstellung oder Zielgruppe gute Erfolge zeigte, ist dies doch kein Garant dafür, dass es für andere Zielgruppen oder Problemkonstellationen gleichermaßen passt. In Umkehrung

dieser Frage sollten sich Evaluationen damit befassen, unter welchen Bedingungen Projekte auf andere Kontexte übertragen werden können.

- Eine weitere Frage, die Evaluationen in Bezug auf die Projektumsetzung berücksichtigen sollten, ist die nach der Konzepttreue. Nach Ansicht der Expert/innen ist es nicht immer der Fall, dass entwickelte und evaluierte Konzepte in der Praxis tatsächlich in der vorgesehenen Form umgesetzt werden.
- Darüber hinaus stellt sich die wichtige Frage, wie der Transfer der bei der Zielgruppe – im Rahmen der Projektdurchführung – erreichten Wirkungen in den Alltag gelingen kann.

Neben diesen übergreifenden Fragestellungen sollten Evaluationen immer auch die Spezifik des Handlungsfelds und die sich hieraus ergebenden Fragestellungen berücksichtigen. So ist es bspw. für die Evaluation eines Projekts, das in der Schule umgesetzt wird, nicht nur wichtig, danach zu fragen, welche gewaltpräventiven Wirkungen bei den Schüler/innen, also auf individueller Ebene, erzielt werden können, sondern auch Veränderungen des Schulklimas in den Blick zu nehmen, die durch das Projekt angestoßen werden und die auf einer generelleren Ebene ebenfalls gewaltpräventiv wirken. Ein anderes Beispiel für eine handlungsfeldspezifische Frage ist die nach der Beziehungsqualität bzw. -stabilität, die, so die Meinung der Expert/innen, im Handlungsfeld Kinder- und Jugendhilfe Voraussetzung für die Anregung von Verhaltensänderungen ist. Im Handlungsfeld Polizei sollte in einer Evaluation u. a. gefragt werden, wie Kooperationen bei ungleichen Strukturbedingungen, die in diesem Handlungsfeld häufig gegeben sind, gelingen können.

Diese hier nur beispielhaft dargestellten handlungsfeldspezifischen Fragen werden in einem weiteren Schritt systematisiert und fließen in das Evaluationskonzept ein.



## Zusammenfassung der Bewertungskriterien

Folgende Abbildung stellt die Kriterien, die zur Bewertung von Evaluationsergebnissen herangezogen werden, geordnet nach methodischen und inhaltlichen Kriterien dar.

**Abbildung 2: Kriterien zur Bewertung von Evaluationen<sup>21</sup>**



Quelle: (DeGEval 2008; Stufflebeam 1999, 2007; UNIFEM Evaluation Unit 2009)

<sup>21</sup> Dafür wurden, wie bereits im vorangegangenen Abschnitt beschrieben, die übergeordneten Standards der DeGEval DeGEval 2008 inhaltlich übernommen, mit Hilfe anderer Modelle oder Kataloge jedoch weiter ausformuliert. Herangezogen wurden die „Meta-Evaluation Checklist“ Stufflebeam 1999 und die Checkliste des CIPP-Evaluationsmodells Stufflebeam 2007 sowie die für den United Nations Development Fund for Women durchgeführte Meta-Evaluation UNIFEM Evaluation Unit 2009.

## Literatur

## Literaturverzeichnis

Baumgartner, Peter/Häfele, Hartmut/Maier-Häfele, Kornelia (2002): E-Learning Praxishandbuch. Auswahl von Lernplattformen. Marktübersicht – Funktionen – Fachbegriffe. Innsbruck/Wien/Bozen.

Bechmann, Armin (1982): Nutzwertanalyse. In: Albers, Willi (Hg.): Handwörterbuch der Wirtschaftswissenschaft. Stuttgart/New York, S. 799–811.

Beywl, Wolfgang/Speer, Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armut- und Reichtumsberichterstattung. Eine Perspektivstudie. [http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/forschungsprojekt-a323\\_wirkungsorientierte\\_evaluation\\_im\\_rahmen\\_der\\_armutsberichterstattung.pdf?\\_\\_blob=publicationFile](http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/forschungsprojekt-a323_wirkungsorientierte_evaluation_im_rahmen_der_armutsberichterstattung.pdf?__blob=publicationFile), 17.12.2013.

Bliesener, Thomas/Kindlein, Andrea/Riesner, Lars/Schulz, Jan Felix/Thomas, Jana (2010): Eine Prozess- und Wirkungsevaluation polizeilicher Konzepte zum Umgang mit jungen Mehrfach-/Intensivtätern in NRW. Abschlussbericht. [http://www.polizei-nrw.de/media/Dokumente/Behoerden/LKA/CAU\\_Internet\\_%20Abschlussbericht\\_MIT\\_lang.pdf](http://www.polizei-nrw.de/media/Dokumente/Behoerden/LKA/CAU_Internet_%20Abschlussbericht_MIT_lang.pdf), zuletzt aktualisiert am 2010, 17.12.13.

Borković, Vladimir (2010): Evaluation kommunaler Sportprojekte zur sozialen Integration von Heranwachsenden. Potsdam.

Bowi, Ulrike/Ott, Gudrun/Tress, Wolfgang (2008): Faustlos – Gewaltprävention in der Grundschule. In: Praxis der Kinderpsychologie und Kinderpsychiatrie, H. 7, S. 509–520.

Bundesministerium des Inneren (Hg.) (2013): Handbuch für Organisationsuntersuchungen und Personalbedarfsermittlung. [www.orghandbuch.de](http://www.orghandbuch.de), 17.12.2013.

Caspari, Alexandra (2010): Lernen aus Evaluierungen. Meta-Evaluation & Evaluationssynthese von InWEnt-Abschlussevaluierungen 2009. <http://www.giz.de/de/downloads/giz2010-de-meta-evaluation-synthese-lernen-aus-evaluierungen.pdf>, 11.12.2013.

Cooksy, Leslie/Caracelli, Valerie J. (2005): Quality, Context and Use. Issues in Achieving the Goals of Meta-Evaluation. In: American Journal of Evaluation, H. 1, S. 31–43.

Cooper, Harris/Hedges, Larry V. (2009): Research Synthesis as a Scientific Process. In: Cooper, Harris, Hedges, Larry V./Valentine, Jeffrey C. (Hg.): The Handbook of Research Synthesis and Meta-Analysis. New York, S. 3–16.

DeGEval (2008): Standards für Evaluation. <http://www.alt.degeval.de/calimero/tools/proxy.php?id=19074>, 11.12.2013.

- Grieger, Katja/Schroer, Miriam (2002): GAG – Was ist geiler als Gewalt? Anti-Aggressions-Trainings für gewaltbereite Jugendliche. Evaluation eines Modellprojekts. Berlin.
- Harth, Michael (2006): Multikriterielle Bewertungsverfahren als Beitrag zur Entscheidungsfindung in der Landnutzungsplanung unter besonderer Berücksichtigung der Adaptiven Conjoint-Analyse und der Discrete Choice Experimentes. Dissertation, Halle an der Saale.
- Hermann, Dieter (2008): Zur Wirkung Kommunaler Kriminalprävention. Eine Evaluation des Heidelberger Modells. In: Trauma & Gewalt, H. 3, S. 220–233.
- Hermann, Dieter/Dölling, Dieter/Fischer, Sabine/Haffner, Johann/Parzer, Petra/Resch, Franz (2011): Schulsozialarbeit contra Gewalt – eine Evaluationsstudie. In: Bannenberg, Britta/Jehle, Jörg-Martin (Hg.): Gewaltdelinquenz, lange Freiheitsentziehung, Delinquenzverläufe. Mönchengladbach, S. 129–146.
- Jungbluth, Thomas (1989): Beurteilung nicht quantifizierbarer Verfahrensparameter von Feuchtgetreidekonservierungsverfahren mit der Nutzwertanalyse. In: Grundlagen der Landtechnik, H. 2, S. 48–51.
- Kirchner, Andreas/Sing, Dorit (2012): Die Nutzwertanalyse als Instrument für die Sozialplanung. Vortrag in Stuttgart auf der Tagung des Vereins für Sozialplanung (VSOP).  
[http://www.vsop.de/files/RG\\_S\\_2012\\_06\\_22\\_Nutzwertanalyse\\_Instrument\\_Sozialplanung.pdf](http://www.vsop.de/files/RG_S_2012_06_22_Nutzwertanalyse_Instrument_Sozialplanung.pdf), 11.12.2013.
- Körner, Jürgen (2006): Wirksamkeit ambulanter Arbeit mit delinquenten Jugendlichen: Erste Ergebnisse einer vergleichenden Studie. In: Zeitschrift für Jugendkriminalrecht und Jugendhilfe, H. 3, S. 267–275.
- Krems, Burkhard (o.J.): Stichwort Nutzwertanalyse. Online Verwaltungsexikon.  
<http://www.olev.de/n/nwa-kurz.htm>, 11.12.2013.
- Landesinstitut für Lehrerbildung und Schulentwicklung der Behörde für Schule und Berufsbildung Hamburg (2009): Mobbing in der Schule. Evaluationsbericht.  
<http://li.hamburg.de/contentblob/3852762/data/pdf-mobbing-evaluationsbericht.pdf>, 17.12.2013.
- Mayring, Philipp (2001): Kombination und Integration qualitativer und quantitativer Analyse. <http://www.qualitative-research.net/index.php/fqs/article/view/967/2110>, 17.12.2013.
- Müller-Herbers, Sabine (2007): Methoden zur Beurteilung von Varianten. Arbeitspapier der Fakultät Architektur und Stadtplanung der Universität Stuttgart.  
[http://www.igp.uni-stuttgart.de/lehre/1\\_Vorlesung/download/Reader/Methoden.pdf](http://www.igp.uni-stuttgart.de/lehre/1_Vorlesung/download/Reader/Methoden.pdf), 18.12.2013.

Palloks, Kerstin (2004): Cool sein auf Kommando? Konfrontative Pädagogik in der Praxis. Evaluierung einer Umsetzung des Projekt-Formates „Coolness-Training“. In: Unsere Jugend, H. 4, S. 158–170.

Peters, Malte/Zelewski, Stephan (2004): Möglichkeiten und Grenzen des „Analytical Hierarchy Process“ als Verfahren der Wirtschaftlichkeitsanalyse. In: Zeitschrift für Planung und Unternehmenssteuerung, H. 03, S. 295–324.

Pruin, Ineke (2013): Übergangmanagement im Jugendstrafvollzug: Die Evaluation des Projekts BASIS in der JVA Adelsheim. In: Dölling, Dieter/Jehle, Jörg-Martin (Hg.): Täter, Taten, Opfer. Grundlagenfragen und aktuelle Probleme der Kriminalität und ihrer Kontrolle. Mönchengladbach, S. 691–714.

Scheithauer, Herbert/Hess, Markus (2010): fairplayer.sport – Soziale Kompetenzen spielerisch fördern. In: Landeskommision Berlin gegen Gewalt (Hg.): Berliner Forum Gewaltprävention. Nr. 41: Evaluation und Qualitätsentwicklung in der Gewalt- und Kriminalitätsprävention. Berlin, S. 76–84.

Schenkel, Kerstin/Schwenzer, Victoria (2010): Evaluation der Modellprojekte der Landeskommision Berlin gegen Gewalt zur Steigerung der Erziehungskompetenz von Eltern mit Migrationshintergrund und zur Steigerung der Konfliktlösungskompetenz von männlichen Kindern und Jugendlichen mit Migrationshintergrund. In: Landeskommision Berlin gegen Gewalt (Hg.): Berliner Forum Gewaltprävention Nr. 43. Berlin, S. 4–64.

Schwenzer, Victoria (2010): TESSYA family. Tandem-Trainings für aggressiv auffällige Jugendliche und ihre Eltern. Berlin.

Spiel, Christiane/Strohmeier, Dagmar/Schiller, Eva Maria/Stefanek, Elisabeth/Schultes, Marie Therese/Hoffmann, Christine et al. (2011): WiSK Programm: Förderung sozialer und interkultureller Kompetenzen in der Schule. WiSK Evaluationsstudie. <http://www.gemeinsam-gegen-gewalt.at/resources/files/812/wisk-evaluation-endversion.pdf>, 18.12.2013.

Stockmann, Reinhard (2006): Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement. Münster/New York/München/Berlin.

Stufflebeam, Daniel L. (1999): Program Evaluations Metaevaluation Checklist. Western Michigan University. [http://www.wmich.edu/evalctr/archive\\_checklists/program\\_metaeval.pdf](http://www.wmich.edu/evalctr/archive_checklists/program_metaeval.pdf), 11.12.2013.

Stufflebeam, Daniel L. (2001): The Metaevaluation Imperative. In: American Journal of Evaluation, H. 2, S. 183–209.

Stufflebeam, Daniel L. (2007): CIPP Evaluation Model Checklist. Second Edition. [http://www.wmich.edu/evalctr/archive\\_checklists/cippchecklist\\_mar07.pdf](http://www.wmich.edu/evalctr/archive_checklists/cippchecklist_mar07.pdf), 11.12.2013.

U.S. General Accounting Office (1992): The Evaluation Synthesis. <http://www.gao.gov/special.pubs/pemd1012.pdf>, 17.12.2013.

UNIFEM Evaluation Unit (2009): Meta-Evaluation. UNIFEM Evaluations 2004–2008. <http://gate.unwomen.org/unifem/evaluationadmin/manageevaluation/viewevaluationdetail.html?evalid=4592>, 11.12.2013.

Widmer, Thomas (1996): Meta-Evaluation. Kriterien für die Bewertung von Evaluationen. Bern/Stuttgart/Wien.

Wustmann, Cornelia/Stiehler, Steve/Seyfarth, Petra/Wendt, Daniel (2007): Evaluierung des Interventions- und Präventionsprojektes für straffällig gewordene Kinder, Jugendliche und Heranwachsende der Jugendgerichtshilfe Dresden (IPP). [http://jugendgerichtshilfe.dresden.de/media/pdf/jgh/Gg7\\_Ergebnisbericht\\_Evaluierung.pdf](http://jugendgerichtshilfe.dresden.de/media/pdf/jgh/Gg7_Ergebnisbericht_Evaluierung.pdf), 13.15.2013.